Fall 11-30-2017

Transcriptomic Regulation of Alternative Phenotypic Trajectories in Embryos of the Annual Killifish *Austrofundulus limnaeus*

Amie L. Romney
Portland State University

Transcriptomic Regulation of Alternative Phenotypic Trajectories in embryos of the Annual Killifish *Austrofundulus limnaeus*

by

Amie Lynn Thomas Romney

A dissertation submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Biology

Dissertation Committee
Jason Podrabsky, Chair
Suzanne Estes
Bradley Buckley
Todd Rosenstiel
Dirk Iwata-Reuyl

Portland State University
2017

ABSTRACT

The Annual Killifish, *Austrofundulus limnaeus,* survives the seasonal drying of their pond habitat in the form of embryos entering diapause midway through development. The diapause trajectory is one of two developmental phenotypes. Alternatively, individuals can "escape" entry into diapause and develop continuously until hatching. The alternative phenotypes of *A. limnaeus* are a form of developmental plasticity that provides this species with a physiological adaption for surviving stressful environments. The developmental trajectory of an embryo is not distinguishable morphologically upon fertilization and phenotype is believed to be influenced by maternal provisioning within the egg based on observations of offspring phenotype production. However, incubation temperature may override any such maternal pattern suggesting an environmental influence on the regulation of developmental trajectory.

We hypothesize that maternally packaged gene products coordinate the cellular events prior to the maternal-to-zygotic transition (MTZ) that determine developmental trajectory in embryos of *A. limnaeus*. In addition, we propose that environmentally responsive gene expression after the MTZ can sustain or override any such maternal provisioning. Using high-throughput RNA-sequencing, we have generated transcriptomic profiles of protein-coding messenger RNA and noncoding RNA during development in *A. limnaeus*. Embryos destined for either the diapause or escape phenotypes display unique expression profiles immediately upon fertilization that support hormone synthesis, well before the stage when phenotypes are morphologically distinct. At stages when the

trajectories diverge from one another, differential expression of the vitamin D receptor signaling pathway suggests that vitamin D signaling may be a key regulator of developmental phenotype in this species. These data provide a critical link between maternal and environmental influences on the genetic regulation of phenotypic plasticity.

These results will not only impact our understanding of the genetic mechanisms that regulate entrance into diapause, but also provide insight into the epigenetic regulation of gene expression and development. Uncovering genetic mechanisms in a system exhibiting alternative developmental trajectories will elucidate the role of maternal packaging in regulating developmental decisions, and in sustaining metabolic depression during diapause.

DEDICATION

This work is dedicated to my parents, and my parents' parents.

ACKNOWLEDGEMENTS

My dissertation research in the Podrabsky Lab would not have been possible without the support of many people. I am indebted to Jason Podrabsky, my research advisor, who was supportive and encouraging throughout my project. This body of work is a testament to his contribution and insight on all aspects of my research. I would like to thank my graduate advisory committee: Suzanne Estes, Todd Rosenstiel, Dirk Iwata-Reuyl, Bradley Buckley, and Deborah Lutterschmidt. Their advice throughout the years enriched my scientific development. Further research support came from Rahul Raghavan and his advice on bioinformatics analysis. I would also like to thank the many lab mates I have had throughout my dissertation project: Josiah Wagner, Claire Riggs, Daniel Zajic, S. Cody Woll, Ian Garrett, and Kristin Culpepper. Our collaborations fostered a better approach to my own research and our personal relationships further enhanced my doctoral experience. I am grateful to the faculty and graduate students of the Biology department at Portland State for creating an environment that invoked growth and achievement.

TABLE OF CONTENTS

ABSTRACT i

DEDICATION iii

ACKNOWLEDGEMENTS iv

LIST OF TABLES vi

LIST OF FIGURES vii

CHAPTER 1: The influence of genetics and environment on phenotype 1

CHAPTER 2: The maternally provisioned transcriptome in embryos of *A. limnaeus* 10

CHAPTER 3: Small noncoding RNAs along alternative developmental trajectories in *A. limnaeus* 59

CHAPTER 4: Vitamin d is (not) for diapause 89

CHAPTER 5: Connecting gene expression associated with alternative developmental trajectories across life-history stages 118

CHAPTER 6: Ecologically relevant developmental programs 135

REFERENCES 142

APPENDIX A: Tissue sampling and RNA-sequencing from diapause- and escape-bound embryos 162

APPENDIX B: Tissue sampling and small RNA-sequencing from temperature induced diapause- and escape-bound embryos 172

APPENDIX C: WGCNA Analysis 176

APPENDIX D: Supplemental files 208

LIST OF TABLES

Table 2.1 Frequency of poly-A genes at expression bins 51

Table 2.2: Top 20 most abundant poly-A RNA transcripts expressed in 1-2 cell stage embryos of *A. limnaeus*.....52

Table 2.3: Most differentially expressed orthologous genes between *A. limnaeus* and *D. rerio*.....53

Table 2.4: Mitochondrial genes of *A. limnaeus* and *D. rerio* and their expression summary in the 1-2 cell stage transcriptome 54

Table 2.5: Top twenty most abundantly expressed small RNA transcripts expressed across all libraries in 1-2 cell stage embryos of *A. limnaeus*.....55

Table 2.6: Number of genomic alignments for the top 8 most abundant small RNA sequences as a function of the number of bases that must match identically starting from the 5' end.....56

Table 2.7: Differentially expressed sncRNAs in diapause- and escape-bound embryos of *A. limnaeus*.....57

Table 2.8: Number of genes designated in the minimal stress proteome (or MSP) identified in the *A. limnaeus* and *D. rerio* transcriptomes at the 1-2 cell stage 58

Table 3.1: Summary of potential *miR-430* gene targets 88

Table 4.1: Summary of modules and hub genes 117

Table 5.1: Differentially expressed exons in 1-2 cell stage embryos of *A. limnaeus*.....134

LIST OF FIGURES

Figure 2.1. Alternative splicing of poly-A RNA in embryos of *A. limnaeus* that will develop along two alternative developmental trajectories 45

Figure 2.2: Top 10 genes with developmental trajectory-specific splice variants based on statistical significance 46

Figure 2.3: Comparative analysis of poly-A transcriptomes in 1-2 cell stage embryos of *D. rerio* and *A. limnaeus* 47

Figure 2.4: The relationship of the maternally packaged transcriptome of *A. limnaeus* to other teleosts 48

Figure 2.5: Maternally packaged sncRNA transcriptome of *A. limnaeus* 49

Figure 2.6: Rfam database annotation of the maternally packaged sncRNA transcriptome of *A. limnaeus* 50

Figure 3.1: Experimental design for temperature-induced developmental phenotypes 80

Figure 3.2: Small RNA transcriptome characterization 81

Figure 3.3: *miR-10* family miRNA variants in *A. limnaeus* 82

Figure 3.4: *miR-10* family miRNA genes in *A. limnaeus* 83

Figure 3.5: *miR-430* family miRNA variants in *A. limnaeus* 84

Figure 3.6: *miR-430* family miRNA genes in *A. limnaeus* 85

Figure 3.7: miRNA mimic oligonucleotide microinjections 86

Figure 3.8: miRNA knockdown oligonucleotide microinjections 87

Figure 4.1: Transcriptome analysis of genes throughout development of *A. limnaeus* 108

Figure 4.2: Network analysis of gene co-expression along the diapause trajectory 109

Figure 4.3: The light yellow module is not preserved between networks at the cell dispersion stage 110

Figure 4.4: DAF gene homologs in *A. limnaeus* 111

Figure 4.5: Vitamin D3 (VD3) gene homologs in *A. limnaeus* 112

Figure 4.6: Vitamin D Synthesis.....113

Figure 4.7: Vitamin D Receptors 114

Figure 4.8: Light and temperature effects on developmental phenotype.....115

Figure 4.9: Embryonic exposures to Vitamin D precursors 116

Figure 5.1: Intron and exon structures of genes and their annotated mRNA isoforms ...131

Figure 5.2: Frequencies of putative *miR-430* targets throughout development.....132

Figure 5.3: Transcriptomic profiles of *miR-430* targets throughout development 133

Figure 6.1: A model for developmental plasticity in *A. limnaeus* 141

CHAPTER 1:

The influence of genetics and environment on phenotype

The environment is a significant factor in the developmental program of all organisms. Environmental variation represents both an opportunity for and a threat to successful development. On one hand, normal development requires buffering mechanisms against possible environmental perturbations that might interfere with critical molecular and physiological processes. Conversely, meaningful environmental cues can signal need for and induce production of alternative phenotypes that increase the likelihood of organismal success and reproduction based on environmental conditions. In addition, fertilized eggs are supplied with the molecular instructions that coordinate the events of early development, and may even program embryos for a specific phenotype. Thus, developmental mechanisms have evolved to respond to external conditions in at least two competing ways, and phenotypic outcome is a result of integration of inherited factors and the environmental context of the developing embryo. Organisms that can change their phenotype during early life stages (developmental plasticity) offer powerful examples for embryologists to explore the relationship between ecology and developmental biology.

In this thesis I explore the genetic mechanisms and environmental interactions that regulate developmental plasticity in the annual killifish, *Austrofundulus limnaeus*. My dissertation research has revealed that these killifish rely on conserved and likely ancient genetic mechanisms that are directly responsive to environmental influences and

can coordinate the molecular events that result in alternative developmental trajectories. This work highlights the fundamental and conserved genetic networks and programs that underlie animal diversity and adaptation to altered or changing environments.

Alternative developmental phenotypes

Animals across all domains of life have evolved strategies to optimize their developmental program to prevailing environmental conditions. While many developmental effects from environmental cues result in the production of specific structural or functional changes to the organism, the most interesting adaptations are quite possibly those where ontogenetic timing is modified. When conditions are temporarily unfavorable, a delay in developmental progression until a later time can offer an organism a better chance for survival. This is an especially important strategy in communities exposed to periodic and stressful conditions. Diapause, an endogenously cued state of developmental arrest, provides a population with the opportunity to dramatically alter the timing of development to better match with environmental conditions favorable for growth and reproduction.

Annual killifishes (Cyprinidontiformes, Aplocheiloidei) inhabit temporary ponds that dry on an annual basis. While the adults cannot survive without access to liquid water, species persists by the survival of diapausing embryos that are drought-tolerant and resistant to the other harsh conditions imposed when the ponds dry. Survival through the dry season is achieved by increasing the time to complete development through programmed developmental arrest at up to three stages of diapause: diapause I, II, and III.

Diapause I can occur early in development prior to embryonic axis formation during an annual killifish-specific developmental phase known as dispersion and reaggregation of blastomeres. Diapause II occurs after completion of somitogenesis but prior to the major stages of organogenesis. Diapause II is accompanied by extreme metabolic depression and increased tolerance to desiccation and anoxia stresses. Finally, diapause III can occur in late pre-hatching embryos that are essentially fully formed fish larvae ready to hatch.

Entrance into diapause II is an alternative developmental trajectory that is unique in terms of the timing of morphological, physiological, and biochemical traits. Alternatively, a proportion of embryos within the population can bypass diapause II and instead 'escape' from diapause and develop directly to the pre-hatching stage. These escape embryos can presumably complete their entire lifecycle within a single rainy season, while embryos that enter diapause II will have to wait for several months or perhaps over a year to hatch. Having mixed proportions of either phenotype within a clutch of embryos can be advantageous for individual reproductive success and the population as a whole as it provides a mechanism to survive unpredictable or fluctuating conditions as are common in their ephemeral pond habitat.

Predicting Phenotype in A. limnaeus

The regulation of developmental trajectory associated with entrance into diapause II can be influenced by: 1) maternal age and 2) incubation temperature and light experienced by the embryo during early development. During early development, phenotypic trajectory is indistinguishable morphologically (Podrabsky, Garrett et al.

2010). It is not until mid-somitogenesis that embryos developing along either trajectory display obvious traits of metabolic and ontogenetic arrest in diapausing embryos compared to the continued and rapid development of escape embryos. While many physiological aspects of development along the two trajectories have been explored, no investigations have successfully identified the key regulators that induce the diapause phenotype and alter developmental trajectory.

1) Maternal age and endocrine physiology

Development along the diapause and escape trajectory in *A. limnaeus* can be influenced by maternal age (Podrabsky, Garrett et al. 2010). Typically, younger females produce a high proportion of escape embryos and older females produce almost exclusively diapausing embryos. However, offspring from a single spawning event, even when incubated under identical conditions, can differ in developmental outcomes. Moreover, there are a few females within the population who, regardless of age, will consistently produce escape or diapausing embryos throughout their lifetime when the embryos are incubated under standard laboratory conditions of 25˚C.

Previous studies found that maternal production of escape embryos is correlated with increased ratios of 17 ß-estradiol to testosterone in ovary tissue, although estrogen levels were not correlated (Pri-Tal, Blue et al. 2011). Additionally, exposure to high amounts of exogenous estrogen and cortisol can cause embryos to shift from a diapausing to an escape developmental trajectory (Pri-Tal, Blue et al. 2011). A preliminary microarray transcriptome analysis of ovarian tissue from females producing exclusively

escape embryos identified a number of differentially expressed genes involved in steroid hormone signaling, such as nuclear receptor and steroid metabolic enzymes that may contribute to phenotypic regulation (Pri-Tal 2010). Steroid hormones strongly influence organismal physiology and can dictate the developmental outcomes of vertebrates regarding metabolism, sexual dimorphism, and reproduction. As critical signaling molecules that drive large transcriptional changes throughout any life stage, they present a convincing contender for developmental regulation of trajectory in *A. limnaeus*. The levels of steroid hormones required to alter developmental trajectory were much higher than physiological levels measured in embryos, and thus the biological validity of these results under natural conditions are suspect. However, because steroid hormone receptors and other nuclear receptors are known to bind a variety of compounds with various specificities, the non-physiological concentrations of steroid hormones used in these studies could have activated other nuclear receptors yet to be identified (Beildeck, Gelmann et al. 2010).

From an ecological perspective, young females producing a higher proportion of escape embryos would appear to be a form of bet hedging. The earlier in the rainy season that a fish reproduces, the higher the chance that their offspring could complete development and have the opportunity to reproduce again within a single rainy season. Also, many ponds inhabited by *A. limnaeus* are inundated and subsequently dry multiple times during the rainy season. In both of these scenarios, having embryos in diapause III and ready to hatch within a single rainy season could provide a distinct fitness advantage.

Alternatively, entering diapause II can provide a fitness advantage under conditions that are not conducive to supporting multiple generations within a single rainy season.

2) Environmental Cues

While developmental phenotype of *A. limnaeus* can be determined by maternal influences, the embryonic environment can override this programming. The strongest effectors of phenotype are exposure to temperature and light. When eggs from a single female are incubated at temperatures of 30˚C and 20˚C, embryos develop exclusively along escape and diapause trajectories, respectively, regardless of the maternal phenotypic profile (Podrabsky, Garrett et al. 2010). Exposure to light can also drive embryos to develop along the escape trajectory, and there is a synergistic effect of light and temperature (McCracken and Podrabsky, personal observations). Further, there is a critical window when these environmental effects are most potent, or become irreversible. Exposure to light can also break diapause in *A. limnaeus* (Podrabsky and Hand 1999). Investigations in *Nothobranchius*, a genus of African annual killifish, found that during the winter months embryos incubated at 25°C spontaneously entered diapause II, while during the summer months they bypassed diapause and developed along the escape trajectory (Markofsky and Matias 1977, Markofsky and Matias 1977, Markofsky, Matias et al. 1979, Levels and Denuce 1988). Levels and Denuce (1988) concluded that the effect of light is due to the photoperiod experienced by the embryos, rather than the adults. These laboratory studies suggest a pivotal role for photobiology in the regulation of diapause in annual killifishes from both Africa and South America.

The ecological significance of light and temperature in the regulation of developmental trajectory is not fully understood in *A. limnaeus* or any other species of annual killifish. However, in *A. limnaeus* from the Maracaibo Basin of Venezuela, the rainy season is associated with warm and wet conditions while the dry season is typified by cooler temperatures. Thus, higher temperatures favoring the escape trajectory may be advantageous. However, the significance for photobiological control of diapause is more difficult to rationalize given that the embryos exist at the bottom of muddy ponds during the rainy season and buried in clay soils during the dry season. It is important to note that very little is known about the spawning behavior of annual killifishes, and it is possible that females may choose spawning sites, for instance in the shallow pond periphery, where the chances for exposure to higher temperatures and light are more likely. Even in embryos buried in the soil, it is uncertain how much light they may receive. Light may penetrate for several mm or more into even dense clay soils depending on a number of environmental and soil characteristics (Tester and Morris 1987). Perhaps more convincing is the fact that the seeds of many species of plant require light for germination and root growth is known to be affected by light cues (Tester and Morris 1987, Benvenuti 1995, Pons and Fenner 2000, Mo, Yokawa et al. 2015). Thus, photobiology clearly plays a role in the ecology of soil dwelling organisms and should not be dismissed as irrelevant to the regulation of diapause in annual killifishes.

Dissertation Framework

The alternative developmental phenotypes of *A. limnaeus* are examples of complex traits that are almost certainly controlled by multiple molecular factors that are influenced by environmental conditions. In order to investigate these genotypic and environmental interactions, I designed a set of experiments that exploited modern genomics tools combined with the ability to control developmental phenotype using both maternal cues and environmental exposures. This approach required a deep analysis of gene transcription across multiple developmental stages, in embryos exposed to different environmental conditions. I generated profiles of both protein coding and non-coding RNA transcripts in these experiments in order to maximize the ability to identify patterns of gene expression that may explain the biology of diapause in *A. limnaeus*. Using a systems level approach, I was able to identify the gene networks that, when influenced by environmental cues, regulate entrance into diapause in embryos of *A. limnaeus*. Further, I identified a number of potential mechanisms that may explain maternal control of developmental trajectory, and may provide a link between maternal programming and environmental control over development.

This work is an exploration of gene expression using RNA-seq technology to identify the potential key players that regulate developmental trajectory in embryos of *A. limnaeus*. In Chapter 2, the maternally provisioned transcriptomes of fertilized embryos that are destined to develop along either the escape or diapause trajectory are characterized. Expression of both poly-adenylated (poly-A) and small noncoding classes of RNA is explored. These results are compared to other teleosts to better understand the genetic underpinnings of the unique developmental attributes of annual killifish

development and their tolerance to stressful environments. In Chapter 3, the possible roles of small noncoding RNAs in the control of developmental phenotype are explored by characterizing their expression in embryos developing along both developmental trajectories as influenced by incubation temperature. In particular, microRNAs that have the potential to modify the maternally provisioned transcriptome in embryos are explored. In Chapter 4, patterns of poly-A RNA gene expression are profiled in embryos developing at temperatures that produce embryos of both developmental phenotypes. Network bioinformatics analysis is used to identify the key genes that instruct developmental trajectory rather than simply to identify gene expression patterns that are associated with each phenotype. In total, this thesis explores transcriptomic data crosses two life stages of development: the maternally packaged transcriptome at fertilization and the zygotic transcriptome during early development as the two developmental trajectories diverge. To investigate any possible interactions between various classes of RNA and developmental stage, I review the findings from Chapters 2-4 in Chapter 5, and in Chapter 6 I attempt to integrate these studies into a molecular model for the control of diapause in this species. This model suggests mechanisms for how developmental outcome may rely upon maternally inherited cues that are regulated by environmentally responsive networks in the developing embryo.

CHAPTER 2:

The maternally provisioned transcriptome in embryos of *A. limnaeus*

This chapter has been previously published:

Romney, A. L., and Podrabsky, J. E. (2017). Transcriptomic analysis of maternally provisioned cues for phenotypic plasticity in the annual killifish, *Austrofundulus limnaeus*. *EvoDevo*, *8*(1), 6.

Abstract

Genotype and environment can interact during development to produce novel adaptive traits that support life in extreme conditions. The development of the annual killifish Austrofundulus limnaeus is unique among vertebrates because the embryos have distinct cell movements that separate epiboly from axis formation during early development, can enter into a state of metabolic dormancy known as diapause, and can survive extreme environmental conditions. The ability to enter into diapause can be maternally programmed, with young females producing embryos that do not enter into diapause. Alternately, embryos can be programmed to "escape" from diapause and develop directly by both maternal factors and embryonic incubation conditions. Thus, maternally packaged gene products are hypothesized to regulate developmental trajectory, and perhaps the other unique developmental characters in this species. Using high-throughput RNA-sequencing, we generated transcriptomic profiles of mRNAs, long non-coding RNAs (lncRNAs) and small non-coding RNAs (sncRNAs) in 1-2 cell stage embryos of A. limnaeus. Transcriptomic analyses suggest maternal programming of

embryos through alternatively spliced mRNAs and antisense sncRNAs. Comparison of these results to those of comparable studies on zebrafish and other fishes reveal a surprisingly high abundance of transcripts involved in the cellular response to stress, and a relatively lower expression of genes required for rapid transition through the cell cycle. Maternal programming of developmental trajectory is unlikely accomplished by differential expression of diapause-specific genes. Rather, evidence suggests a role for trajectory-specific splice variants of genes expressed in both phenotypes. In addition, based on comparative studies with zebrafish, the A. limnaeus 1-2 cell stage transcriptome is unique in ways that are consistent with their unique life history. These results not only impact our understanding of the genetic mechanisms that regulate entrance into diapause, but also provide insight into the epigenetic regulation of gene expression during development.

Introduction

Development in the annual killifish *Austrofundulus limnaeus* is unique for four major reasons. First, the embryos develop slowly for their size and can enter into a state of developmental and metabolic arrest termed diapause at three distinct developmental stages (Wourms 1972, Wourms 1972). Second, during early development the cell movements associated with epiboly are separated temporally and spatially from gastrulation and formation of the embryonic axis (Wourms 1972, Wagner and Podrabsky 2015). Third, embryonic development is plastic and embryos can develop along at least two alternative pathways based on an interaction of maternal programming and

incubation environment (Podrabsky, Garrett et al. 2010). Finally, embryos of *A. limnaeus* can tolerate and survive extreme environmental stresses, such as long-term anoxia and dehydration (Podrabsky, Riggs et al. 2016). Despite these unique characters, the development of *A. limnaeus* is quintessentially vertebrate, and appears to utilize the same conserved genetic networks that govern development of the typical vertebrate body plan (Wagner and Podrabsky 2015). The mix of unique and apparently conserved developmental characteristics of this species make it an excellent model for examining the evolutionary and mechanistic adaptations of novelty in vertebrate development.

In all vertebrates, the activation of the embryonic genome is delayed for several to many cell divisions following fertilization (Baroux, Autran et al. 2008, Tadros and Lipshitz 2009). During this time, cellular processes are directed by maternal products (RNA transcripts, proteins, ribosomes, and hormones) packaged in the egg during oogenesis (Bowerman 1998, Pelegri 2003, White and Heasman 2008, Shen-Orr, Pilpel et al. 2010). In the zebrafish *Danio rerio*, and most other organisms, maternally derived mRNA transcripts direct early gene expression in the zygote through cleavage and blastulation, but are targeted for degradation coincident with activation of embryonic genome transcription (Kane and Kimmel 1993, Pelegri 2003, Giraldez, Mishima et al. 2006, Mishima, Giraldez et al. 2006, Aanes, Winata et al. 2011). This process has been termed the maternal-to-zygotic transition and coincides with what is known as the mid-blastula transition in some vertebrates (Kane and Kimmel 1993). While many studies have reported on the existence of the maternal-to-zygotic transition during vertebrate development, very few have actually profiled the contents of newly fertilized eggs to

identify maternally contributed gene products (Aanes, Winata et al. 2011, Harvey, Sealy et al. 2013).

Maternally packaged RNAs underlie cellular programming in vertebrate embryos that ensures proper early development such as cleavage, formation of the blastula, and gastrulation (Baroux, Autran et al. 2008, Harvey, Sealy et al. 2013, Paranjpe, Jacobi et al. 2013). Patterns of early cleavage determine distribution of yolk resources, maternally derived factors, and establish the morphogenetic fields that define the vertebrate body plan. Thus, maternally packaged RNAs and their regulation, modification and stability are especially important during the earliest phases of embryonic development. Understanding the key transcripts that must be packaged into an oocyte, and the mechanisms that determine their stability, translatability, and thus their ultimate expression and action could help explain a diversity of developmental phenomena.

There are two major avenues for altered expression of the maternal transcriptome during early development: alternative packaging of transcripts, and modifications of stability or translatability to existing transcripts. For example, alternative mRNA splicing has been found in many vertebrate systems as part of cellular responses to environmental stimuli (Guilgur, Prudêncio et al. 2014, Cotter, Nacci et al. 2016). Alternatively, small non-coding RNAs (sncRNAs) have been implicated in the regulation of gene expression by altering mRNA stability or translation in a variety of contexts including embryonic development (Lee and Ambros 2001, Hannon 2002, Wienholds, Kloosterman et al. 2005, Zhao and Srivastava 2007, Bartel 2009, Barrey, Saint-Auret et al. 2011). The role of sncRNAs in the highly conserved processes of early vertebrate development has received

far less attention than other potential mechanisms of regulation. The ability of a single sncRNA to target multiple transcripts and effect global alterations in gene expression makes this an attractive mechanism for control of gene networks in the absence of an active genome. Many recent studies suggest a critical role of regulatory RNAs in early vertebrate development (Morris and Mattick 2014).

Annual killifish (Aplocheiloidei) are represented by hundreds of species of small tropical and subtropical fishes in Africa and South America (Murphy and Collier 1997). The annual killifish, *Austrofundulus limnaeus*, is native to ephemeral ponds on the coast of Venezuela (Weitzman and Wourms 1967, Taphorn and Thomerson 1978, Hrbek, Taphorn et al. 2005). These ponds are short-lived (several weeks to several months), typically small (3-200 m^2), and offer a harsh developmental environment with large fluctuations in key environmental parameters such as temperature, oxygen partial pressure, pH, and water availability (Podrabsky, Hrbek et al. 1998, Podrabsky, Riggs et al. 2016). These fish grow rapidly to sexual maturity and spawn continuously during their typically short (a few months) adult life, leaving their embryos to survive the dry season. Importantly, an expanding set of genomic tools, including a draft genome assembly (Wagner, Warren et al. 2015) are now available for *A. limnaeus* making it possible to explore genetic and epigenetic mechanisms during development.

Annual killifish embryos are a unique system for examining developmental physiology because they are capable of entering an endogenously cued metabolic dormancy termed diapause as an adaptive phenotype to survive the seasonal drying of their pond habitats (Podrabsky, Tingaud-Sequeira et al. 2010, Furness, Lee et al. 2015).

Diapause can occur at three distinct developmental stages, diapause I, II, III (Wourms 1972). There are unique physiological traits associated with each stage of diapause; however, diapause II embryos show the greatest degree of tolerance to environmental stresses such as desiccation and anoxia (Podrabsky, Riggs et al. 2012, Podrabsky, Riggs et al. 2016). Entrance into diapause II (from here forward referred to as diapause) is one of two possible trajectories during the embryonic development of annual killifish (Podrabsky, Tingaud-Sequeira et al. 2010, Furness, Lee et al. 2015). While a large proportion of embryos enter diapause as their normal mode of development, others are capable of "escaping" diapause and instead develop continuously until the pre-hatching stage (Wourms 1972). Early embryos on either trajectory are indistinguishable, however during somitogenesis the trajectories diverge in both morphological and physiological characters such that the timing of developmental events is unique for each trajectory (Podrabsky, Garrett et al. 2010). The mechanisms that regulate these trajectories are currently unknown, but recent studies suggest possible maternal provisioning during oogenesis or another form of epigenetic programming (Podrabsky, Garrett et al. 2010, Podrabsky, Romney et al. 2016). Typically, younger females produce almost exclusively escape embryos and older females produce almost exclusively diapausing embryos (Podrabsky, Garrett et al. 2010). Yet there is a great deal of inter-individual variation and some females will consistently produce escape or diapausing embryos independent of age. Interestingly, embryonic incubation temperature can override maternal influences and lead to embryos that develop exclusively along the escape and diapause trajectories. Incubation at 20°C results in 100% diapausing embryos, while incubation at 30°C results

in 100% escape embryos (Podrabsky, Garrett et al. 2010). Thus, both the pre-fertilization and post-fertilization environment can affect developmental trajectory. Given the harsh conditions in which these embryos exist and the short duration of pond inundation, the developmental trajectory of an embryo will likely have a profound effect on its survival, and on the survival of the local population. The mechanisms that mediate this type of critical genome-environment interaction are unknown but critically important for understanding the basic mechanisms of development in the natural world.

The events that occur in early development are thought to be some of the most conserved processes in biology. For example, there is striking conservation of function in the genes that regulate blastulation and gastrulation in all animals (Gould 1992, Carroll, Grenier et al. 2013). A great deal of work has gone into characterizing these shared molecular pathways, while relatively few studies have focused on gene expression changes that may underlie plasticity during vertebrate development (Irie and Kuratani 2011). In fact, the vast majority of gene expression studies on developing vertebrates have focused on systems that exhibit little to no intraspecific plasticity in development (Wei, Salichos et al. 2012, Lanes, Bizuayehu et al. 2013, Tingaud-Sequeira, Lozano et al. 2013). Here, we report on the transcriptome of newly fertilized eggs of the annual killifish *A. limnaeus* collected from females that are known to produce 100% escape and 100% diapausing embryos. This paper describes for the first time the maternally derived transcriptome of early embryos of *A. limnaeus*, explores the possibility of maternal control of entrance into diapause through differential packaging of RNA, and uses a comparative approach to identify aspects of the transcriptome that may explain some of

the unique attributes of development in this species compared to more typical teleosts. Evidence is presented that suggests splice variants of genes common to both trajectories and differential packaging of sncRNAs are both possible routes for maternal control of developmental trajectory. Further, comparative analysis of *A. limnaeus* to zebrafish suggests a unique maternally-packaged transcriptome in *A. limnaeus* that is consistent in many ways with the unique developmental patterns observed in annual killifishes.

Methods

Animals

Adult *A. limnaeus* were housed in the PSU aquatic vertebrate facility and cared for according to standard laboratory methods established for this species (Podrabsky 1999) that were approved by the PSU Institutional Animal Care and Use Committee (PSU Protocol #33). This laboratory strain of *A. limnaeus* was originally collected near the town of Quisiro in Venezuela (Podrabsky, Hrbek et al. 1998) and has been cultured continuously in the lab since 1995. Embryos were collected twice a week from 42 pairs of fish during controlled spawning events. Clutches of embryos (20-200 embryos/clutch) were kept separate by spawning date and mating pair. Approximately two hours after each spawning event (1-2 cell stage), all but 10-15 embryos from each clutch were flash-frozen in liquid nitrogen. The remaining embryos were used to determine the proportion of diapause and escape embryos produced by that spawning pair on that date. These embryos were transferred into embryo medium and maintained at 25°C in darkness (Podrabsky 1999) for 17-21 days post-fertilization (dpf) when they were scored as

escape- or diapause-bound embryos using previously established criteria (Podrabsky, Garrett et al. 2010). Total RNA samples were prepared from pools of embryos (30-50 embryos per sample) representing 6 females producing either 100% diapausing or escape embryos (n=6 pooled samples for each trajectory, see APPENDIX A.1 for details).

RNA extraction

Total RNA was extracted from whole embryos using Trizol Reagent (Life Technologies; 50 µl/embryo) according to the manufacturer's instructions for tissues containing polysaccharides. RNA concentration and purity were assessed by UV absorbance at 260 nm and calculating A_{260}/A_{280} ratios (Tecan Infinite M200 Pro with NanoQuant plate, Switzerland). RNA integrity was assessed by agarose gel electrophoresis. RNA samples were maintained at -80°C until use.

Poly-A RNA sequencing

cDNA libraries were prepared using the Illumina TruSeq RNA Sample Preparation kit (v2, Illumina, San Diego, CA, USA) following the manufacturer's instructions with 1 µg of total RNA as starting material. The purified cDNA libraries were quantified by qPCR and their quality confirmed on a 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA) using a DNA 1000 chip. The libraries were sequenced (100 nt paired-end reads, 4 samples multiplexed per lane on the flow cell) on an Illumina HiSeq 2000 at the Oregon Health & Science University (OHSU).

sncRNA sequencing

sncRNA libraries were prepared using the Illumina TruSeq small RNA sample preparation kit following manufacturer's instructions with 1 μg of total RNA as starting material. The RNA samples used to prepare these libraries are the same samples used to prepare the poly-A RNA sequencing libraries. Eleven cycles of PCR amplification were used for library production. Libraries were quantified by qPCR and quality checked on an Agilent Bioanalyzer 2100 DNA 1000 chip. The libraries were sequenced (100 cycles, paired–end, 12 samples multiplexed per flow cell lane) at OHSU using an Illumina HiSeq 2000 (APPENDIX A.2).

Bioinformatics pipeline

A variety of software packages were used to process and analyze the transcriptomic data. CLC genomics workbench (version 6.5) for a Mac Pro (2 x 3.06 GHz 6-Core Intel Xeon and 64 GB RAM) was used for genomic mapping and generation of count data for the sncRNA libraries. The rest of the analyses were performed in a UNIX environment on the Portland State University computing cluster (Dell PowerEdge R730; 2 x Intel(R) Xeon(R) CPU E5-2695 v3 with 28 cores @ 2.30GHz and 256 GB RAM). Differential gene expression and exon usage analyses were performed using the R Bioconductor packages DESeq2 and DEXSeq (Anders, Reyes et al. 2012, Love, Huber et al. 2014). Gene ontologies were assigned using software tools on the PANTHER website (Mi, Poudel et al. 2016).

Analysis of poly-A RNA sequence data

Sequence quality was initially assessed using FastQC, version 0.10.1 (Andrews 2010) to ensure high quality data. Sequence reads were filtered on quality scores and trimmed for the presence of adapter sequences using Trimmomatic (Bolger, Lohse et al. 2014) with the settings "ILLUMINACLIP:2:30:7:1:true", "SLIDINGWINDOW:5:15", "LEADING:20", "TRAILING:20", and "MINLEN: 25". Quality reads were mapped to the *A. limnaeus* genome 1.0 using the *--very-fast-local* preset in Bowtie2 (Langmead and Salzberg 2012). Preserved paired reads after trimming were aligned in paired-end mode and any orphaned mates after trimming were aligned in single-end mode. Reads that aligned to the *A. limnaeus* genome with 0 mismatches were used for expression analyses (APPENDIX A.3). Gene counts (union mode) were generated for all samples using the *summarizeOverlaps* function of the GenomicAlignments package from Bioconductor (Gentleman, Carey et al. 2004) and the NCBI *A. limnaeus* genome annotation Release 100 (Wagner, Warren et al. 2015). Count matrices were filtered for genes with 1 or more normalized counts summed across all replicates. Ontologies of overrepresented genes were determined using PANTHER software with Bonferroni corrected P value < 0.01 (Mi, Poudel et al. 2016).

Gene abundance (FPKM) and differential expression analysis were performed using DESeq2 in the R Bioconductor package. Differential gene expression between diapause- and escape-destined embryos was determined on gene count data using the negative binomial distribution and estimations of mean-variance dependence (Love, Huber et al. 2014). Differential exon usage was tested using DEXSeq (Anders, Reyes et

al. 2012). In both cases, differential expression was evaluated using a Benjamini-Hochberg multiple comparisons adjusted FDR of 10%.

Analysis of sncRNA sequence data

Sequence files were preliminarily examined using FastQC, version 0.10.1 (Andrews 2010) to explore sequencing quality. Small RNA reads were trimmed for quality and adapters with Trimmomatic using the settings "ILLUMINACLIP:2:30:7:1:true", "SLIDINGWINDOW:5:15", "LEADING:20", "TRAILING:20", and "MINLEN: 15". Trimmed reads that aligned to the *A. limnaeus* genome with 0 mismatches were retained and counted using the "annotate and merge" function in CLC-workbench (version 6.5; CLCbio, Arhus, Denmark; APPENDIX A.4). Read counts were normalized across all libraries by DESeq2 to generate a catalogue of small RNAs per treatment (average of 2 counts per million or greater) as well as differential expression based on a $\log_2$ fold change of 1 or greater using the Benjamini-Hochberg multiple comparisons adjusted FDR of 10%.

Sequences were annotated by alignment to small RNA databases allowing up to two mismatches to sequences for *Danio rerio, Fugu rubripes, Tetraodon nigroviridis,* and *Oryzias latipes* using miRBase (v.21), Rfam (v. 12.1) and to all *D. rerio* piRNAs downloaded from Ensembl (Zebrafish V10). Small RNA sequences were mapped to the *A. limnaeus* genome using Bowtie and evaluated for proximity to annotated genes. For those sncRNA sequences annotated as miRNAs, consensus genomic sequences were determined by alignment to the *A. limnaeus* genome with perfect matches. Precursor

stem-loop secondary structures were predicted from expanded regions of 80-100 nucleotides upstream in the 5' direction using the Vienna package RNAfold prediction tool in Geneious software (R 8.1.6).

Comparative Transcriptomics

Orthologous transcript pairs between *A. limnaeus* and *D.* rerio were identified by using NCBI BLASTn to align all RNA sequences of one species to the other species and vice versa. Transcript pairs were determined as Reciprocal Best Hits (RBH) when the sequence alignment between two transcripts resulted in the best matching score for both comparisons. The resultant list was cross referenced to the maternally packaged transcripts dataset of Harvey et al. for *D. rerio* (Harvey, Sealy et al. 2013), and thus gene names for both species are based on the zebrafish annotation for this analysis.

Results

Maternally derived poly-A RNA transcriptome

Details regarding poly-A RNA transcriptome library sequencing and bioinformatics can be found in APPENDIX A.1. RNAseq methods detected 12,329 transcripts in the 1-2 cell stage transcriptome of *A. limnaeus* with a mean expression of 2 or greater fragments per kilobase of transcript per million mapped reads (FPKM), representing about 60% of all sequences in the libraries (Table 2.1, see Supplemental File 2.1 for the complete transcriptome). The 20 most abundant transcripts included nuclear- and mitochondrially-encoded protein-coding, ribosomal RNA, and long non-coding RNA genes (Table 2.2). Gene Ontology (GO) analysis revealed enrichment for highly expressed genes (500 most abundant transcripts; >180 FPKM) in categories that include RNA binding proteins, cytoskeletal proteins as well as redox reaction enzymes and pathways enriched for ATP-synthesis and G-protein coupled signaling (APPENDIX A.5).

Differential expression of poly-A RNA

Gene-level analysis determined that none of the 12,329 genes were differentially packaged in diapause- and escape-destined embryos (*t*-test, FDR > 0.10). However, 57 genes showed differential exon usage between diapause- and escape-destined embryos (Figure 2.1A). GO analysis suggests that the differentially expressed exons are enriched in a number of pathways including glycolysis and insulin signaling (Figure 2.1B). These

exons reside within genes with a variety of functions including: heat shock protein (hsp), *hspa14*; pre-mRNA splicing, *snrnp200*; cytoskeletal proteins, *vinculin*; transmembrane ion transport, *sideroflexin*; cell adhesion, *svep1*; and mTOR signaling of cell growth and proliferation, *ribosome protein s6 kinase.* The 10 genes with the most significant differentially expressed exons (based on FDR adjusted p-values) are presented in Figure 2.2, and the entire list of 57 genes can be found in APPENDIX A.6.

Comparative transcriptomics

Comparison of the transcriptome of 1-2 cell stage embryos from *A. limnaeus* to the same stage embryos of *D. rerio* (as reported by Harvey et al Harvey, Sealy et al. 2013) revealed thousands of species-specific transcripts that likely identify major differences in developmental programs between annual killifish and zebrafish (Figure 2.3). Of the 9,018 transcripts present in the 1-2 cell stage transcriptome of *D. rerio* (≥2 FPKM, protein coding only Harvey, Sealy et al. 2013) less than 10% (841) were identified as orthologous to the transcripts expressed in *A. limnaeus* (Figure 2.3). Comparatively, 89% (10,997) and 91% (8,174) of the expressed transcripts (≥2 FPKM, protein-coding only) were unique to *A. limnaeus* and *D. rerio*, respectively. The group of 841 genes that were expressed in both species shared similar patterns of abundance. However, there were significant differences in expression of these transcripts between the two species, with the most differentially expressed transcripts coming from mitochondrially encoded genes, *claudin*, and *ubiquitin*, among others (Table 2.3). Transcripts encoded in the mitochondrial genome are 2-4 fold higher in most cases for *D.*

rerio compared to *A. limnaeus* (Table 2.4). While there are only 3 comparable datasets available in the literature for 1-2 cell stage embryos (*Hippoglossus hippoglossus* (Bai, Solberg et al. 2007), *D. rerio* (Harvey, Sealy et al. 2013) and data presented here for *A. limnaeus*), it appears that each species has a unique expression pattern for the 20 most abundant maternally packaged transcripts (Figure 2.4A). When comparing the 100 most abundantly expressed transcripts in the *D. rerio* and *A. limnaeus* 1-2 cell stage transcriptomes (Figure 2.4B-D), the *A. limnaeus* transcriptome is enriched in GO terms for ion binding and transport as well as cytoskleletal structure and function (Figure 2.4D). Alternatively, the *A. limnaeus* transcriptome is underrepresented compared to *D. rerio* in transcripts with GO terms for metabolic processes (Figure 2.4D, APPENDIX A.7).

Small RNA transcriptome

Details of the small RNA library sequencing and bioinformatics results can be found in APPENDIX A.4. The *A. limnaeus* 1-2 cell stage sncRNA transcriptome is characterized by 3,379 sncRNA transcripts expressed at a level of ≥ 2 normalized counts per million reads (Supplemental File 2.2). To better examine the diversity of sequence reads we grouped only identical sequences, not making any assumptions about mature sequence length, clustering, or shared gene origin. At fertilization, embryos possess a wide diversity of sncRNA sequences. The majority of the sequence reads were short in length (<20 nt) while longer sequences (>25 nt) were less abundant (Figure 2.5A). However, analysis of unique sncRNA sequence abundance as a function of sequence

length revealed two dominating size classes; the highest abundance size class was 16 nt while the second most abundant was 26 nt (Figure 2.5B).

Between 57-73% of sncRNA sequence reads in each sample library could be annotated to known RNAs based on sequence alignment with miRBase Release 21 and Rfam version 12.1 (Griffiths-Jones, Grocock et al. 2006, Nawrocki, Burge et al. 2014). All sncRNAs were also annotated based on sequence alignment to a database of piwiRNA (piRNA) sequences in RNAdp v. 2.0 (Pang, Stephen et al. 2005). Annotations of miRNA sequences were similar in Rfam and miRBase, and the following summary of annotations is based on the Rfam results only. Of the 3,379 unique sncRNA sequences, the majority had lengths below 20 nt and were not annotatable, while only 21% (722) annotated to known RNAs (Figure 2.5C). The sncRNA sequences with the highest abundance annotated as antisense RNAs (55% of total reads, Figure 2.6A) with the remainder including fragments of ribosomal RNAs (<1% of total reads), and small nucleolar RNAs (<1% of total reads; Figure 2.6B,C). Surprisingly, miRNA annotations comprised <1% of the sncRNA transcriptome in 1-2 cell stage embryos (Figure 2.6A-C). Of the 22 unique mature sequence variants that annotated as miRNAs, consensus sequences were generated for *miR-181* and *miR-10*. High confidence precursor sequences, based on sequence and secondary structure modeling were prepared for submission to miRBase for *Alim-miR-10* and *Alim-miR-181* (Figure 2.6D).

The 3,379 unique sncRNA sequences mapped to approximately 33,000 locations in the genome, with 61% mapping to intergenic regions and 39% within exons (Supplemental File 2.2). The remaining alignments were positioned within introns (8%)

or as a combination of categories (1%). Most sequences (2,796) mapped to the genome in multiple locations (2-100). Only 583 mapped to exactly one position, with 51% of these in intergenic regions, 21% in introns, 27% in exons. Interestingly, 41% aligned in an antisense orientation, and 4 sequences mapped to the mitochondrial genome (3 of 4 to the *ND3* gene).

The two most abundant sncRNA sequences had average counts per million reads of 168,769 and 79,972 and annotated as antisense RNAs in Rfam, while the third-most abundant (43,648 counts), annotated as rRNA based on genome alignments (Table 2.5). The top twenty most abundantly expressed sncRNAs are dominated by sequences unknown to Rfam. However, they appear to be unique and non-repetitive sequences as determined by a low number of alignments to the genome (1-2 alignments each with perfect match of leftmost 15 bases). However, loosening the parameters just slightly increases the genomic alignments to over 100 locations (Table 2.6).

Differential expression of sncRNAs

Only four sncRNA sequences (≥ 2 normalized counts) were differentially expressed between embryos on the escape and diapause trajectories (Table 2.7). These four sequences do not match any known RNA sequences in Rfam or miRBase. The most-differentially expressed gene sequence, *ACAACGTGTGATACA*, aligned once to the genome in an intronic region of a gene predicted to be *zinc finger protein 646*-like (*LOC106517361*; APPENDIX A.8). The other differentially expressed sequences had two alignment positions in the genome (using strict parameters aligning the left-most 15

nt with 0 mismatches). Exploring the role of these sncRNAs is critical to the understanding of their possible involvement in post-transcriptional modification of other components of the 1-2 cell stage transcriptome.

Discussion

This study demonstrates for the first time a profile of maternally packaged RNAs for *A. limnaeus*. Due to the critical nature of early developmental events across multiple time scales (individual lifespan to evolutionary time) it is highly likely that each of the transcripts identified in this study plays an essential role in supporting early development and diapause in this species. This work serves as a foundation of information from which to build working hypotheses concerning the maternal control of the unique developmental attributes of annual killifishes.

General patterns in the poly-A RNA transcriptome

Perhaps not surprisingly, the protein-coding transcriptome of 1-2 cell stage embryos from *A. limnaeus* shares many similarities with fertilized eggs from other teleosts when analyzed on a presence-absence basis (Bai, Solberg et al. 2007, Vesterlund, Jiao et al. 2011, Pauli, Valen et al. 2012, Harvey, Sealy et al. 2013, Lanes, Bizuayehu et al. 2013) and GO analysis indicates similar function between the transcriptomes of *A. limnaeus* and *D. rerio*. Early vertebrate development is thought to be highly constrained and conserved, and thus it is reasonable to hypothesize that similar gene expression patterns would be required in all species. However, for the 3 species for which data are directly comparable (same developmental stage and sequencing methodology) each species appears to have a unique pattern of highly abundant transcripts with only 2 of the top 20 transcripts shared across all three species. In a direct comparison of *A. limnaeus* and *D. rerio*, each species expresses an order of magnitude more transcripts that are

unique, compared to those that are shared. This species-specific expression in the maternally derived transcriptome likely reflects the different developmental strategies and environments of each species.

Stress tolerance

Early development, specifically the cleavage stages, tends to be sensitive to environmental stress. It is certainly the case in *A. limnaeus* that stress tolerance is lowest during the first 4 days of development (Machado and Podrabsky 2007, Podrabsky, Lopez et al. 2007, Podrabsky, Riggs et al. 2016). It has been proposed that over-expression of stress tolerance genes may have a negative effect on development in some species. For example, heat shock that induces HSP expression in zebrafish embryos leads to a high proportion of abnormal embryos (Krone, Lele et al. 1997). Also, over-expression of the molecular chaperone HSP 70 in *Drosophila* larvae increased thermal tolerance, but also caused a notable decrease in developmental rate and had other negative effects (Feder, Cartaño et al. 1996, Krebs and Feder 1998). In contrast, reduction of two small molecular weight chaperones, *p26* and *artemin*, appears to slow development in diapause-bound embryos of *Artemia* (King and MacRae 2012, King, Toxopeus et al. 2014). In other studies, inhibition of HSP 90 led to teratogenic effects due to an unmasking of cryptic variation in *Drosophila* and zebrafish (Rutherford and Lindquist 1998, Queitsch, Sangster et al. 2002, Yeyati, Bancewicz et al. 2007). Thus, buffering the effects of environmental stresses is likely an essential component of development, but may affect other processes such as developmental rate, or phenotypic expression. The transcriptome of *A. limnaeus*

has a number of characters that suggest an increased ability to mitigate the potentially negative impacts of environmental stress. A better understanding of these characters may shed light on how embryos can evolve to tolerate high levels of environmental stress while preserving essential processes required for normal development.

The small molecular chaperone, *hsp 27* (*hspb1*) is one of the top 20 expressed transcripts (>3000 FPKM) in *A. limnaeus* and is represented at a substantially higher level than in zebrafish (2.1 FPKM; Harvey et al (2013), and was not detected in Atlantic halibut (Bai, Solberg et al. 2007). In another report, *in situ* hybridization did not detect *hspb1* in zebrafish embryos until the gastrula stage (Mao and Shelden 2006). In a study of *Fundulus heteroclitus*, Tingaud-Sequiera et al. (2013) identified the transcriptome of pooled embryonic stages (including the 2-4 cell stage) during exposure to air and addressed the expression of larger inducible heat shock molecular chaperones such as 40, 70, and 90 kDa classes, but made no mention of any small HSPs. In zebrafish, HSP 27 possesses the ability prevent protein precipitation, and is induced under heat stress in post-blastula stage embryos (Haslbeck, Franzmann et al. 2005, Rupik, Jasik et al. 2011). Further, small HSPs have been shown to play a critical role in the survival of heat stress in adult desert fishes (Norris, Brown et al. 1997, Hightower, Norris et al. 1999) and are highly differentially expressed in response to fluctuating daily temperatures in adult *A. limnaeus* (Podrabsky and Somero 2003). Thus, the high abundance of *hspb1* in the *A. limnaeus* 1-2 cell stage transcriptome appears to be unique, and may contribute to their ability to survive stresses imposed by their harsh developmental environment.

Protection from reactive oxygen species is essential for normal cell function, and is thought to be critical in mediating survival of oxygen deprivation (Vanden Hoek, Becker et al. 1998, Levraut, Iwase et al. 2003). Embryos of *A. limnaeus* are the most anoxia tolerant vertebrate (Podrabsky, Lopez et al. 2007, Podrabsky, Riggs et al. 2012, Podrabsky, Riggs et al. 2016), and can develop normally even under extreme hypoxia (Anderson and Podrabsky 2014). Thus, it may not be surprising that maternal provisioning of antioxidant systems would be elevated in this species. Some of the most highly expressed unique sequences in the *A. limnaeus* transcriptome are involved with glutathione metabolism, a key mechanism to deal with oxidative stress (Pompella, Visvikis et al. 2003). In addition, transcript abundance for the superoxide dismutase genes (*sod1 and sod2*) are packaged at a much higher level in embryos of *A. limnaeus* (863 and 46 FPKM) compared to *D. rerio* (87 and 19 FPKM). Maternal packaging of *sod1* is thought to protect embryos from oxidative damage during development, when metabolic demands are high (Lin, Tseng et al. 2009). However, the low metabolic rate of early *A. limnaeus* embryos may suggest other reasons for maternal packaging of *sod1* that are more consistent with survival of environmental stress, perhaps playing a role in their unique tolerance of oxygen deprivation.

All organisms possess cellular mechanisms that help maintain homeostasis in the face of environmental stress; many of these mechanisms are activated in the cellular stress response (CSR). A minimal stress proteome (MSP) was described by Kültz (2004) that included 44 proteins and protein families that participate in the CSR. Analysis of the 1-2 cell stage transcriptomes from *A. limnaeus* and *D. rerio* suggests a similar number of

genes represented in the MSP that are maternally packaged (Table 2.8). These transcripts represent 2% or less of the transcriptome, and less than 1% of the abundant 25% of transcripts. Based on this admittedly limited evaluation, neither species appears to be particularly enriched for potentially stress-responsive gene expression. However, the nature of the most abundant transcripts from this list is different in the two species. While mitochondrial genes critical for metabolism are highly abundant in both species, they dominate the most abundant transcripts in *D. rerio*, while in *A. limnaeus* the two most highly abundant MSP genes are a molecular chaperone and a thioredoxin (Supplemental File 2.3). These data suggest some potentially important functional differences in the stress responsive transcripts that are maternally packaged in these two species that warrant future attention.

Slow developmental rate

A relatively slow developmental rate is a general character of all Atherinid fishes including the cyprinodonts (Parenti 1981). Even within the Aplocheiloid fishes (annual killifish and their relatives), embryos of annual killifish tend to develop much more slowly compared to closely related non-annual relatives. Recently, rates of cleavage were found to be significantly slower in annual killifish compared to non-annual killifish in three lineages that are thought to have independently evolved the annual life history (Dolfi, Ripa et al. 2014). Thus, when compared to rapidly developing species such as zebrafish, the transcriptome of 1-2 cell stage embryos of *A. limnaeus* may provide clues

to the molecular underpinnings of slow development in Atherinids and specifically for annual killifishes.

A simple comparison of the *D. rerio* and *A. limnaeus* transcriptomes suggests a higher abundance of cell cycle associated transcripts in zebrafish. For example, of the top 20 unique and highly abundant transcripts in zebrafish, there are 4 cyclins, 2 histones, and a number of other transcripts that are expected to be high in proliferating cells. In contrast, these transcripts are not found among the top 20 unique and highly abundant transcripts in *A. limnaeus*. In addition, 9 of the 20 most differentially expressed transcripts between *D. rerio* and *A. limnaeus* are mitochondrially-encoded and are statistically more abundant in *D. rerio*. While these transcripts may be maternally provisioned, it has also been demonstrated that active transcription of the mitochondrial genome occurs very early in development, and well before activation of the nuclear genome (Heyn, Kircher et al. 2014). Thus, the observed differences in mitochondrially-derived transcripts could arise either through maternal provisioning, differences in the rates of mitochondrial transcription, or perhaps both. While the abundance rank of these transcripts is high in both *D. rerio* and *A. limnaeus*, they are generally 2-4 times more abundant in the *D. rerio* transcriptome. This higher representation of mitochondrially-encoded transcripts is consistent with a more active transcription of the mitochondrial genome in *D. rerio*, compared to *A. limnaeus* and this may perhaps lead to higher rates of mitochondrial activity.

One remarkable difference between embryos of either species is in the profile of transcripts encoding for zinc finger proteins. A total of 462 zinc finger protein transcripts

were present in *A. limnaeus* compared to 165 in *D. rerio*. Many of the genes in both datasets corresponded to counts less than 100 FPKM. However, the RNA-binding protein, *zinc finger protein 36*, *C3H1 type-like 2* (*LOC106511212* or *Zfp36l2* in other species), was particularly high in *A. limnaeus* with a gene count of approximately 3,500 FPKM while in *D. rerio*, the paralogs of this gene (*zfp36l1*, and zfp36l2) have counts of 8 and 36 FPKM, respectively. The gene *zfp36l2* is shared among all vertebrates and is known to promote mRNA decay through interactions with the 3'-untranslated region (UTR) (Zhang, Prak et al. 2013). This protein has been described as part of a common mechanism that can induce cellular quiescence through posttranscriptional regulation of mRNA stability. Recently it has been shown that the mRNAs degraded through this mechanism code for proteins that promote cell cycle progression into the S-phase (Galloway, Saveliev et al. 2016). The high level of expression for this particular transcript in *A. limnaeus* could be a mechanism for slowing progression through the cell cycle during development. It is also reasonable to hypothesize that this transcript could play a critical role in the regulation of developmental arrest associated with diapause in this species.

The homeobox transcription factor, *nanog* is known to play a role in the maintenance of pluripotency in a number of animal models (Camp, Sánchez‐Sánchez et al. 2009). However, in the Medaka, *nanog* acts instead as a critical regulator of cell proliferation during early development that supports cell-cycle progression into the S-phase (Camp, Sánchez‐Sánchez et al. 2009). Transcripts for *nanog* are abundant (1,011 FPKM) in fertilized embryos of zebrafish (Harvey, Sealy et al. 2013). When the protein sequence

for *D. rerio nanog* (AEZ64150.1) was compared to protein sequences of *A. limnaeus*, (NCBI BLASTp) the best resulting match was *LOC106530841,* which annotated as *homeobox protein DLX-1-like* (36% identity, e-value=$3e^{-31}$). Interestingly, expression of this transcript in *A. limnaeus* is much lower (317 FPKM) than that of *nanog* in zebrafish. The apparent altered role for *nanog* in fish development and the lower abundance in embryos of *A. limnaeus* point to another potential mechanism that could slow cell cycle progression and potentially contribute to the ability of these embryos to arrest development in diapause.

Evolutionary biologists have been discussing the potential importance of developmental rate, and conducting laboratory selection experiments to alter this rate for decades (Hubbs 1926, Marien 1958, Chippindale, Alipaz et al. 1997, Prasad, Shakarad et al. 2000, Strathmann, Staver et al. 2002). However, mechanistic studies that can explain differences in developmental rate are lacking with only a few studies focusing on heterozygosity in a variety of metabolic enzyme isoforms (Danzmann, Ferguson et al. 1986, DiMichele and Powers 1991, DiMichele and Westerman 1997). Thus, the differences pointed out here between *A. limnaeus* and *D. rerio* may be the first attempt to evaluate, on a global level, the potential cellular mechanisms for altering developmental rate. We do not have enough data at this point to infer causality, but this study certainly points to a number of likely candidates for future evaluation.

Regulation of developmental trajectory through alternative mRNA splice variants

There are 57 exons that are differentially expressed in the two alternative developmental trajectories in *A. limnaeus*. These exons are found in genes with a myriad of functions including, DNA/RNA binding, protein degradation, intracellular signaling, posttranscriptional and posttranslational modifications, the cellular stress response, cytoskeletal and transport properties, as well as cellular adhesion. Most of the exons (86%) are more abundant in diapause-destined embryos and absent or only rarely expressed in escape-bound embryos. It is important to note that the presence or absence of an exon changes not the gene dosage, but rather likely alters the regulation or activity of the gene product. Given the wide range of cellular pathways in which these gene products function, changes in their regulation could have profound effects on cellular and organismal physiology. GO analysis indicates an enrichment in these genes with roles in a number of signaling pathways that are known to regulate rates of cell proliferation. It is especially interesting that the insulin/IGF pathway is enriched, given its known role in regulating diapause in *C. elegans* and arthropods (Kimura, Tissenbaum et al. 1997, Sim and Denlinger 2013, Sim, Kang et al. 2015). This may indicate a role for IGF signaling in the regulation of diapause in *A. limnaeus*, which would suggest a perhaps universal role for IGF signaling during animal diapause. It is also interesting that genes important for glycolysis are represented because diapausing embryos of *A. limnaeus* are known to be poised for anaerobic metabolism compared to escape embryos (Chennault and Podrabsky 2010). A role for alternative splicing in the regulation of fish physiology is supported by previous work on estrogen receptors in killifish exposed to estrogenic compounds (Cotter, Nacci et al. 2016).

While it is beyond the scope of this paper to explore every gene that is represented in this list, one of the mRNA splice variants highly packaged in diapause-bound embryos (over 60-fold higher than in escape-bound embryos) is found within the gene *phkb* (*phosphorylase kinase, beta subunit*). A recent study of this protein suggests it is a modulator of *hsp 27* (Chebotareva, Makeeva et al. 2010), one of the most abundant transcripts in the *A. limnaeus* 1-2 cell stage transcriptome. In the milkfish, expression of this gene was reduced following exposure to thermal and salinity stress (Hu, Kang et al. 2015). In the Medaka, *phkb* is a key regulator of glycogen synthesis and contributes to calcium and insulin signaling pathways (Kanehisa, Sato et al. 2016). Given the noted importance of insulin signaling to the regulation of diapause across a variety of animal species, this transcript holds exceptional promise as a target for future functional studies, and suggests a potential avenue for alteration of developmental trajectory by incubation temperature.

The discovery of these differentially packaged mRNA variants suggest that alternative splicing rather than differential gene expression may be critical for determining maternal influence on developmental trajectory in *A. limnaeus*. This is a novel result and suggests a subtle role for changes in gene regulation in the control of diapause in this species. Future studies will be required to test for the influence and function of these genes in response to temperature and other factors that are known to affect developmental trajectory.

The small non-coding RNA transcriptome

A high diversity of sncRNAs, such as demonstrated here, is common in early development stages of animals and is likely important for the proper regulation of gametogenesis and fertilization (Bourc'his and Voinnet 2010, García-López, Alonso et al. 2015). To our knowledge this is the first report of the sncRNA transcriptome of a 1-2 cell stage fish embryo. The profiles of abundance, diversity and sequence length distribution presented here for *A. limnaeus* are similar to those reported for the zebrafish 256-cell stage embryo (Wei, Salichos et al. 2012). However, it is worthy of note that few mature miRNA sequences were identified here in *A. limnaeus*, while a great variety of miRNAs are present in zebrafish at the 256-cell stage, and in later developmental stages of *A. limnaeus* (Romney and Podrabsky, unpublished observations). This may indicate that mature miRNAs are not a common or major component of the maternally packaged transcriptome in fish embryos.

In *C. elegans*, miRNAs are known to be key regulators of the developmental switch associated with entrance into the diapause-like dauer state (Rougvie and Moss 2013). Recent studies suggest diapause-specific miRNAs in flesh flies as well (Reynolds, Peyton et al. 2017). In addition, miRNAs are critical for reactivation of delayed implanting mouse embryos (Liu, Pang et al. 2012, Cheong, Pang et al. 2014). The lack of miRNA diversity in 1-2 cell stage embryos of *A. limnaeus*, and the fact that no miRNA transcripts appear to be differentially expressed in association with either developmental trajectory suggests that maternal control over entrance into diapause is not likely mediated through any currently described miRNA. It is possible that some of the novel

sequences discussed below may act as miRNAs, but more work is needed to explore this possibility.

The high diversity of sncRNAs identified in embryos of *A. limnaeus* fall into two basic categories: those having many sequence variants with low abundance, and those with few sequence variants that are highly abundant. While a functional role for the low abundance sequences cannot be ruled out, it is also possible that these are the products of RNA degradation (Harding, Horswell et al. 2013). However, it is important to note that degradation is a perfectly acceptable means to generate biologically active molecules and thus these sequences should not be summarily dismissed. However, given the large number of sequences involved, we have chosen to focus on the highly abundant transcripts that have few variants. We propose a role for these RNAs similar to sncRNA such as endogenous small interfering RNAs (endo-siRNAs) or regulatory genes derived from longer RNAs such as tRNAs, rRNAs and small nucleolar RNAs (snoRNAs). Understanding the role of these potentially novel sncRNAs could elucidate gene regulatory pathways that support alternative developmental trajectories in this species.

The most abundant sncRNA sequences annotated as antisense RNAs, specifically the ST7 antisense RNA 1 conserved region 1 (RF02179). In humans, this long non-coding RNA is involved in the cellular response to DNA-damage (Liu, Sun et al. 2015), and is described as a tumor suppressor gene (Vincent, Petek et al. 2002, Haverty, Lin et al. 2016), but there is a paucity of information available. When aligned in an antisense orientation in the *A. limnaeus* genome these sequences map to RNA genes, zinc finger proteins, as well as a gene annotated as synaptophysin-like (involved in synaptic

vesicular transport). When mapped in a sense direction, the sequences align to regions in rRNA genes suggesting they could be generated from excision of internal spacer RNAs from rRNA transcripts. Small RNAs derived from rRNAs and tRNAs have recently been described as important regulators of cell differentiation (Harding, Horswell et al. 2013, García-López, Alonso et al. 2015). It is particularly interesting that these sncRNA sequences may target zinc finger proteins, which appear to be enriched in the transcriptome of *A. limnaeus*. Obviously, the complete picture of the significance of these sncRNAs cannot be determined by expression profile alone, and functional studies are needed to draw additional conclusions.

Only four sncRNAs were packaged into embryos at statistically different levels in embryos developing on the two developmental trajectories. All of these sequences are novel, and thus we currently do not know enough about their biology to make valid predictions for their mode of action or potential targets. However, if we make the assumption they are acting as antisense RNAs to block translation or induce degradation of mRNA targets, a cohesive story emerges that is consistent with a role for these sncRNAs in the regulation of cell proliferation in escape embryos and the potential for chromatin remodeling control of gene expression in diapause-bound embryos. For instance, three of the potential target sequences for sncRNAs that are more abundant in the escape-bound embryos are mRNA transcripts that encode for proteins known to regulate cell cycle progression. One sequence is antisense to an intron in a zinc finger protein 646-like sequence that contains a SFP1 domain. SFP1 in yeast is a known transcriptional repressor that regulates ribosomal protein expression and blocks the G_2/M

transition of the cell cycle (Xu and Norris 1998, Marion, Regev et al. 2004). Another sequence could specifically target two *A. limnaeus* genes. One target, RNA-binding motif, single-stranded-interacting protein 3 (RBMS3), is known to bind the c-Myc promotor and reduce cell proliferation through alteration of β-catenin expression in at least two types of human cancer (Chen, Kwong et al. 2012, Liang, Liu et al. 2015). Another target for this sncRNA could be the transcript for endosialin-1, also known as TEM-1 in humans. In humans and mice, high TEM-1 expression is associated with pericyte proliferation and angiogenesis and acts through platelet-derived growth factor receptor pathways that lead to activation of immediate early genes like *c-fos* (Tomkowicz, Rybinski et al. 2010, Maia, DeVriese et al. 2011). The function of endosialin-1 is currently not understood during early development, and has not been characterized in *A. limnaeus* or any other fish. Reduction of this protein could reduce proliferation if it acts as it does in mammals. However, what we know about this protein is restricted to its role in pericyte cells of adult mammals, and it is possible that a context-dependent role for endosialin-1 during early development could be critical for other essential functions such as cell migration or differentiation. The important aspect of this protein is that it contains the functional domains necessary to interact with cell proliferation signaling pathways and thus has the potential for regulation of cell proliferation or perhaps migration. Another potential targeted mRNA transcript in escape-bound embryos codes for a neurexin-2-like protein. The transcript for this protein is maternally packaged in *Xenopus*, and is highly expressed during early development in zebrafish, although its function is unknown (Zeng, Sharpe et al. 2003, Rissone, Monopoli

et al. 2007), and thus the role that this protein my play in determination of developmental trajectory in *A. limnaeus* remains unclear. The one sncRNA sequence that is upregulated in diapause-bound embryos targets an uncharacterized protein with SANT and reverse transcriptase domains. Proteins with SANT domains are known to be highly expressed in proliferating cells especially during early development and have a role in the regulation of chromatin organization through histone acetylation (Boyer, Langer et al. 2002, Mo, Kowenz-Leutz et al. 2005). Thus, reduction of this protein could have large-scale effects on gene expression. Unfortunately, we do not know when or where in the developing embryo that these sncRNAs may be active, and future studies to inhibit or modify their expression are needed to evaluate their potential role in altering developmental trajectory in *A. limnaeus*.

Conclusion

This study reports for the first time the maternally-packaged transcriptome of an annual killifish, and represents one of only a handful of similar studies on early vertebrate embryos. The data support a transcriptomic poise that could support the unique characters of development in *A. limnaeus* compared to other fishes. Surprisingly, the transcriptome of *A. limnaeus* and *D. rerio* are more unique than similar, even at the 1-2 cell stage. Unlike the transcriptomes of other teleost species, *A. limnaeus* embryos contain an abundance of gene transcripts that support the slow developmental rates and stress tolerance specific to their life history. For the first time, splice variants of mRNAs are identified that are differentially packaged into oocytes in a vertebrate with alternative

developmental trajectories. In addition, the sncRNAs that are differentially packaged suggest an important mechanism for post-transcriptional control that may contribute to the regulation of developmental trajectory, and support the unique embryonic characters in this lineage. Studies such as this are essential to broaden our understanding of the evolution of developmental and phenotypic plasticity.

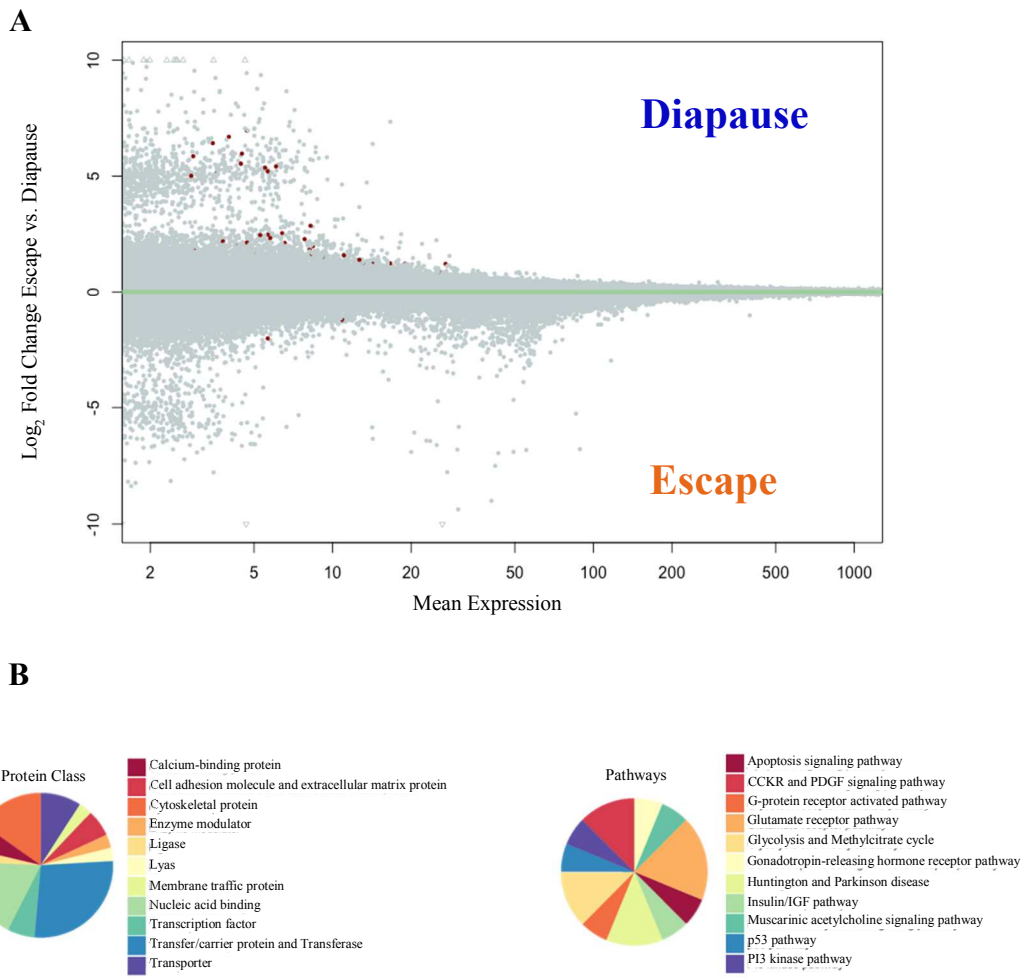


Figure 2.1. Alternative splicing of poly-A RNA in embryos of A. limnaeus that will develop along two alternative developmental trajectories. (A) Differential exon usage in mRNA gene transcripts that are packaged into diapause- and escape-destined 1-2 cell stage embryos of A. limnaeus. Of the 57 exons that were significantly different between trajectories (red symbols, FDR <0.1, t-test) 49 are upregulated in diapause-bound embryos, while only 8 are upregulated in escape-bound embryos. (B) GO term analysis for transcript variants between diapause- and escape-bound embryos of A. limnaeus ($P < 0.05$) suggests enrichment for exons expressed in genes for a variety of molecular and metabolic pathways including glycolysis and the Insulin/IGF signaling pathway.

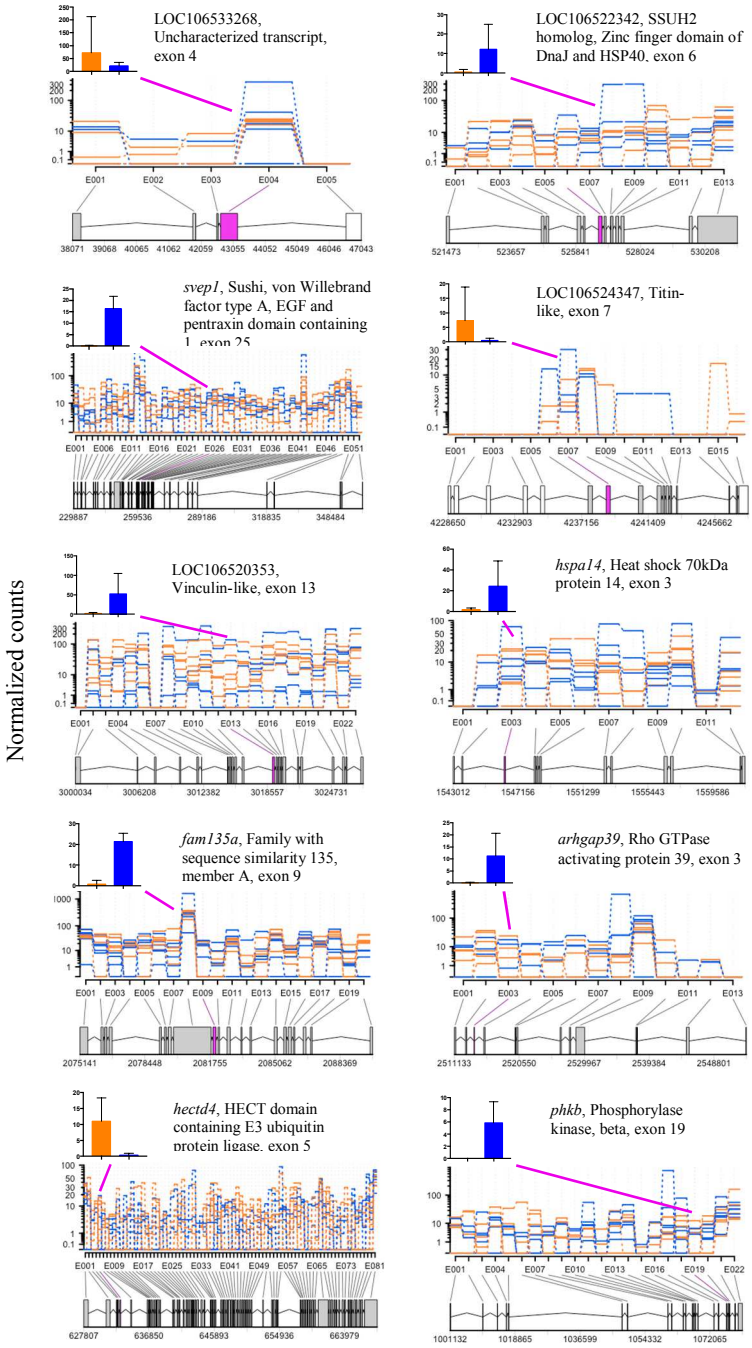


Figure 2.2: Top 10 genes with developmental trajectory-specific splice variants based on statistical significance. Each biological replicate is graphed separately in the exon usage graphs with orange lines indicating escape-bound embryos and blue lines indicating diapause-bound embryos. The x-axis on the plots indicates the exon number and the mapping location of the exon on the appropriate contig from the *A. limnaeus* genome file. Note that the y-axis is a log scale, which tends to mask the differential expression of the exons, and thus we have provided a bar graph on a linear scale to better illustrate the mean (± S.D.) levels of expression for the differentially expressed exon within each gene. Blue bars indicate diapause-bound embryos, while orange bars represent escape-bound embryos.

A. Twenty most abundant genes uniquely expressed in *A. limnaeus* and *D. rerio*.

FPKM	Gene	Description
6,652	*phc2b*	polyhomeotic homolog 2b
3,445	*siva1*	SIVA1, apoptosis-inducing factor
3,129	*hnrnpa0l*	heterogeneous nuclear ribonucleoprotein A0, like
2,925	*thy1*	Thy-1 cell surface antigen
2,558	*EPCAM*	epithelial cell adhesion molecule
2,523	*cldng*	claudin g
2,412	*h1m*	linker histone H1M
2,349	*ccnb1*	cyclin B1
2,261	*hmgb2a*	high mobility group box 2a
2,244	*CCNA2 (2 of 2)*	cyclin A2
2,207	*cth1*	cysteine three histidine 1
1,896	*hnrnpa1*	heterogeneous nuclear ribonucleoprotein A1
1,826	*HNRNPAB (2 of 2)*	heterogeneous nuclear ribonucleoprotein A/B
1,759	*h3f3a*	H3 histone, family 3A
1,668	*nme2b.2*	NME/NM23 nucleoside diphosphate kinase 2b, tandem duplicate 2
1,603	*ctsl1a*	cathepsin La
1,554	*zgc:112425*	cold inducible RNA binding protein a
1,428	*nasp*	nuclear autoantigenic sperm protein (histone-binding)
1,382	*ccnb2*	cyclin B2
1,371	*ccna1*	cyclin A1

FPKM	Gene	Description
6,063	*gsr*	glutathione reductase
5,095	*gstcd*	glutathione S-transferase, C-terminal domain containing
4,425	*acads*	acyl-CoA dehydrogenase, C-2 to C-3 short chain
3,735	*pot1*	protection of telomeres 1
3,711	*gstk1*	glutathione S-transferase kappa 1
3,569	*pou2f1*	POU class 2 homeobox 1
3,229	*gstz1*	glutathione S-transferase zeta 1
3,190	*ppap2a*	phosphatidic acid phosphatase type 2A
3,158	*gtf2b*	general transcription factor IIB
2,949	*nudt13*	nudix (nucleoside diphosphate linked moiety X)-type motif 13
2,891	*cnp*	2',3'-cyclic nucleotide 3' phosphodiesterase
2,803	*pld6*	phospholipase D family, member 6
2,540	*ppap2c*	phosphatidic acid phosphatase type,
2,217	*gtf2e1*	general transcription factor IIE, polypeptide 1, alpha 56kDa
2,134	*gtf2e2*	general transcription factor IIE, polypeptide 2, beta 34kDa
2,115	*ppapdc1b*	phosphatidic acid phosphatase type 2 domain containing 1B
2,059	*ppara*	peroxisome proliferator-activated receptor alpha
2,001	*gtf2f1*	general transcription factor IIF, polypeptide 1, 74kDa
1,879	*gtf2f2*	general transcription factor IIF, polypeptide 2, 30kDa
1,775	*gtf2h1*	general transcription factor IIH, polypeptide 1, 62kDa

B

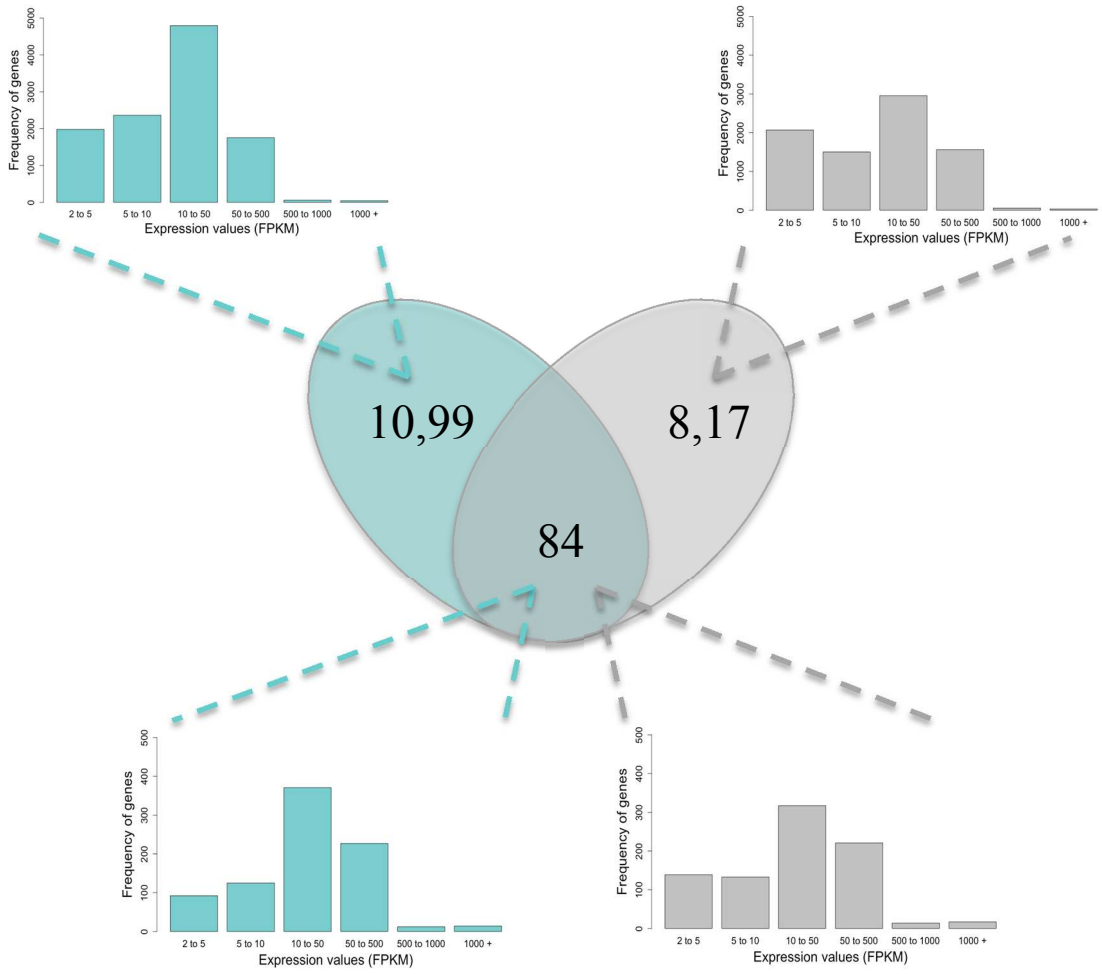


Figure 2.3: Comparative analysis of poly-A transcriptomes in 1-2 cell stage embryos of D. rerio and A. limnaeus. (A) The 20 most abundant transcripts and their FPKM values that are unique to the transcriptome of either *A. limnaeus* or *D. rerio*. (B) Venn diagram depicting the number of shared (orthologous) and non-shared transcripts in 1-2 cell stage embryos of *A. limnaeus* (turquoise) and *D. rerio* (grey). Frequency histograms show the distribution of expression values of shared (orthologous) and non-shared genes and indicate similar patterns of transcript abundance between the two species.

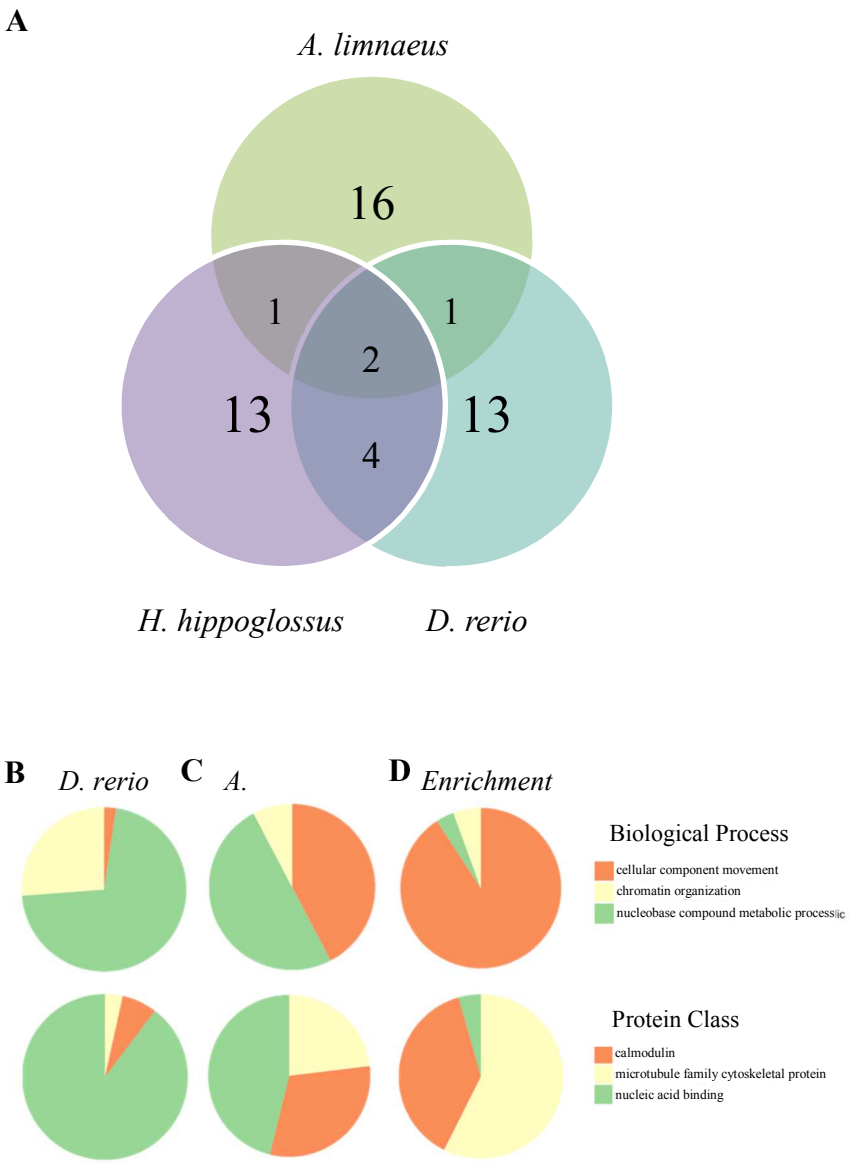


Figure 2.4: The relationship of the maternally packaged transcriptome of A. limnaeus to other teleosts. (A) Venn diagram showing shared numbers of the 20 most abundant transcripts in 3 species of fish with 1-2 cell stage Illumina sequenced transcriptomes*: A. limnaeus* (this study), *D. rerio (Harvey, Sealy et al. 2013)* and *H. hippoglossus (Bai, Solberg et al. 2007)*. (B-D) Gene ontology analysis comparing the top 100 transcripts in the 1-2 cell stage transcriptomes of *A. limnaeus* and *D. rerio*. Pie charts represent the quantities of GO categories in (B) *D. rerio* and (C) *A. limnaeus* for biological process classification (top) and protein class (bottom). (D) Go terms enriched in the *A. limnaeus* compared to *D. rerio* (P <0.05) transcriptome. For more details see the text and Appendix A.7.

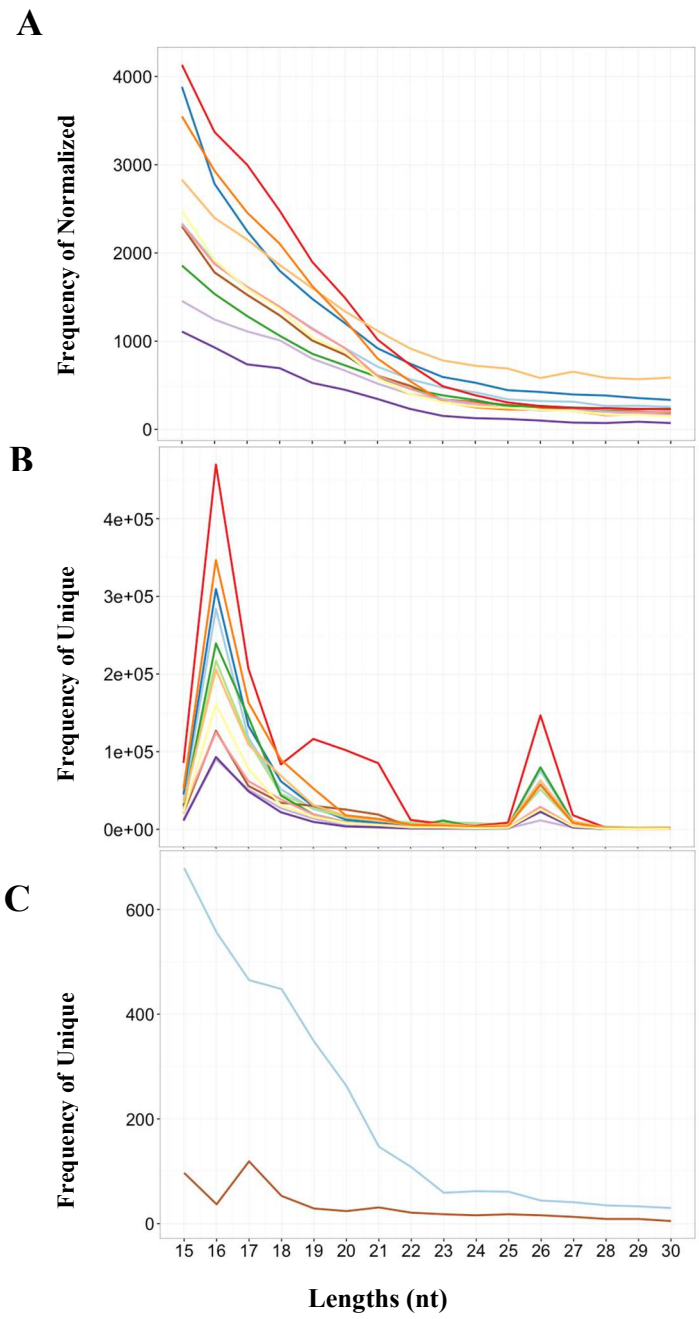


Figure 2.5: Maternally packaged sncRNA transcriptome of A. limnaeus. Frequency distribution of (A) normalized sequence reads and (B) unique sequences as a function of sequence length in the sncRNA libraries (n=12). Each library is a different color. (C) There is a high diversity of sncRNA sequences with lengths between 15 and 23 nucleotides that are unknown (blue line) compared to those that could be annotated (red line) by sequence similarity to those cataloged in RNA databases.

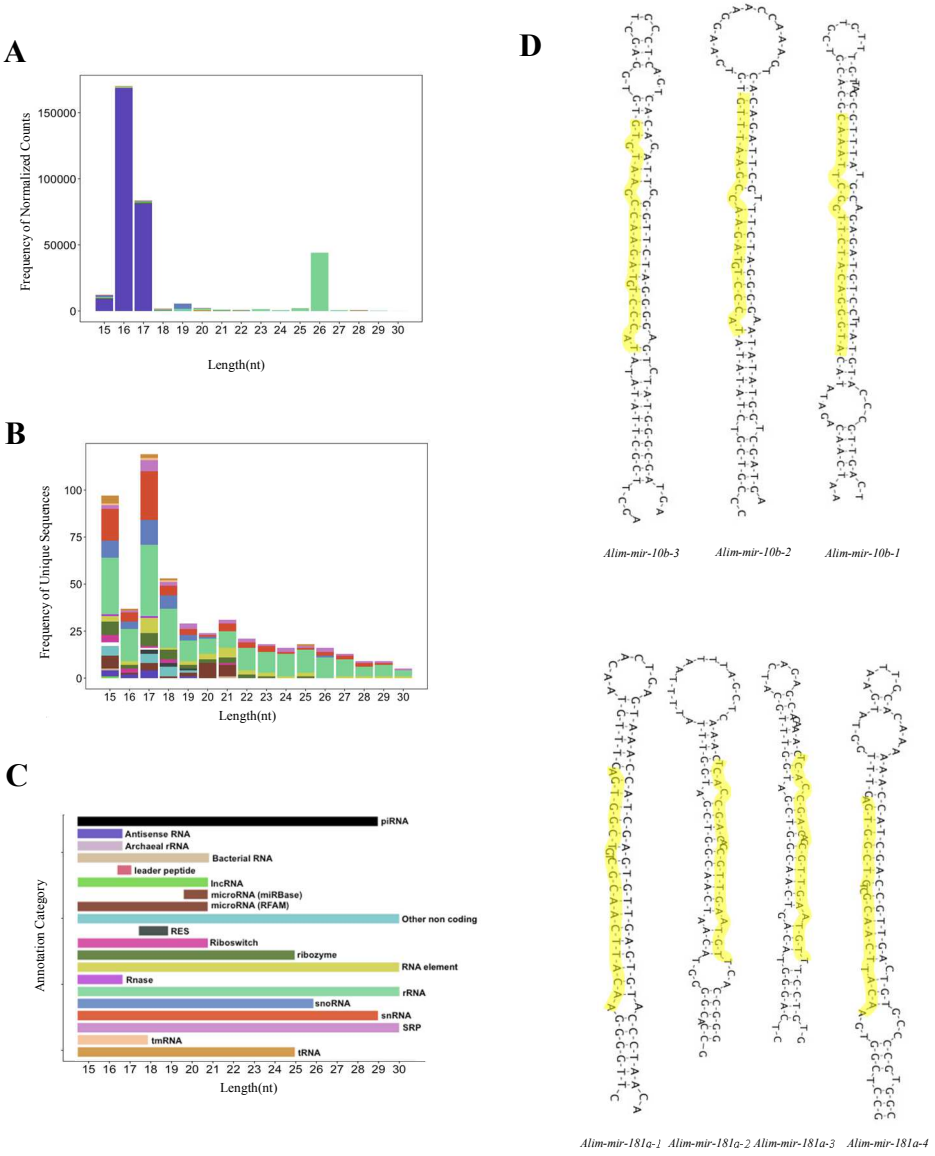


Figure 2.6: Rfam database annotation of the maternally packaged sncRNA transcriptome of *A. limnaeus*. (**A**) The most abundant sncRNAs are 16-17 nucleotides in length and annotate as antisense RNAs, while the second most-abundant group are sequences that are 26 nucleotides in length that annotate as ribosomal RNA (see panel C for a color key to annotation category). (**B**) The highest diversity of unique sncRNA sequences are in the 15 and 17 nucleotide length categories. Note the enrichment of miRNA sequences in the 20-22 nucleotide range as expected, even though miRNAs are not a dominant part of the sncRNA transcriptome. (**C**) Some annotation categories have distinct size ranges, while others span the entire range of sizes explored in this study. (**D**) Putative microRNA precursor structures and consensus mature sequences (highlighted in yellow) annotated as miR-181a (*Alim-miR-181a1-4*) and miR-10b (*Alim-miR-10b1-3*).

Table 2.1 Frequency of poly-A genes at expression bins. Sequences that were detected with a cumulative (across all samples) abundance at >2 FPKM were determined as biologically relevant maternally packed transcripts.

Mean Expression (FPKM)	Cumulative Frequency
>0	20,739 - 100%
>1	13,872 - 67%
>2	12,329 - 59%
>5	9,758 - 47%
>10	7,276 - 35%
>20	4,729 - 23%
>50	2,143 - 10%
>100	1,010 - 5%
>200	453 - 2%
>500	140 - 1%
>1000	68 - 0.33%
>5000	5 - 0.02%

Table 2.2: Top 20 most abundant poly-A RNA transcripts expressed in 1-2 cell stage embryos of *A. limnaeus.*

Gene Symbol	Gene description	Gene Type	FPKM
COI	Cytochrome C oxidase I*	Protein-coding	8,491
16S	16s ribosomal RNA*	rRNA	7,635
LOC106526789	Ferritin, middle subunit	Protein-coding	5,782
LOC106522651	Uncharacterized	lncRNA	5,697
COIII	Cytochrome C oxidase III*	Protein-coding	5,017
LOC106523488	Claudin-like protein, ZF-A89	Protein-coding	4,858
LOC106532313	Uncharacterized	lncRNA	4,619
LOC106518911	Uncharacterized	lncRNA	4,331
LOC106512296	Uncharacterized	lncRNA	4,322
LOC106533547	Late histone H2B.L4-like	Protein-coding	4,219
LOC106530205	Ribosyldihydronicotinamide dehydrogenase [quinone]	pseudogene	3,774
LOC106512721	Uncharacterized	lncRNA	3,752
LOC106511212	Zinc finger protein 36, C3H1 type-like 2	Protein-coding	3,561
LOC106527827	Peptidyl-prolyl cis-trans isomerase-like	Protein-coding	3,539
LOC106519960	Tubulin alpha chain-like	Protein-coding	3,403
ATPase6	ATP synthase 6*	Protein-coding	3,270
ubb	Ubiquitin B	Protein-coding	3,259
LOC106532441	Ribosyldihydronicotinamide dehydrogenase [quinone]-like	Protein coding	3,078
LOC106526526	Tumor-associated calcium signal transducer 2-like	Protein coding	3,042
hspb1	Heat shock 27kDa protein 1	Protein coding	3,011

* Mitochondrially encoded gene

FPKM – Fragments per kilobase of transcript per million mapped reads

Table 2.3: Most differentially expressed orthologous genes between *A. limnaeus* and *D. rerio*.

Gene Name	Percent difference	Species dominantly expressed
mitochondrially encoded ATP Synthase 6	61%	*D. rerio*
claudin-4-like/d	58%	*D. rerio*
mitochondrially encoded cytochrome c oxidase subunit II	45%	*D. rerio*
ubiquitin B/C	30%	*A. limnaeus*
mitochondrially encoded cytochrome c oxidase subunit I	23%	*D. rerio*
mitochondrially encoded NADH dehydrogenase 3	23%	*D. rerio*
calmodulin-like/3a	22%	*A. limnaeus*
mid1-interacting protein 1-like	20%	*A. limnaeus*
mitochondrially encoded cytochrome B	20%	*D. rerio*
translation elongation factor 1 alpha 1	19%	*D. rerio*
mitochondrially encoded NADH dehydrogenase 1	18%	*D. rerio*
mitochondrially encoded NADH dehydrogenase 6	17%	*D. rerio*
heat shock cognate 70 kDa protein/heat shock protein 8 (hspa8), mRNA	12%	*D. rerio*
H3 histone, family 3B/3d	12%	*D. rerio*
mitochondrially encoded NADH dehydrogenase 2	11%	*D. rerio*
mitochondrially encoded cytochrome c oxidase subunit III	10%	*D. rerio*
small ubiquitin-like modifier 3b	10%	*D. rerio*
osteoclast stimulating factor 1	10%	*A. limnaeus*
heterogeneous nuclear ribonucleoprotein A/B	9%	*D. rerio*
ribosomal protein L23	9%	*A. limnaeus*
acidic (leucine-rich) nuclear phosphoprotein 32 family, member B	8%	*D. rerio*

Percent difference – Difference in ortholog rank order abundance in either transcriptome, expressed as percent of total transcripts

Table 2.4: Mitochondrial genes of *A. limnaeus* and *D. rerio* and their expression summary in the 1-2 cell stage transcriptome.

		A. limnaeus		*D. rerio*		*D.r./A.l.*
Gene Code	**Gene Name**	**Rank**	**FPKM**	**Rank**	**FPKM**	**Fold difference**
mt-co1	Cytochrome c oxidase subunit 1	1	8,491	4	7,127	0.8
mt-co3	Cytochrome c oxidase subunit 3	5	5,017	7	6,417	1.3
mt-atp6	ATP synthase subunit a	16	3,270	3	9,270	2.8
mt-co2	Cytochrome c oxidase subunit 2	26	2,436	5	6,798	2.8
mt-nd4	NADH-ubiquinone oxidoreductase chain 4	28	2,309	26	1,916	0.8
mt-nd2	NADH-ubiquinone oxidoreductase chain 2	39	1,738	16	2,924	1.7
mt-nd1	NADH-ubiquinone oxidoreductase chain 1	48	1,419	12	3,208	2.3
mt-cyb	Cytochrome b	63	1,057	14	2,991	2.8
mt-nd4l	NADH-ubiquinone oxidoreductase chain 4L	93	755	126	501	0.7
mt-nd6	NADH-ubiquinone oxidoreductase chain 6	100	714	21	2,373	3.3
mt-nd5	NADH-ubiquinone oxidoreductase chain 5	105	688	47	1,197	1.7
mt-nd3	NADH-ubiquinone oxidoreductase chain 3	106	674	17	2,858	4.2
mt-atp8	ATP synthase protein 8	633	149	89	684	4.6

Rank – rank order of abundance in the transcriptome

FPKM – Fragments per kilobase of transcript per million mapped reads

Fold difference – expression value in *D. rerio* divided by the expression value in *A. limnaeus*

Table 2.5: Top twenty most abundantly expressed small RNA transcripts expressed across all libraries in 1-2 cell stage embryos of *A. limnaeus*.

Sequence	Length[1]	Mean Expression[2]	Rfam ID	Class of RNA
TCAGACAACTCTTAGC	16	168,769	RF02179	antisense RNA
CTCAGACAACTCTTAGC	17	79,972	RF02179	antisense RNA
GTCGCCTGAATACCGCAGCTAGGAAT	26	43,648	not annotated	
GAGCGCCGCGACTCCTCA	18	12,557	not annotated	
GAGCGCTGCGACTCCTCA	18	10,491	not annotated	
TCTCAGACAACTCTTAGC	18	9,808	not annotated	
ACGAGAGCTTTGAAGACCGA	20	7,987	not annotated	
AACGAGAGCTTTGAAGACCGA	21	6,983	not annotated	
CGAGAGCTTTGAAGACCGA	19	6,119	not annotated	
CAGACAACTCTTAGC	15	5,280	RF02179	antisense RNA
GGAGCGCCGCGACTCCTCA	19	5,198	not annotated	
TCAGACAACTCTTAGA	16	4,734	not annotated	
GGTCGCCTGAATACCGCAGCTAGGAAT	27	4,638	not annotated	
GAGCGCCGCGACTCCT	16	3,836	not annotated	
TCAGACAACTCTTAG	15	3,806	RF02179	antisense RNA
ATGTCAAAGTGAAGAAATT	19	3,729	not annotated	
TCACTCTGCATCTCTC	16	2,158	not annotated	
CTCAGACAACTCTTAG	16	1,902	not annotated	
AAACGAGAGCTTTGAAGACCGA	22	1,616	not annotated	
GGAGCGCCGCGACTCCT	17	1,528	not annotated	

1. Length in base pairs

2. Expression in normalized counts per million reads

Table 2.6: Number of genomic alignments for the top 8 most abundant small RNA sequences as a function of the number of bases that must match identically starting from the 5' end.

Sequence	FASTA ID	Length	Mean	5'-15bp	5'-12bp	5'-10bp	5'-8bp
TCAGACAACTCTTAGC	sm_HPF_680	16	168,769	4	77	105	175
CTCAGACAACTCTTAGC	sm_HPF_1236	17	79,972	5	18	27	65
CAGACAACTCTTAGC	sm_HPF_1	15	5,280	4	81	188	338
TCAGACAACTCTTAG	sm_HPF_2	15	3,806	4	128	289	633
TCAGACAACTCTTAGCG	sm_HPF_1239	17	941	4	12	15	27
TCAGACAACTCTCAGC	sm_HPF_769	16	20	30	67	133	221
CTCAGACAACTCTCAGC	sm_HPF_1371	17	9	3	18	34	70
ATCAGACAACTCTTAGC	sm_HPF_1409	17	7	3	12	22	48

Table 2.7: Differentially expressed sncRNAs in diapause- and escape-bound embryos of *A. limnaeus*.

Sequence	Bp[1]	Exp.[2]	Log_2FC[3]	Sig.[4]	Pheno.[5]	15bp[6]	12bp[6]	10bp[6]	8bp[6]	Alignment[7]	Orientation[8]	Position[9]
ACAACGTGTGATACA	15	26.9	5.1	0.008	escape	1	23	103	183	LOC106517361 (Zinc finger protein 646-like)	A	intragenic, intronic
TAGTATATAGGACTA	15	46	4.1	0.036	escape	2	13	60	179	LOC106525548 (neurexin-2-like)/Unannotated region	A/S	intragenic, intronic/ Intergenic
GGCTCTGAATACATTAG	17	5.8	4.5	0.036	escape	2	126	131	159	LOC106520853 (RNA-binding motif, single-stranded-interacting protein 3)/LOC106533203 (endosialin-like)	A/A	intragenic, intronic/ Intragenic, exonic
TCGGAACTCACCCAGTC	17	15.8	4	0.044	diapause	2	4	6	6	Unannotated region/LOC106512419 (un-characterized LOC-106512419)	S/A	intergenic/ Intragenic, exonic

1. Length of the sncRNA in base pairs
2. Mean counts per million mapped reads across all samples
3. Log_2 fold change
4. Statistical significance (FDR < 0.1)
5. Developmental phenotype with the higher abundance of the sncRNA
6. Number of alignments in the *A. limnaeus* genome for the sequence if the number of bases indicated are used starting at the 5' end of the sequence.
7. The locations in the *A. limnaeus* genome where the 15bp sequence aligned.
8. The orientation of each alignment location for the 15bp sequence in reference to the *A. limnaeus* genome annotation: A = antisense; S = sense.
9. Position within the alignment location that the sncRNA aligned to.

Table 2.8: Number of genes designated in the minimal stress proteome (or MSP) identified in the *A. limnaeus* and *D. rerio* transcriptomes at the 1-2 cell stage.

	Number of MSP genes		% of transcriptome	
Abundance Rank*	***A. limnaeus***	***D. rerio***	***A. limnaeus***	***D. rerio***
Top 1% of transcripts	12	8	0.10	0.09
Top 5% of transcripts	20	15	0.16	0.17
Top 25% of transcripts	73	62	0.60	0.68
Top 50% of transcripts	128	116	1.05	1.28
Top 75% of transcripts	163	159	1.33	1.75
Total	188	198	1.54	2.18

*Abundance rank – transcript abundance as a percentage of the total transcriptome

CHAPTER 3:

Small noncoding RNAs along alternative developmental trajectories in *A. limnaeus*

Abstract

Embryonic development of *Austrofundulus limnaeus* can occur along two alternative phenotypic trajectories that are physiologically and biochemically distinct. Phenotype appears to be influenced by maternal provisioning based on the observation that young females produce predominately escape embryos and older females produce mostly diapausing embryos. Additionally, embryonic incubation temperature can override this maternal pattern and alter trajectory. Maternally provisioned gene transcripts were described in the previous chapter, yet no phenotype-specific regulator of trajectory has been discovered. Alternative developmental programs may be regulated by post-transcriptional modification via noncoding RNAs. Using RNA-seq, I have generated transcriptomic profiles of small noncoding RNAs (sncRNAs) in embryos developing along the two alternative trajectories. I hypothesize that the expression of environmentally responsive microRNAs (miRNA) during development can target maternal transcripts and override maternal programming. We find distinct profiles of miRNA families, *miR-10*, expressed in high abundance throughout development and, *miR-430*, that are highly enriched in escape trajectory embryos. *Mir-430* family members are known to target maternally provisioned mRNAs in zebrafish, and may play the same role in *A. limnaeus*. This expression pattern and known function for *miR-430* sequences presents a novel model for maternal-embryonic-conflict in gene regulation that provides the embryo with the ability to override maternal programming in the face of altered

environmental conditions. These results will not only impact our understanding of genetic mechanisms that regulate entrance into diapause, but will also provide insight into epigenetic control of development.

Introduction

The genetic toolkit for vertebrate development provides the evolutionary foundation that supports diverse phenotypic outcomes (West-Eberhard 2005, Carroll, Grenier et al. 2013). However, the contents of this toolkit remain poorly understood. Patterns of gene transcription during early development are highly conserved, but alterations in the timing and especially the level of expression of genes (i.e. control of gene *activity*) during early developmental stages are thought to underlie species-specific morphogenetic processes, while still allowing for "normal" development of key features and structures (Yanai, Peshkin et al. 2011). Conserved gene expression programs can be altered post-transcriptionally through a variety of mechanisms including gene regulatory networks of small RNAs. Regulation of gene expression by small non-coding RNAs (sncRNAs) has emerged as an important driver in the generation of phenotypic complexity (Karp, Hammell et al. 2011, Reynolds, Peyton et al. 2017). Here we explore the potential for sncRNAs to contribute to the developmental program and phenotypic plasticity during development of the annual killifish *A. limnaeus*.

Modification of mRNA structure and stability are especially important steps at which gene expression can be regulated, and can lead to profound variations in sequence and structure of mRNAs from a single genomic region (gene). Understanding the

mechanisms of how RNAs are processed, what factors determine their stability, translatability, and thus their ultimate expression could explain such phenomena as the expression of alternative developmental phenotypes generated from a single genetic background. sncRNAs can regulate chromatin structure as well as induce RNA degradation and translational repression by serving as RNA scaffolds that can target specific nucleotide sequences for alteration by a variety of partner protein complexes (Holoch and Moazed 2015). This unifying mechanism for the action of sncRNAs, and an apparent high degree of evolutionary conservation of function, allows for the identification of potential gene targets of sncRNAs during vertebrate development.

A small number of sncRNAs classes have widely recognized roles in directing gene regulation throughout animal development and include microRNA (miRNA), antisense RNA (asRNA), and piwiRNA (piRNA) (Mattick and Gagen 2001, Wienholds, Kloosterman et al. 2005, Thatcher, Flynt et al. 2007, Bourc'his and Voinnet 2010, Aanes, Winata et al. 2011). Among the various classes, miRNAs are perhaps best known for their role controlling developmental timing of the nematode, *Caenorhabditis elegans* (Lee and Ambros 2001). After being processed from ~80 nt double-stranded precursor RNAs, mature miRNAs maintain a short length of approximately 22 nt. They bind to an Argonaute protein to form the RNA-induced silencing complex (RISC) (Bartel 2004, Bartel 2009, Agarwal, Bell et al. 2015, Mohr and Mott 2015). Acting as sequence specific guides, complimentary base pairing occurs between the miRNA seed (nucleotides 2–7 on the 5' end) and sites within the 3'untranslated regions (3' UTRs) of target RNAs (Bartel 2009). Appropriate docking of the RISC initiates transcript

degradation or translation repression of the targeted mRNA (Bartel 2009, Bazzini, Lee et al. 2012)

In vertebrates, investigations more recently aim to understand the functional complexity within families of miRNAs sequences, which may share a similar core or "seed" sequence but possess a large variety of sequence variants. In zebrafish (*Danio rerio*), members of the *miR-10* family cooperate with closely positioned *Hox* gene clusters that enable a greater precision in anterior-posterior body patterning (Woltering and Durston 2008). The genomic proximity of *miR-10* miRNAs to their target *Hox* genes further supports the critical importance of gene silencing and post-transcriptional regulation in highly evolutionary conserved processes.

Another miRNA family, *miR-430*, has been shown to function in the degradation of maternally packaged transcripts to support a rapid and efficient maternal-to-zygotic transition (MTZ) as the embryonic genome becomes active and assumes control over the developmental process (Giraldez et al., 2006). Maternal mRNAs that are deposited into the egg prior to fertilization are known to control numerous processes during early vertebrate development, including cleavage, blastulation, and formation of the embryonic axis. By definition, the activity of these transcripts must be regulated by post-transcriptional processes such as alterations in translational efficiency or mRNA stability. Members of *miR-430* family are known to target maternal mRNAs and block their expression at the MTZ in order to avoid conflict between maternal and zygotic gene regulation (Giraldez, Mishima et al. 2006, Bazzini, Lee et al. 2012).

The development of *A. limnaeus* and other annual killifishes is unique among vertebrates because two alternative developmental trajectories that differ morphologically, physiologically, and biochemically are known to exist (Podrabsky, Garrett et al. 2010, Pri-Tal, Blue et al. 2011, Woll 2016). One trajectory leads to discontinuous development and the production of embryos that enter into a profound state of metabolic depression known as diapause, while the other trajectory supports continuous development to the pre-hatching stage in embryos that "escape" arrest in diapause. Development along these two trajectories can be controlled by both maternal provisioning (see Chapter 2) and embryonic incubation temperature. An incubation temperature of 30°C induces 100% escape embryos, while 20°C induces 100% diapausing embryos. Intermediate temperatures lead to a mix of phenotypes. The mechanisms by which temperature can alter gene expression and developmental trajectory are currently unknown.

I hypothesize that miRNA expression can regulate developmental trajectory in *A. limnaeus* by altering the stability of maternally packaged transcripts. The first step in evaluating this hypothesis is to identify ontogenetic patterns in miRNA expression along the two developmental trajectories to identify candidate regulators of known phenotypic differences. I propose that environmental signals received by the developing embryo may induce expression of specific miRNAs that target maternally packaged sequences and thus alter developmental trajectory. Importantly, this proposed regulatory mechanism allows for developmental trajectory to be regulated across multiple life stages and integrates both maternal and zygotic information. Using high-throughput RNA

sequencing, I have generated detailed sncRNA transcriptomes for embryos at the same developmental stage that are developing along the two developmental trajectories as controlled by incubation temperature. This global analysis identifies a large contribution by miRNAs to the total sncRNA transcriptome during early development. Specifically, the sncRNA transcriptome is dominated by *miR-430* and *miR-10* family genes. To further test the possible contribution of *miR-430* as a regulator of developmental phenotype, I performed oligonucleotide microinjections of *miR-430* agonists and antagonists. This study documents the diversity of sncRNAs during early development, and suggest a major role for *miR-430* family members in the environmental determination of developmental phenotype in embryos of *A. limnaeus*.

Methods

Experimental design

Details regarding the maintenance of adult fish and collection of embryos can be found in Chapter 2. Embryos were collected and pooled together from 42 pairs of fish during controlled spawning events and maintained for 4 days post fertilization (dpf) in an incubator at 25°C in darkness (Podrabsky 1999). Once reaching 100% epiboly, embryos were transferred to incubation temperatures of 20°C to induce the diapause phenotype and 30°C to induce the escape phenotype (Podrabsky, Garrett et al. 2010). Embryos remained at alternative temperatures until being flash-frozen at seven morphological stages of development that cover the relevant window when the commitment to either the escape or diapause trajectory is determined: cell dispersion (24 hours after temperature

transfer), neural keel, and 6, 10, 16, 20, and 24 pairs of somites (Podrabsky, Riggs et al. 2017) (Figure 3.1). Groups of embryos in specific developmental stages (20-40 embryos per sample) were used for RNA extraction and small RNA-sequencing (n= 6 samples per stage, n = 3 samples per treatment).

Small RNA transcriptomes

The details of RNA extraction, small RNA sequencing, and bioinformatics analysis of sncRNA data can be found in the Methods section of Chapter 2. The details of the samples used and the bioinformatics results of sncRNA profiles are available in APPENDIX B.1 and B.2.

Analysis of miR-430 and miR-10 *family genes*

Abundance patterns of *miR-430* and *miR-10* family sequences were organized and visualized using hierarchical clustering and heat maps of median-centered (by sequence) abundance data expressed as normalized counts. Clustering was accomplished using uncentered Pearson correlation with average linkage in Gene Cluster 3.0 (Eisen, de Hoon et al. 1998, Eisen, Spellman et al. 1998). Heat maps were generated with Java Treeview 1.1.6r4 (Saldanha 2003). Mature sequences that annotated as either *miR-430* or *miR-10* variants by miRBase (Chapter 2) were used to investigate the organization of these genes for these transcripts in the *A. limnaeus* genome. Positions for miRNA genes were identified by the alignment of mature sequences with perfect match to the draft assembly of the NCBI *A. limnaeus* genome (1.0; GenBank accession GCA_001266775.1) with

reference to the genome annotation Release 100 (Wagner, Warren et al. 2015). The majority of individual variants, with only a few exceptions, aligned in clusters of similar miRBase annotations (paralogs, for example: *miR-430b*, *miR-430c*), which aided identifying *A. limnaeus* specific paralogs. Stretches of expanded genomic regions representing putative miRNA genes were evaluated for hairpin secondary structures using the Vienna package RNAfold prediction tool from Geneious (R 8.1.6). These potential miRNA genes were annotated as *A. limnaeus*-specific paralogs from the *miR-430* or *miR-10* family and were used for identifying variant-specific expression profiles in this data set.

Identification of putative gene targets for miR-430 *sequences*

Seed sequences of 6 and 7 nucleotides in length (6- and 7-mer, respectively) on the 5' arm of the *A. limnaeus miR-430* family members were determined as: AAGUGCU and AGUGCU. The complement of these sequences was identified within the 3' UTR of all coding genes in the *A. limnaeus* genome by inferring the differences between exon and gene CDS features annotated by NCBI using a python script available at (ftp://ftp.ncbi.nlm.nih.gov/genomes/TOOLS/add_utrs_to_gff/). Functional annotation of predicted *miR-430* gene targets was determined by generating protein homologs for all gene identities to *Homo sapiens* gene IDs using BLASTp with a threshold e value less than 1.0E-5. Using the Database for Annotation, Visualization and Integrated Discovery (DAVID) bioinformatics software, clusters of *A. limnaeus* genes that were identified by a

H. sapiens homolog gene ID, were tested for GO term enrichment with Bejamani-Hochberg adjusted *P* values < 0.05.

Microinjection of oligonucleotides

A microinjection protocol for the effective introduction of RNA molecules into *A. limnaeus* embryos was developed by adaptation of methods described for other fish species (Xu 1999, Matson, Clark et al. 2008, Hartmann and Englert 2012, Allard, Kamei et al. 2013). Fertilized eggs were embedded into a petri dish of 1% w/v low melting point agarose at 28°C and cooled to its gelling point (24 °C). Embryos were visualized using a dissecting microscope at 20-40X total magnification. Microinjections were performed using an MPPI-3 Pressure injector (Applied Scientific Instrumentation, Inc., Eugene Oregon) attached to a MX130 4-axis micromanipulator (Siskiyou Corporation, Grants Pass, OR). Embryos at the 1-2 cell stage were either injected with 1-2 pulses into the yolk mass or directly into the blastomeres with *miR-430* solutions (see below). Afterwards, embryos were briefly stored in the dark for 1-3 hours before being removed from the agarose and transferred to incubation media and maintained at 25 or 30°C for further observations depending on the experiment.

Microinjection needles were prepared using borosilicate glass capillaries with an internal filament (1.0 mm O.D. and 0.58 mm I.D.). Capillaries were pulled into micropipette needles using a Sutter Micropipette Puller, Model P-80/PC (California, USA). Similar to previous reports for other killifish species, the needle profiles that are best for penetrating *A. limnaeus* chorions have a shorter and more robust tip than that of a

typical zebrafish microinjection needles (Hartmann and Englert 2012). The program used has two loops with the following parameters: Heat: 700, Pull: 50, Velocity: 20, Time: 44.

A morpholino oligonucleotide sequence for miRNA expression inhibition (*miR-430a* knockdown) was designed and manufactured by Gene Tools (Philomath, OR) in the reverse complement for *alim-miR-430a*: 5'-ACTACCCCAACAATCAGCACTTACT-3', and labeled with a 3'-carboxyfluorescein. The standard control oligonucleotide sequence offered by Gene Tools was used as an injection control: 5'CCTCTTACCTCAGTTACAATTTATA-3', and also labeled with fluorescein. An RNA mimic to induce enhanced miRNA activity (*miR-430a* mimic) was manufactured by Integrated DNA Technologies as an RNA duplex oligonucleotide with one arm identical to the mature seq of *alim-miR-430a*: 5'-AGUAAGUGCUGAUUGUUGGGGUAG-3' (active anti-sense strand) and the arm as 3'-TCAUUCACGACUAACAACCCCA-5' labeled with a 5' 6-FAM™ fluorophore. All oligonucleotides were diluted to 500 nM and 750 nM injection stocks in RNase-free water.

Results

Small RNA transcriptome

After processing the small RNA sequence reads, including trimming and genome annotation, our methods detected approximately 350,000 to 600,000 unique sequences across developmental stages with ≥ 2 normalized counts (Figure 3.2A). The length distribution of unique small RNAs identified two highly abundant class sizes with peaks

at 22 and 17 nucleotides (nt) (Figure 3.2B). These transcriptomes were annotated to databases of known small RNAs and further categorized as ribosomal RNA (rRNA), miRNA, transfer RNA (tRNA), or small nuclear RNA (snRNA) among many others (Figure 3.2C). Across developmental stages, miRNAs represented 13-20% of annotated sequences. As a major contributing RNA class to post-transcriptional regulation of mRNA, we analyzed these sequences in further detail. The most abundant miRNA annotations (normalized counts) across all developmental stages were *miR-430* and *miR-10* variants (Figure 3.2D). Other highly abundant miRNAs include *miR-92*, and *miR-181*.

A. limnaeus miR-10 *family genes*

A total of 144 sequences were annotated by miRBase as variants of *miR-10*, a miRNA family present at fertilization (Chapter 2) and the most abundantly expressed family detected in the present study. The abundance of these variants increases substantially during somitogenesis in *A. limnaeus*, contributing to over 50% of annotated reads in embryos with 16 pairs of somites and greater (Figure 3.2D). To investigate if the expression of *miR-10* variants is specific to phenotype, we analyzed their relative expression in escape- (30°C) and diapause-bound (20°C) embryos (Figure 3.3A). Interestingly, most variants appear to increase in abundance earlier in escape-embryos (10 pairs of somites) than in diapause-trajectory embryos (16 pairs of somites). However, only a handful of *miR-10* family variants are differentially expressed between the two developmental trajectories (Figure 3.3B).

Computational analysis of the genomic positions for *miR-10* family genes identified 4 of the 5 already established variants of *miR-10*: *b-1*, *b-2*, *c*, and *d*. The variants aligned in high abundance to 4 scaffolds of the *A. limnaeus* genome, NW_013952411.1, NW_013952405.1, NW_013952721.1, and NW_013952574.1, in patterns that demonstrate detection of 5' and 3' arms of the precursor sequences. There was no evidence of *miR-10a* detected in this dataset. Secondary hairpin structures were successfully predicted for miRNA precursors for these 4 paralogs further demonstrating their strong candidacy as miRNA genes (Figure 3.4A). Comparable to what is known for other teleosts, *A. limnaeus miR-10* genes are positioned within *Hox* family gene clusters (Figure 3.4B). Consensus sequences of the 4 paralogs identified were highly similar and thus it is impossible to distinguish the gene origin between *alim-miR-10b-1*, *alim-miR-10b-2*, or *alim-miR-10d* mature sequences in our data set. Separate profiles of expression could be characterized for the group *alim-miR10-b/d* and for *alim-miR-10c* (Figure 3.4C). Both groups show an increase in abundance throughout development with greater expression in escape embryos. While the sum of transcripts belonging to *miR-10b/d* showed a marked decrease along the escape trajectory at 20 pairs of somites, none of the individual variants were differentially expressed (*t*-test, FDR > 10%).

A. limnaeus miR-430 *family genes*

To investigate the role of the 85 sequences that annotated as *miR-430* family members, expression profiles for all variants were compared in embryos developing along the escape and diapause trajectories (Figure 3.5A). A majority of *miR-430* mature

sequences were highly abundant throughout development in the escape trajectory with significant differential expression between the two trajectories (Figure 3.5B).

Analysis of the genomic positions for *miR-430* family genes identified 5 possible variants that mapped to two large scaffolds, NW_013952456.1 and NW_013952881.1, in the *A. limnaeus* genome assembly (Figure 3.6A). All variants were investigated for the ability to form hairpin secondary structures (Figure 3.6B) and named as paralogs *a-e* of the *alim-miR-430* gene family. The consensus sequences of the 5 paralogs share a 5'- and 3'- end sequence but exhibit a unique 4-mer in the central part of the sequence (Figure 3.6A). Each paralog can be further identified by isoforms within the paralog family. Mature sequences of the 5 *miR-430* paralogs of *A. limnaeus* were evaluated for variant-specific expression profiles (Figure 3.6C). The sequences belonging to *alim-miR-430a* are highly abundant in comparison to other family members by 1-2 orders of magnitude in most cases. This variant, along with *alim-miR-430b*, *alim-miR-430d* had a greater abundance at stages of cell dispersion, neural keel, and 6 pairs of somites in embryos developing on the escape trajectory in comparison to the diapause trajectory. In contrast, expression of *alim-miR-430c* was similar in the two trajectories, while variants of *alim-miR-430e* demonstrated a strong diapause-specific expression pattern. All paralogs decline in abundance during early development and all but the *miR-430e* family members reach roughly equivalent abundance in embryos with 20 pairs of somites or more.

Potential gene targets were identified in the *A. limnaeus* genome with complementary binding sites within the 3'UTR for the 6- and 7-mer *miR-430* seed sequences (Table 3.1A). These genes were enriched with molecular functions of protein

interactions including binding, signaling, and regulatory roles (Benjamini-Hochberg adjusted *p* values < 0.05; Table 3.1B). Gene targets for *miR-430* often display multiple binding sites (Bazzini, Lee et al. 2012), and thus we used the 6-mer binding site data to be as inclusive as possible. Interestingly, the transcript with the most binding sites comes from the gene *argonaute* (Table 3.1C). Other genes of interest with a high number of binding sites include the UV radiation resistance associated gene (*uvrag*) and nuclear receptors such as retinoic-acid-receptor (RAR) related orphan receptor A (*rora*) and nuclear receptor ROR-beta-like, (LOC106513868). The resulting gene targets for *miR-430* represent potential key regulators of normal development, and trajectory-specific regulation of development that will require further functional analysis (see Supplemental File 3.1 for the full list).

Microinjection of oligonucleotides

The RNA microinjection protocol resulted in optimal survival rates at the 1-2 cell stage based on multiple rounds of microinjections with a *miR-430a* mimic, a *miR-430a* knockdown oligonucleotide, and a scramble oligonucleotide. Delivery and persistence of the RNA was confirmed via fluorescence microscopy. Injections delivered directly into the first blastomere appeared to be the best way to achieve positively labeled embryonic cells and tissues.

To examine if over-expression of *miR-430a* would induce development strictly along the escape trajectory under diapause-inducing laboratory conditions (25°C), we injected double stranded RNA (500nM) with one arm matching the *miR-430a* sequence

either into the yolk or directly into the blastomeres (Figure 3.7). *MiR-430a* mimic microinjected embryos experienced a slight lengthening of the dispersion and reaggregation phase of development compared to control embryos and embryos mostly developed along the escape trajectory at 25°C.

We also injected RNA molecules with a sequence in the reverse compliment of *miR-430a* to induce the under-expression of *miR-430a* and stimulate embryos to develop along the diapause trajectory under escape-inducing conditions (30°C). Knockdown microinjected embryos showed either a high a degree of mutation or normal development along the escape trajectory at 30°C (Figure 3.8).

Discussion

This study is the first to characterize the sncRNA transcriptome associated with development along the two alternative developmental trajectories possible in embryos of annual killifish. Consistent with previous work in this species, we have identified miRNAs as major contributors to the sncRNA transcriptome during development with the most abundant sequences annotating to conserved families of vertebrate miRNAs including *miR-10*, *miR-430*, *miR-92*, and *miR-181* (Riggs and Podrabsky 2017, (Romney and Podrabsky 2017). Many classes of miRNA have been implicated in the regulation of gene expression in a variety of contexts during embryonic development (Lee and Ambros 2001, Hannon 2002, Wienholds, Kloosterman et al. 2005, Zhao and Srivastava 2007, Bartel 2009, Barrey, Saint-Auret et al. 2011). Two of these families, *miR-10* and *miR-430*

are highly abundant and differentially expressed during early development in *A. limnaeus*.

The miRNA family, *miR-10*, has been extensively investigated and characterized by its co-evolution with, and genomic proximity to, *Hox* family clusters (Tanzer, Amemiya et al. 2005, Wolter, Le et al. 2017). *Hox* genes code for highly conserved transcription factors that are crucial for anterior-posterior patterning during vertebrate development. Additionally, mounting evidence has demonstrated both the co-expression of *miR-10* and *Hox* genes during development and the targeting of HOX transcripts by *miR-10* family members (Woltering and Durston 2008, Jiajie, Yanzhou et al. 2017). In *A. limnaeus* embryos there is a sharp increase in *miR-10* variant expression at the 10 somite stage in escape embryos incubated at 30°C. A similar sharp increase is observed in the diapause-bound (20°C) embryos at the 16 somite stage. These expression patterns are consistent with a role for *miR-10* family members in helping to establish the anterior and posterior axis and develop appropriate structures during this time. The earlier expression of *miR-10* variants in the escape trajectory embryos is an interesting pattern that warrants further investigation.

Despite these trends in expression, very few of the *miR-10* variants were differentially expressed in the two developmental trajectories. Of note are a few variants that are highly upregulated in escape embryos starting at the neural keel stage and highly expressed through the 16 somite stage. These same variants are expressed at much lower levels in diapause-bound embryos during this window of development. Another interesting feature of *miR-10* variants is their known ability to *increase* the translation of

proteins involved in the support of protein synthesis, especially ribosomal proteins (Lund 2010). This role is rather unique among miRNAs which typically block translation of proteins. Protein synthesis is known to be depressed in embryos developing along the diapause trajectory (Podrabsky and Hand 2000), and perhaps the reduced expression of these *miR-1*0 variants could help to explain the limited ability of these embryos to support high levels of protein synthesis.

The analysis of multiple *miR-10* family members in this study has led to a much more detailed understanding of the *miR-10* family in *A. limnaeus*. This increased knowledge has led to the need to reclassify a previously identified *miR-10-b-3* as *miR-10c* (Chapter 2).

Members of the *miR-430* family are known to be expressed at the MTZ which occurs in most vertebrates prior to axis formation. In *A. limnaeus* the MTZ occurs around the initiation of epiboly during a time known as the midblastula transition (Podrabsky, Riggs et al. 2017). Interestingly, *miR-430* expression is still high in dispersed cell stage embryos several days after the midblastula transition has occurred. Further, many *miR-430* family variants are highly expressed during early somitogenesis in the escape-bound embryos. This continued high abundance and differential expression between the two developmental trajectories is unique compared to other models of fish development.

Developmental trajectory can be influenced by both maternal cues and embryonic incubation temperature (Podrabsky, Garrett et al. 2010). Thus, we investigated if temperature-induced differences in *miR-430* variant abundance could target maternal transcripts that may be critical for programming embryos to develop along the diapause

trajectory. A number of trajectory-specific mRNA splice variants were identified that appear to be maternally packaged (Chapter 2). Two of these variants contain predicted *miR-430* targets sites: sideroflexin-5 like (LOC106524277) and phosphorylase kinase, beta (*phkb*; Romney and Podrabsky 2017). Sideroflexin-5 belongs to a family of mitochondrial tricarboxylate carrier proteins and is known to contribute to brain development in *Xenopus laevis* (Li, Han et al. 2010). This is the first report to our knowledge of any member of the sideroflexin family being a *miR-430* target. The exact function of sideroflexin-5 is unknown, but it has been shown to transport citrate under *in vitro* conditions. Diapause-bound embryos are known to be poised for anaerobic metabolism based on high ratios of lactate dehydrogenase activity compared with citrate synthase activity (Chennault and Podrabsky 2010). A role for a diapause-specific citrate transporter is an intriguing possibility that deserves future attention. Another splice variant that may be targeted by *miR-430*, *phkb*, is highly upregulated in diapause embryos at fertilization (Romney and Podrabsky 2017). This particular gene is a known contributor to regulation of glucose mobilization from glycogen that can be regulated by insulin and insulin-like signaling pathways (Ørngreen, Schelhaas et al. 2008). Reduced insulin-like signaling has been established as an important contributor to the diapause trajectory in *A. limnaeus* (Woll and Podrabsky 2017), and perhaps a diapause-specific variant of *phkb* is required to support the unique metabolic poise of diapausing embryos that would be incompatible with the high metabolic demands of the escape trajectory. In light of these findings, we propose that key maternal mRNA transcripts inherited by

diapause-bound embryos could be targeted by zygotic *miR-430* in response to environmental cues to override maternal signals.

While predicting targets of miRNAs by sequence homology is imprecise and known to be somewhat unreliable, it is still the most common way to interrogate the possible effects of a miRNA at the genomic level. The fact that miRNAs are highly conserved in sequence and function across vertebrate taxa allows for the generation of hypotheses for functions of *miR-430* in *A. limnaeus*. Gene targets of *miR-430* in other species have been predicted and are important to developmental regulation including members of the Wnt and TGFβ/Nodal signaling pathways (Schier and Shen 2000, Chen, Manninga et al. 2005, Giraldez, Mishima et al. 2006). Interestingly, the TGF-β pathway is a known regulator of dormancy in *C. elegans* dauer larvae (Fielenbach and Antebi 2008), and contains a *miR-430* binding site in *A. limnaeus*. In addition, many genes in *A. limnaeus* (2,272, Table 3.1A) have predicted binding sites for *miR-430*. An interesting population of transcription factors occupies the list of genes with the most binding sites (Table 3.1C). In addition, other genes that are important to developmental timing were discovered including a member of the POU family of transcription factors (LOC106534879), period circadian clock 1 (*per1*), multiple members of forkhead box genes, and multiple members of cell cycle regulating cyclin genes (Supplemental File 3.1).

The systemic impact of miRNAs is commonly investigated by either the removal of the miRNA processing enzyme, *dicer*, or by the targeted silencing of mature miRNAs (Giraldez, Cinalli et al. 2005, Harfe, McManus et al. 2005). It has been shown that

mutant embryos lacking functional *dicer* display morphological defects during gastrulation and brain development. The reintroduction of *miR-430* rescues these defects, suggesting that a loss of *miR-430* alone is responsible for these phenotypic outcomes (Giraldez, Cinalli et al. 2005). Therefore, we approached our functional analysis of *miR-430* by enhancing or decreasing the expression of one of the most abundant *miR-430* variants, *alim-miR-430a*, in embryos of *A. limnaeus*. Consistent with previous work in other species (Giraldez, Cinalli et al. 2005), knockdown of *miR-430* led to interruption of normal development in *A. limnaeus* and a large number of abnormal embryos when compared to control embryos (Figure 3.8). However, we observed no alteration in developmental trajectory associated with the knockdown experiments.

Microinjections of a *miR-430* mimic, intending to enhance the abundance of *alim-miR-430a*, demonstrated curious results at stages around epiboly, where positive fluorescence signal was largely purged from the labeled blastomeres and either degraded or somehow transferred to the yolk (Figure 3.6). This was an unexpected result that implies these embryos have a mechanism for purging foreign RNA from cells, or purging cells that contain foreign RNA. Importantly, the loss of signal is associated with the initiation of epiboly, a time where the future embryonic cells lose their attachments to other cells and migrate across the yolk eventually leading to a total dispersion of the embryo-forming blastomeres in the dispersed cell phase of development. The occurrence of the dispersed cell phase of development in annual killifishes has been suggested as a stage that can buffer development against cellular damage in order to support survival in stressful environments (Wourms 1972, Wagner and Podrabsky 2015). This curious and

unexpected result warrants future studies. Based on these experiments, disruption of miRNA expression may have to rely on transgenic or transient expression methodologies to overcome the robust ability of these embryos to purge foreign RNA from the embryo (Hartmann and Englert 2012, Pereiro, Loosli et al. 2017).

In summary, this study suggests a role for temperature-induced miRNA regulation of developmental phenotype in embryos of *A. limnaeus*. While many additional studies are needed to fully support this hypothesis, the expression patterns and potential targets of *miR-430* family members described in this study strongly suggest a role for this miRNA in the integration of environmental signals to alter developmental trajectory. The embryonic alteration or regulation of maternally provisioned genes could provide this species with a strategy to override inherited phenotypic cues in an environment that often fluctuates between tolerable and intolerable conditions. These data suggest a tangible molecular mechanism for maternal-zygotic conflict in life history decisions that could have far-reaching implications beyond the annual killifish model. Further investigations of the role of miRNAs in regulating alternative trajectories may shed light on the role of maternal packaging in the regulation of phenotype, and lead to a better of understanding of the role of epigenetic gene regulation in vertebrate development.

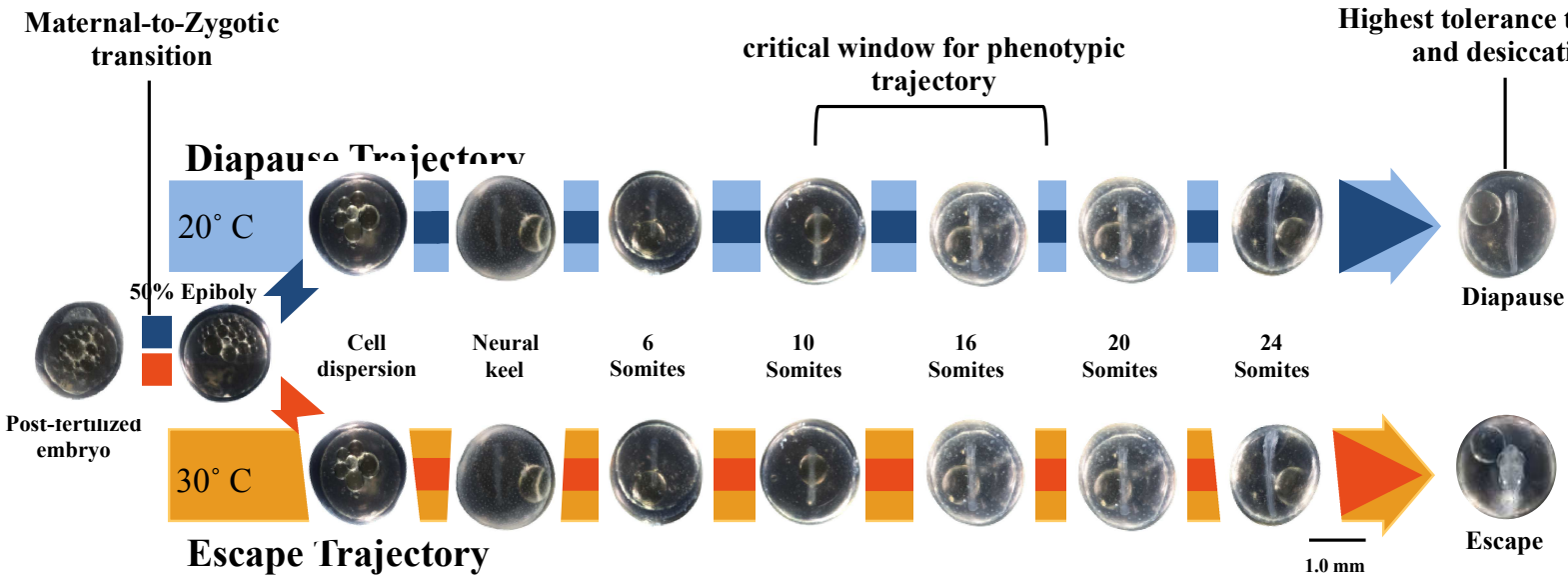


Figure 3.1: Experimental design for temperature-induced developmental phenotypes. Stages along diapause and escape trajectories at which embryos were sampled for sequencing. Groups of related individuals were incubated at 25˚C until 4 days post-fertilization (100% epiboly), when they were transferred to either 20˚C or 30˚C.

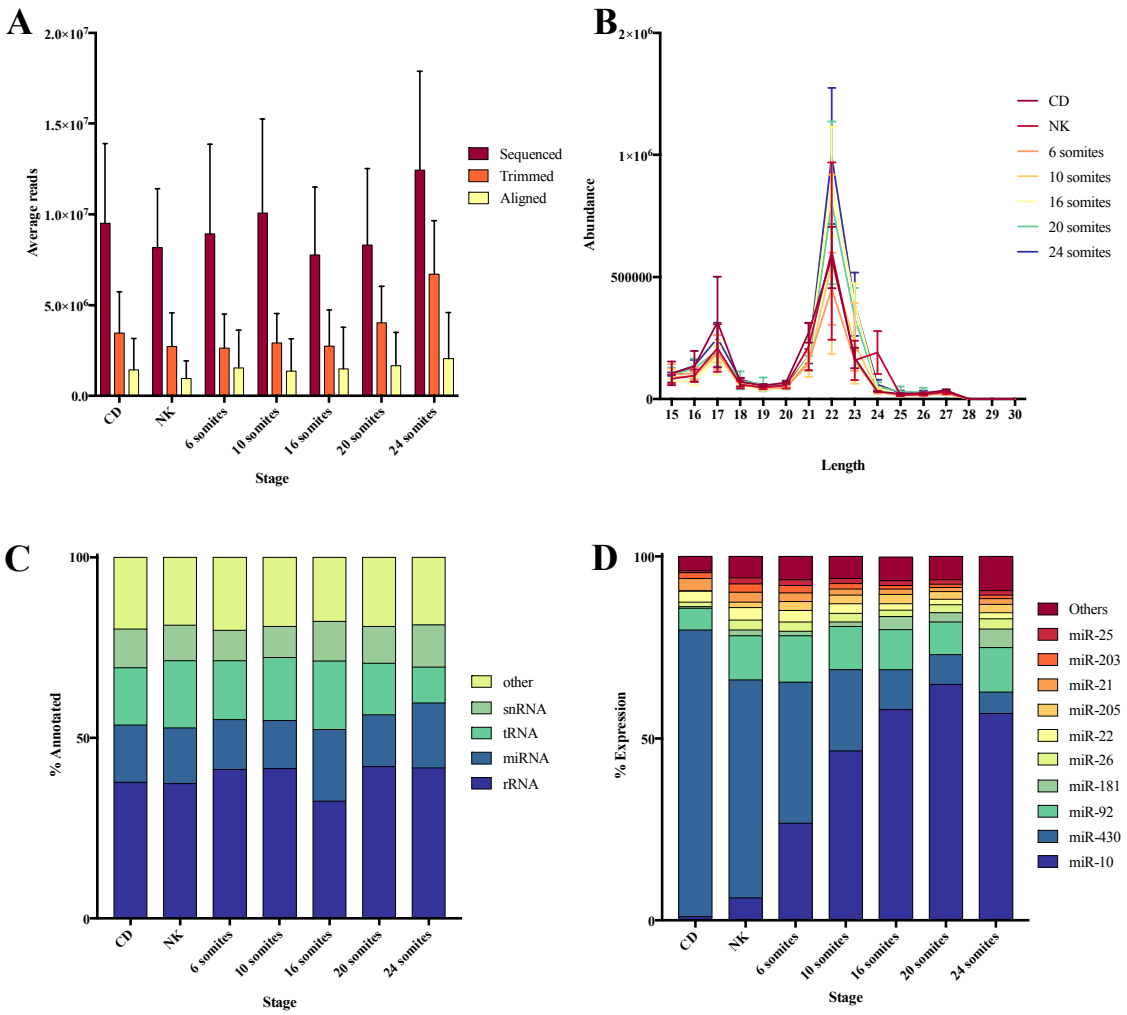


Figure 3.2: Small RNA transcriptome characterization. (A) Average frequency of small RNA sequencing, trimming, and annotation from all temperature-induced libraries. (B) Length distribution of the mean frequency of small RNA sequences from all treatments groups (error bars are standard error of the mean; SEM). (C) Category of small RNAs annotated by miRBase and RFAM across developmental stages. Values are represented as percent of total RNAs annotated. (D) Percent of total expression of the ten most abundant small RNAs annotated as miRNA by miRBase across developmental stages.

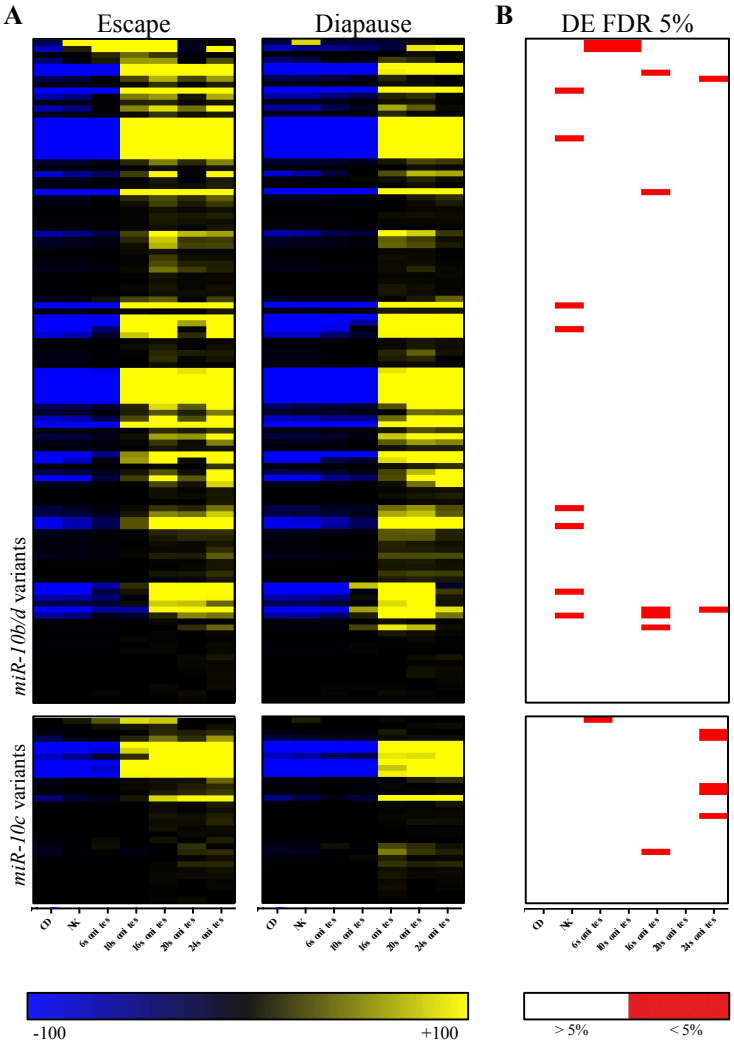


Figure 3.3: ***miR-10* family miRNA variants in *A. limnaeus*** (A) Heat maps of relative fold changes (median-centered) of normalized expression values across development and phenotypic trajectory of all miRNAs that annotated as a *miR-10* variant by miRBase. Each row represents genes that are up-regulated genes in yellow and down-regulated genes in blue across developmental stages. Upper panels are variants belonging to *miR10b/d* and lower panels are variants of *miR-10c*. (B) Heat map shows significant differential expression (DE) between phenotypes for each gene and stage of development *p* values < FDR 5% (red).

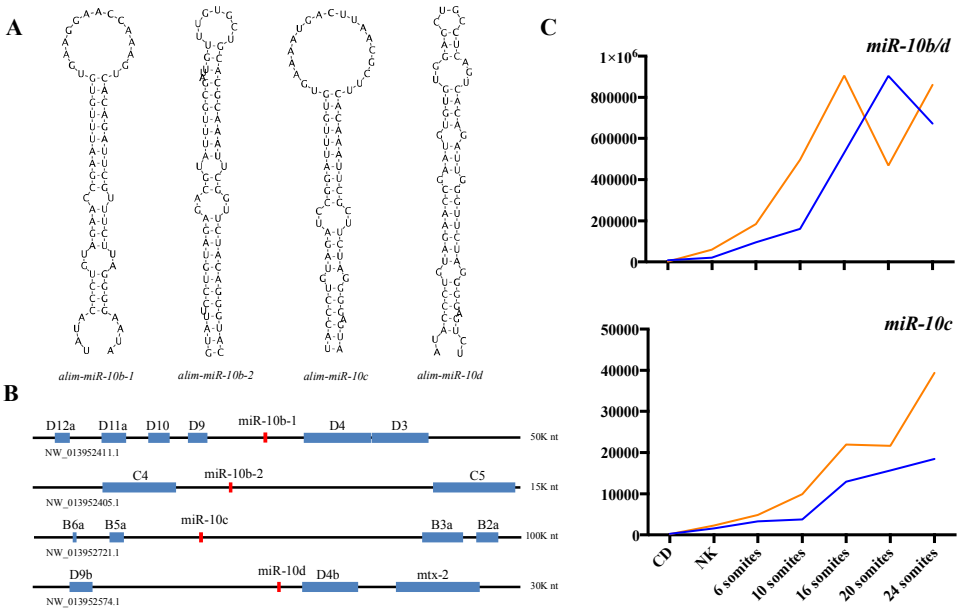


Figure 3.4: ***miR-10* family miRNA genes in *A. limnaeus*.** (A) Genomic organization of *Hox* gene cluster and position of *miR-10* family members. Horizontal black lines represent individual scaffolds, positions of *miR-10* genes are in red and *Hox* genes are in blue. (B) Transcriptomic profiles (normalized counts) of *miR-10* variants determined by paralog (orange is escape trajectory; blue is diapause trajectory). (C) Predicted hairpin structures of precursor *miR-10* paralogs.

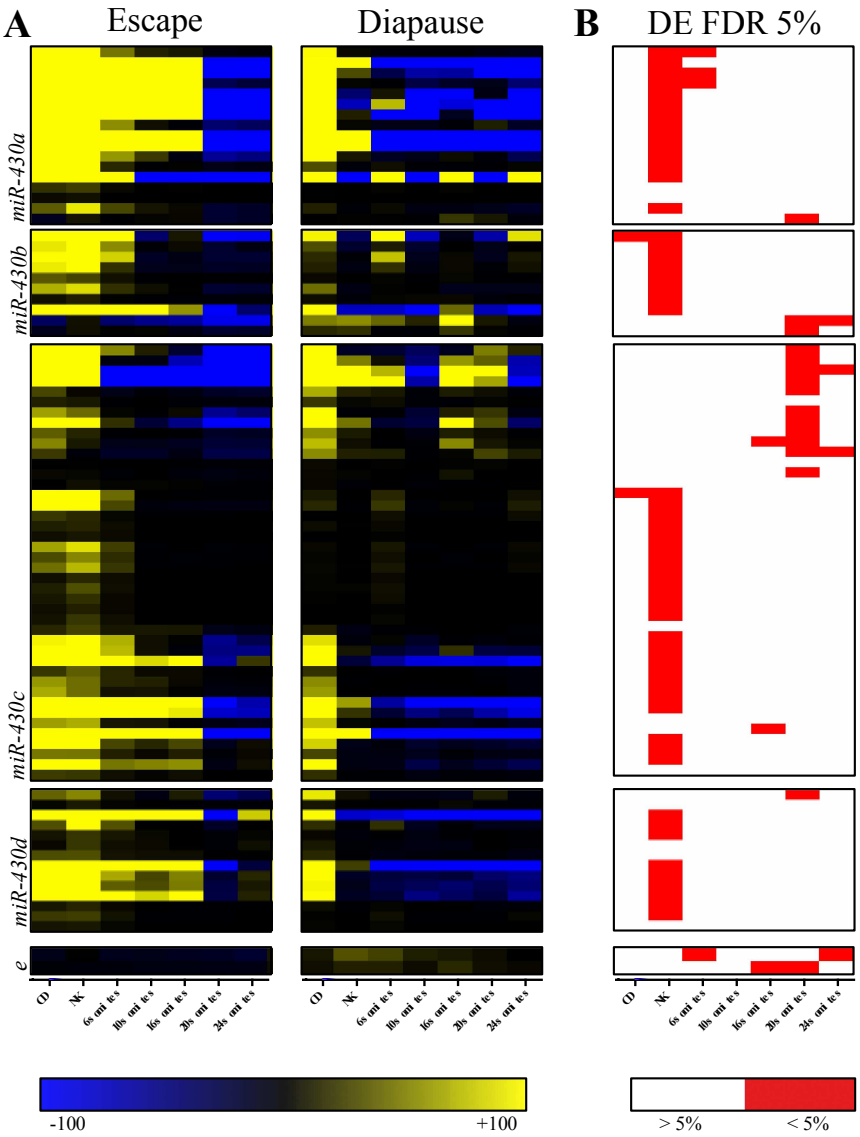


Figure 3.5: ***miR-430* family miRNA variants in *A. limnaeus*** (A) Heat maps of relative fold changes of normalized expression values across development and phenotypic trajectory of all miRNAs that annotated as a *miR-430* variant by miRBase. Each row represents genes that are up-regulated genes in yellow and down-regulated genes in blue across developmental stages. Panels starting at the top are variants *miR-430-a*, *-b*, *-c*, *-d*, and *-e*. (B) Heat map shows significant differential expression (DE) between phenotypes for each gene and stage of development *p* values < FDR 5% (red).

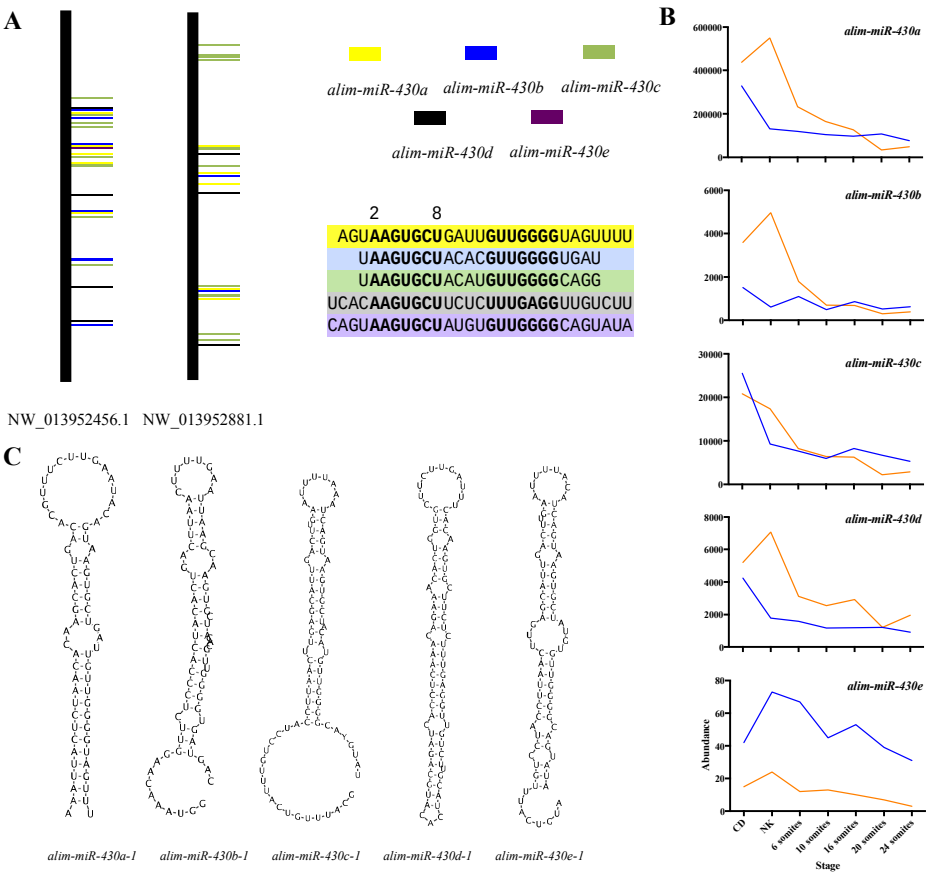


Figure 3.6: ***miR-430* family miRNA genes in *A. limnaeus.*** (A) Genomic organization of the two large clusters of *miR-430* genes and position of *miR-430* family members. Vertical black lines represent individual scaffolds; the positions of *miR-430* genes are color coded and represented by horizontal bars. (B) Transcriptomic profiles of *miR-430* variants determined by paralog (orange is escape trajectory; blue is diapause trajectory). (C) Predicted hairpin structures of precursor *miR-430* paralogs.

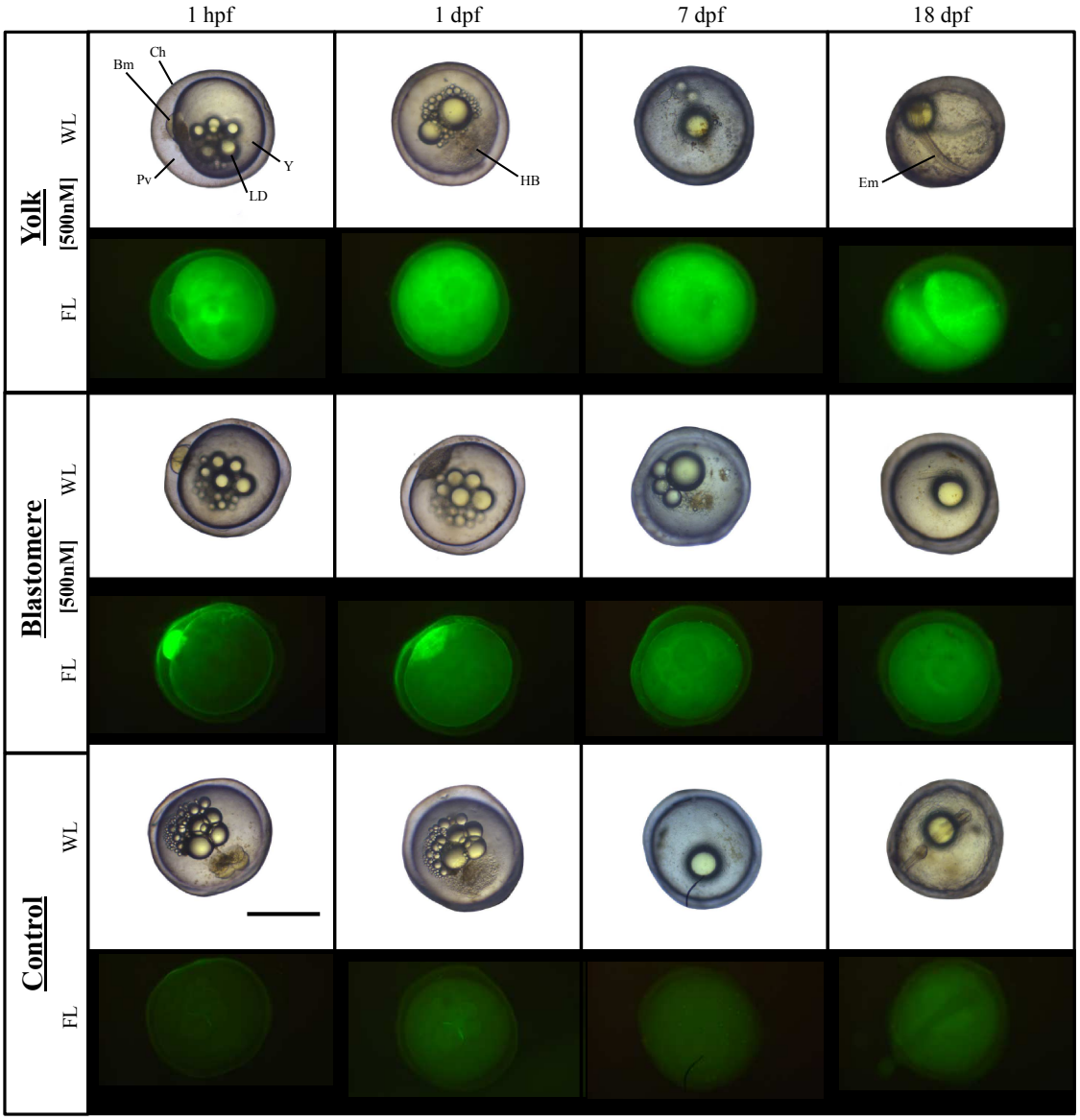


Figure 3.7: miRNA mimic oligonucleotide microinjections. At 1 hour-post fertilization (1 hpf), 1-2 cell stage embryos of *A. limnaeus* for microinjection experiments were separated into three groups depending on treatment: delivery into yolk (Yolk), delivery into the first blastomere (Blastomere), or not injected (Control). Embryos seen here were injected with 500 nM double stranded RNA (*miR-430a* mimic). Embryos were observed under white light (WL) and fluorescent light (FL) throughout development at the Hollow blastula (1 day postfertilization; dpf), reaggregation (7 dpf) and mid-somitogenesis (18 dpf). Scale bar = 1 mm. Ch, chorion, Bm, Blastomere, HB, High Blastula, Em, Embryo, LD, lipid droplets, Pv, Perivitelline space, Y, Yolk

A

750 nM Injections/ 4dpf to 30°C		Diapause-like	Escape-like
	No development	**Delayed development**	**Accelerated/Normal development**
Population- 7dpf	No development	NK	3-6 som
Injected With TKO	12	8	7
Attempted TKO		2	2
Control - Embedded	5	11	8
Control Injected with Scramble	1	1	6
Control - Attempted injection w scramble			1
Population- 13dpf	No development	20-28 som	29+ som
Injected With TKO	4	8, 1*	14
Attempted TKO		1	3
Control - Embedded		3, 1*	21
Control Injected with Scramble			8
Control - Attempted injection w scramble			1

* twinning, mutant

B

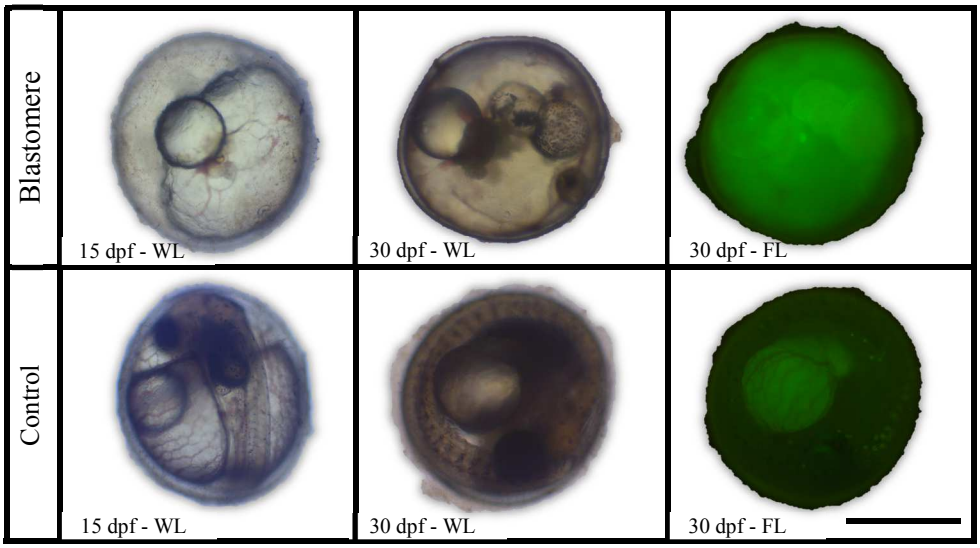


Figure 3.8: miRNA knockdown oligonucleotide microinjections. At 1 hour-post fertilization (1 hpf), 1-2 cell stage embryos of *A. limnaeus* for microinjection experiments separated into three groups depending on treatment: delivery of reverse compliment oligonucleotide for *miR-430* into the first blastomere (Blastomere), or not injected (Control), or injected with a control sequence (scramble). Embryos with injections attempted but not confirmed for delivery were categorized separately. Embryos were transferred to 30°C after injections and observed throughout development (7 and 13 dpf). At observation, populations of embryos were categorized into three age groups with the most accelerated following escape-like traits and less accelerated showing diapause-like traits. (B) Representative embryos showing high mutations (upper panel) and control growth patterns (lower panel). Embryos were observed under white light (WL) and fluorescent light (FL) throughout development. Scale bar = 1 mm.

Table 3.1: Summary of potential *miR-430* gene targets. (A) Number of *A. limnaeus* genes with binding sites for *miR-430* seeds (6mer and 7mer) within the 3' UTR. (B) Functional annotation of potential targets of *miR-430* with molecular functions statistically enriched and having BH adjusted *p* values < 0.05. (C) Genes in *A. limnaeus* genome with the most binding sites

A

Genes with *miR-430* binding sites		sites/gene				
		5	**4**	**3**	**2**	**1**
7-mer seed	822	-	-	4	50	768
6-mer seed	2,272	1	7	54	281	1,929

B

Molecular Function Gene Ontology	**BH *p* value**
protein binding	9.10E-18
protein serine/threonine kinase activity	3.80E-08
magnesium ion binding	6.20E-06
protein kinase activity	7.20E-05
cadherin binding involved in cell-cell adhesion	8.30E-05
receptor signaling protein serine/threonine kinase activity	6.30E-04
ATP binding	1.60E-03
phosphoprotein phosphatase activity	2.10E-03
protein kinase binding	3.20E-03
Rab GTPase binding	3.60E-03
GDP binding	1.10E-02
ubiquitin-protein transferase activity	2.10E-02
protein tyrosine phosphatase activity	4.20E-02

C

Gene symbol	**Description**	**7-mer sites**	**6-mer sites**
LOC106532230	protein argonaute-1	2	5
arid4b	AT rich interactive domain 4B (RBP1-like)	1	4
btg1	B-cell translocation gene 1, anti-proliferative	0	4
LOC106519102	excitatory amino acid transporter 2-like	1	4
pten	phosphatase and tensin homolog	2	4
rora	RAR-related orphan receptor A	3	4
ugt8	UDP glycosyltransferase 8	0	4
uvrag	UV radiation resistance associated	0	4
ano8	anoctamin 8	0	3
atxn1	ataxin 1	2	3
bace1	beta-site APP-cleaving enzyme 1	0	3
fbxo33	F-box protein 33	1	3
fign	fidgetin	1	3
fstl4	follistatin-like 4	0	3
herc3	HECT and RLD domain containing E3 ubiquitin protein ligase 3	1	3
kcnq4	potassium channel, voltage gated KQT-like subfamily Q, member 4	0	3
LOC106512958	mothers against decapentaplegic homolog 3	2	3
LOC106513868	nuclear receptor ROR-beta-like	1	3
LOC106514145	tripartite motif-containing protein 3-like	1	3
LOC106515472	LIM domain kinase 1-like	1	3

CHAPTER 4:

Vitamin D is (not) for diapause

Introduction

Diapause in embryos of annual killifishes is one of the most extreme examples of dormancy in vertebrates. Diapause is an environmentally-induced alternative developmental phenotype that is directly related to their ability to exploit seasonal habitats, and thus provides an unprecedented opportunity to study genome-environment interactions in vertebrate development in an ecologically and evolutionarily relevant context (Podrabsky, Lopez et al. 2007, Podrabsky, Tingaud-Sequeira et al. 2010). Entrance into diapause can be regulated by maternal programming cued by as yet unknown factors that may be related to maternal age. Further, incubation temperature can alter developmental trajectory with cooler temperatures favoring entrance into diapause. Diapausing embryos enter into a profound state of metabolic depression, and gain the ability to survive extreme environmental stress. The unique physiological phenotypes associated with diapause will almost certainly result from dramatic changes of gene expression. However, most changes in gene expression associated with entrance into diapause will likely be associated with normal development, or the maintenance and commitment to the diapause phenotype. Thus, identification of the regulatory pathways that initiate diapause may be difficult using gene expression analyses. In this study, we use weighted gene co-expression network analysis, and a broad comparative approach to identify the molecular pathways that regulate entrance into diapause in embryos of the annual killifish *Austrofundulus limnaeus*.

Using gene expression analysis to identify key regulators of phenotype (e.g., transcription factors) during development can be difficult due to their generally low abundance and often transient expression at the transcriptional level. Thus, highly complex gene regulatory networks must be explored both temporally and spatially to develop a complete picture of their actions and effects. This requirement often leads to complex experimental designs that include large groups of individuals or species, increased number of developmental timepoints, and use of highly sensitive gene sequencing technologies; all of which also increase the complexity and variation in the data set. Yet, the most rigorous attempts to discover gene pathways that underlie specific phenotypes, even extreme phenotypes such as diapause, often only identify the downstream effectors that define and maintain the phenotype. In only a few invertebrate examples have the molecular factors that induce developmental dormancy been resolved.

Evidence from work in *C. elegans* and *Drosophila spp.* have demonstrated that integration of extracellular signals, in the form of hormonal signaling, leads to receptor binding and gene signaling cascades that regulate major life history decisions (Baumeister, Schaffitzel et al. 2006, Fielenbach and Antebi 2008, Emerson, Bradshaw et al. 2009). In *C. elegans*, environmental and dietary signals lead to the production of growth factors such as insulin/IGF, TGF-β, and synthesis of environmentally responsive hormones. The hormones are the ligands for nuclear receptors that once ligand-bound activate gene pathways that promote growth and reproduction. During conditions of crowding or starving, a lack of production of growth factors and signaling hormones leave the nuclear receptor unbound—a state that promotes strictly dauer transcription

programs (review in (Fielenbach and Antebi 2008). Drosophila also utilize a strategy that is condition-specific whereby hormone-bound nuclear receptors induce growth and reproductive maturity rather than diapause when unbound (Saunders, Richard et al. 1990). The key genetic regulator of the initiation of diapause in vertebrates remains elusive.

Entrance into diapause is an alternative developmental trajectory in embryos of *A. limnaeus*, and developmental phenotype can be predicted by environmental conditions experienced by the embryo. Importantly, embryos developing along the two trajectories due to incubation at different temperatures are morphologically indistinguishable during early development until about midway through somitogenesis. This provides an experimental opportunity to explore gene expression during early development that is diapause-specific because gene expression associated with the morphogenesis will presumably be similar in the two trajectories, and thus only diapause-specific genes should be differentially expressed. While the exact timing of gene expression that regulates phenotype are unknown, a critical window of development during mid somitogenesis (between 10 and 16 pairs of somites) has been identified where commitment to phenotype is irreversible. In addition, consistent with what is known in *C. elegans*, repression of insulin-like growth factor signaling plays a strong role in supporting the development of the diapause phenotype (Woll and Podrabsky 2017). Here we investigate phenotype-specific gene expression networks using a systems-level approach. This approach allows for the generation of gene co-expression modules built upon transcriptome profiles of embryos developing along the diapause trajectory.

Importantly, these modules can be cross-analyzed using transcriptomes from embryos developing along the escape trajectory to test for shared and unique gene networks across developmental time. Preservation of modules in both trajectories would be expected for networks that underlie developmental processes, while those that are not preserved should highlight networks unique to each trajectory – specifically those that are key to the regulation of diapause. This approach has led to the identification of vitamin D signaling as the major regulatory pathway that controls temperature-induced entrance into diapause. Further, it has generated a number of gene modules that likely hold the key to understanding the extreme metabolic and physiological phenotypes associated with diapause in *A. limnaeus*.

Methods

Experimental design

Details regarding the animal strain and adult maintenance can be found in Chapter 2. Embryo collection, temperature incubations, sampling, and RNA extractions are previously described in Chapter 3.

Transcriptome analysis

RNA extraction and poly-A high throughput sequencing as well as preliminary bioinformatics processing of sequence read data (quality assessment, trimming, and alignment to *A. limnaeus* reference genome) was performed as described previously (Romney and Podrabsky 2017). Gene counts were normalized as fragments per kilobase

of transcript per million mapped reads (FPKM). Count matrices of detected genes from all sequence libraries were filtered for genes with 1 or more normalized counts summed across all replicates. Differential gene expression was calculated at each stage of development between diapause and escape samples using DESeq2 software from the R Bioconductor package (Love, Huber et al. 2014). Significant difference was determined using a Benjamini-Hochberg multiple comparisons adjusted FDR of 10%.

Network Co-expression analysis

A network of co-expressed genes was constructed for each phenotypic trajectory using the Weighted Gene Co-expression Network Analysis (WGCNA) R software package (Langfelder and Horvath 2008). WGCNA first uses all FPKM values to construct a signed co-expression network from a Pearson's correlation matrix. Then, an adjacency matrix is calculated by raising the absolute value of the correlation matrix to a power of 18 (recommended for networks with 21 samples) to amplify high correlations from low background correlations. The interconnectedness (topological overlap) of each gene pair within a network was calculated using the adjacency matrix, which was then used as input for average linkage hierarchical clustering. Groups of genes that have high topological overlap (clustered connectivity values) were assigned to modules to represent patterns of co-expression within the network. These gene modules, the branches of the resulting clustering tree, were cut to remove the effect of noise within the defined modules using the dynamic tree-cut algorithm. We summarized expression profile of each module by representing it as the first principal component; referred to as a module

eigengene. Module eigengenes represent the weighted average of module gene expression. Additionally, a gene membership value (kME) was calculated that represents the correlation between the expression values for each gene and the module eigengene for each sample. We assigned genes to modules if they had a module membership defined as kME>0.7, and genes with a value below this threshold were assigned to the grey module. Genes with high intramodular membership (kME > 0.90) are considered hub genes that are closely associated with the expression profile of the module. Networks of intramodular connections between genes were visualized with Cytoscape version 3.5.1 (Shannon, Markiel et al. 2003) based on topological overlap determined by WGCNA. Using the WGCNA modulePreservation function to generate module preservation statistics, we tested whether the density and connectivity patterns of gene modules defined in the diapause network are preserved within the escape network using the $Z_{summary}$ statistic > 10 as strong evidence for module preservation.

Gene ontology

Functional annotation of *A. limnaeus* genes was determined by generating protein homologs for all gene identities to *Homo sapiens* gene IDs using BLASTp with a threshold e value less than $1.0E^{-5}$. Using the Database for Annotation, Visualization and Integrated Discovery (DAVID) bioinformatics software, clusters of *A. limnaeus* genes that were identified by a *H. sapiens* homolog gene ID, were tested for gene ontology enrichment (Benjamani-Hochberg adjusted *P* values < 0.05).

Dauer and Vitamin D metabolism pathway analysis

To investigate functionality of expression pathways unique to developmental trajectory in *A. limnaeus*, comparisons were made to the well-established DAF gene family pathway of the dauer phenotype in *C. elegans* and the pathway of vertebrate vitamin D metabolism. Using the AmiGO database (version 2.4.26) from the GO Consortium (Carbon, Ireland et al. 2009), a list of genes was collected from *C. elegans* that were functionally annotated with the GO class keywords 'dauer' and 'DAF' as well as genes from all vertebrates with GO class keywords 'vitamin D metabolism'. Homologous genes were determined by comparing protein sequences from both genes lists to all *A. limnaeus* protein sequences using NCBI BLASTp (Altschul, Gish et al. 1990) with a sequence similarity threshold e value of $1.0E^{-05}$ (DAF pathway) and $1.0E^{-15}$ (VD3 metabolism pathways).

Vitamin D precursor treatments

Killifish embryos were collected after fertilization and maintained at 25°C with minimal light in embryo medium (see Chapter 2). At 4 dpf, embryos were transferred to temperatures favorable for the diapause trajectory, 20 and 25°C, and incubated in media supplemented with the following vitamin D precursors (n=12, each): 1 % v/v DMSO (M + DMSO), 1% v/v ethanol (M + EtOH), 1 nM 7-dehydrocholesterol (7-DHC; Sigma Aldrich), 10 µM cholecalciferol (Sigma Aldric),1 µM ergocalciferol (Sigma Aldrich),100 nM calcifidiol (Selleck Chemicals), and 10 nM alfacalcidol (Selleck Chemicals). Treated embryos were monitored throughout development using minimal yellow light under 40 X

total magnification on an inverted compound microscope (Leica DMIRB, Wetzlar, Germany). Data were analyzed and graphically depicted as percentage diapause or escape embryos by 38 dpf. Representative embryos were photographed digitally using a Leica DFC450C camera and Leica Application Suite V4.3 software (Leica Micro- systems, Buffalo Grove, IL, USA, 2013).

Results and Discussion

Transcriptomic analysis

Embryos were reared in groups at 30 or 20°C to induce escape and diapause trajectories and sampled at seven morphological stages of development: cell dispersion (CD), neural keel (NK), 6, 10, 16, 20, and 24 somite pairs (see Chapter 3 Methods; (Podrabsky, Garrett et al. 2010). Pooled samples of 20-40 embryos were used for cDNA library preparation and RNA-sequencing (see Methods and Chapter 3). Reads were trimmed and aligned to transcripts based on their overlap with the 26,729 reference gene and pseudogene models defined in the NCBI *Austrofundulus limnaeus* Annotation Release 100 (Wagner, Warren et al. 2015) APPENDIX C.1 and C.2). Transcriptomes consisted of 12,472 -14,002 total normalized gene transcripts with 1-28% of total transcripts differentially expressed in diapause- and escape-destined embryos. Thus, within 24 h of being transferred to 30°C, the transcriptomes of the embryos were already diverging even though the embryos are morphologically indistinguishable (≥ 2 fragments per kilobase per million (FPKM); FDR 10%; Figure 4.1A). Biotypes of genes in constituting the transcriptomes were predominantly protein-coding mRNA gene

transcripts, with lesser contributions of long non-coding, pseudogenes, and tRNAs (Figure 4.1B). Full transcriptomes per library (n = 42) can be found in detail in Supplemental File 4.1. Gene-level differential expression analysis revealed large differences between phenotypes throughout development. Highly differentially expressed genes were enriched for gene ontology annotations of calcium signaling and cell cycle regulation. Many, and perhaps most of the genes that are identified as differentially expressed via this methodology appear to be largely symptomatic of development phenotype. For example, it is already known and predictable that cell cycle regulation should be unique in the two developmental trajectories. However, the enrichment of differential expression for genes involved in calcium signaling is somewhat surprising given the general importance of calcium signaling in normal development and cell function. This enrichment, while difficult to interpret in isolation, suggests that signaling pathways that regulate calcium homeostasis, such as parathyroid hormone and vitamin D, may be important in the regulation of diapause (APPENDICES C.3-C.9).

Network Co-expression analysis

Genes that induce regulatory signaling cascades can go undetected in large transcriptomic analyses if they are of low abundance or transiently expressed. However, evidence of their downstream effects is often observed in differential gene expression between treatment groups. To overcome the limitations of identifying regulatory genes in large RNA-seq datasets, we applied a systems-level approach. Network analysis by WGCNA calculates the similarity of expression patterns among all pairs of genes across

all treatment conditions, assigning each gene to a 'module' based on the shared expression patterns. Gene-level expression patterns from the diapause trajectory of approximately 16,000 genes were collapsed into 21 modules of correlated expression patterns (Figure 4.2A). Modules ranged in size from 9,474 to 30 genes (the lower limit). Notably, 17 modules showed stage-specific expression values having positive module eigengene correlation values at a single stage of development ($r > 0.50$; $P < 10^{-2}$, Figure 4.2B). These modules consist of a total of 4,957 genes. Testing for the preservation of co-expression in the escape trajectory of diapause-assigned modules can provide crucial insights phenotype-specific genes networks (Langfelder and Horvath 2008, Langfelder, Luo et al. 2011). Network topology of 8 modules were conserved in the escape trajectory ($Z_{summary} >10$) and represent strongly conserved transcriptional pathways of important development programs between trajectories. The remaining modules were not conserved in the escape trajectory ($Z_{summary} <10$) and represent diapause- and escape- specific networks of genes. To explore the functional significance of each module as well as the modules that are specific to each stage, we tested for enrichment of gene ontology (GO). While this can be applied to large modules with genes that are well-annotated with GO identities, we found small modules or modules consisting of relatively unannotated genes to show no statistical enrichment, again leaving us short in the understanding of the transcriptomic profiles that drive phenotype.

From the WGCNA measure of intramodular connectivity (kME), we identified 3,034 intramodular hub genes (kME > 0.90, $P < 10^{-8}$; Table 4.1) in the diapause trajectory as a means to discover key regulators of each module (Langfelder and Horvath

2008). In the earliest stage sampled, CD, there were three modules identified: yellow, turquoise, and light yellow. The light yellow module was much smaller (39 genes) than the others and the diapause trajectory network was not preserved in the escape trajectory ($Z_{summary}$ = 6.2). Gene ontology analysis showed no statistical enrichment for GO identities, but the low number of transcripts in the module likely explains this result. This module is characterized by transcripts of low abundance, that tend to peak in abundance during the cell dispersion stage (Figure 4.3A). Interestingly very few of these transcripts were found to be differentially expressed despite clear trajectory-specific patterns of expression. This result is likely due to the low abundance of these transcripts, which are often expressed with higher variability and thus are more difficult to identify as differentially expressed in large datasets due to the limits of performing thousands of statistical comparisons (Figure 4.3B). One of the hub transcripts with the highest kME values was identified as LOC106533739, which annotates in the *A. limnaeus* genome as cholesterol desaturase daf-36-like (daf-36-like, kME = 0.98, $P < 10^{-14}$). Expression of this transcript is escape-trajectory specific, with almost undetectable levels of expression in diapausing embryos. Further, the daf-36-like transcript is centrally located within the light yellow module gene network in diapause trajectory embryos, and its position shifts in the escape trajectory network suggesting a dramatic restructuring of connections within this network between the two trajectories (Figure 4.3C). This gene is the homolog to daf-36 in *C. elegans*, a protein-coding gene that catalyzes the first step in the production of dafachronic acids, hormones that regulate dauer-arrest (Wang and Kim 2003, Rottiers, Motola et al. 2006, Fielenbach and Antebi 2008). In *C. elegans,*

dafachronic acids bind directly to a nuclear hormone receptor (NHR) and transcription factor, daf-12, that initiates a cascade of signals to suppress dauer formation and induce reproductive growth (Fielenbach and Antebi 2008). Importantly, in the absence of dafachronic acids, the daf-12 receptor partners with other proteins to initiate gene expression networks that induce dauer formation, thus the gene regulatory activity of daf-12 is dependent on the presence or absence of dafachronic acids.

The identification of a daf-36-like transcript that is differentially regulated in diapause and escape trajectory embryos of *A. limnaeus* motivated us to explore the potential role of daf gene networks in the regulation of diapause in *A. limnaeus*. Thus, we identified all *A. limnaeus* genes homologous to dauer arrest and DAF family genes in *C. elegans* (Figure 4.4A and B). Consistent with other vertebrates, the daf-12 homolog in *A. limna*eus annotates as a vitamin D receptor B (VDR). The NHR, daf-12, is particularly critical to the signal transduction pathway integrating environmental cues to nuclear transcriptional pathways to modify life history decisions in *C. elegans* (Antebi, Yeh et al. 2000, Fielenbach and Antebi 2008). It is well understood that environmental induction of insulin/IGF-1, TGFβ, and cGMP signaling pathways converge on DAF-12 as the key regulator of developmental phenotype (Fielenbach and Antebi 2008). Thus, we hypothesized that the vertebrate vitamin D signaling pathway, through the action of VDR and environmental production of vitamin D could be acting to regulate diapause in *A. limnaeus*.

Interestingly, vitamin D signaling pathway genes in *A. limnaeus* have striking phenotype-specific expression profiles (Figure 4.5A and B). There are two VDR genes in

A. limnaeus; α and β. VDR-α has a mean expression of < 5 FPKM across all development stages and so we continued our analysis focusing on VDR- β (referred to hereafter as VDR). Importantly, transcripts for genes that are involved in the synthesis of active vitamin D, 1α25-dihydroxyvitamin D3 (1,25$(OH)_2$D, or calcitriol), show trajectory-specific patterns of expression consistent with increased production of 1,25$(OH)_2$D in escape trajectory embryos (Figure 4.6). The daf-36-like enzyme catabolizes the production of 7-dehydrocholesterol (7-DHC), also known as provitamin D_3. Exposure of the 7-DHC to the ultraviolet B component of sunlight isomerizes 7-DHC into previtamin D_3, which is slowly converted to vitamin D_3 (cholecalciferol) through a thermochemical process (Holick, MacLaughlin et al. 1980). Conversion of vitamin D_3 to 1,25$(OH)_2$D which binds with high affinity to the vitamin D receptor, is catalyzed by two subsequent enzymatic conversion (Figure 4.6; (McDonnell, Mangelsdorf et al. 1987). The first step is the production of 25-hydroxyvitamin D_3 (25(OH)D), or calcifediol) catalyzed by the cytochrome P450 enzyme vitamin D 25-hydroxylase (CYP2R1, homologous to LOC106530854 in *A. limnaeus,* e = 0.0E+00). The final synthetic step is catalyzed by another cytochrome P450 enzyme, 25OHD-1α hydroxylase (CYP27B1, homologous to LOC106518084, 25-hydroxyvitamin D-1 alpha hydroxylase mitochondrial, e = 0.0E+00, in *A. limnaeus*). Both of these enzymes are upregulated in embryos incubated at 30°C during the critical window where increased incubation temperature leads to irreversible commitment to the escape trajectory (Figure 4.6). Thus, escape-specific expression of all enzymes critical for production of 1,25$(OH)_2$D are expressed in a manner consistent with increased production of active vitamin D_3.

The biologically active form of vitamin D_3, $1,25(OH)_2D$, is the ligand for a transcription factor, the VDR. While it is understood that VDRs exist across vertebrate taxa and are essential for calcium homeostasis in adults, its role in developmental progression has been only briefly examined (Kwon 2016). In *A. limnaeus*, five potential VDR homologs were identified and two of these genes (LOC106518359—vitamin D3 receptor B and *nr1i2*, nuclear receptor subfamily 1, group I, member 2)—exhibit trajectory-specific expression patterns (Figure 4.7). Expression of both receptors is similar in CD, NK and 6 somite embryos, after which their expression diverges with a significant decrease in transcript abundance in escape trajectory embryos. This divergence in expression coincides with the critical window of development when embryos are committed to the escape trajectory by incubation temperature. While this pattern of expression may seem counterintuitive for a role of VDRs in determination of the escape phenotype, this pattern of expression is likely due to negative feedback regulation of VDR gene expression by its own activity which is consistent with the known regulation of VDR gene expression in mammals, and is also consistent with patterns of daf-12 gene expression in *C. elegans* (Hammell, Karp et al. 2009). As reported by Hammell et al. (Hammell, Karp et al. 2009) and Bethke et al. (2009) a negative feedback loop of daf-12 expression occurs upon steroid binding committing development towards active growth.

VDR activity is thought to be dependent on the formation of heterodimers with other NHR, such as the retinoid X receptors. In *A. limnaeus*, the receptor LOC106517732, retinoid X receptor beta A like (RXR-beta-A-like) transcript is

expressed in a relatively stable pattern across both developmental trajectories, and thus is one possible partner for regulating VDR binding affinity (Figure 4.7). The VDR/RXR complex is responsible for the activation of gene transcription networks associated with calcium transport, hormone secretion, and cellular proliferation and differentiation (Ramagopalan, Heger et al. 2010). Accordingly, trajectory-specific, highly differentially expressed genes identified in this study are enriched for GO annotations in cell cycle regulation, calcium transport, and other vitamin-D related downstream processes (APPENDICES C.3-C.9).

It is critical to note that most of what is known about the synthesis and signaling of vitamin D has been studied in humans and other mammals. In mammals, the production of vitamin D_3 is dependent on UV light in the skin, and subsequence stepwise conversion to $1,25(OH)_2D$ by 25-hydroxylase activity in the liver, and 1α-hydroxylase activity in the kidneys (Bikle 2014). The canonical requirement for solar UV radiation and environmental temperature in the formation of vitamin D_3 presents an opportunity for environmental factors to directly affect developmental trajectory in *A. limnaeus*. What is known about the effect of light and temperature on entrance into diapause is consistent with a role for active vitamin D_3 signaling promoting the escape trajectory. First, increased temperatures lead to escape embryo development (Podrabsky, Garrett et al. 2010). Second, exposure of embryos to full spectrum light can lead to release from diapause (Podrabsky and Hand 1999), and can induce early embryos to develop along the escape trajectory (Figure 4.8). However, the ecological relevance of UV exposure in fish embryos that exist in muddy waters or buried in the soil is questionable. It is possible that

aerial exposure of embryos occurs as the soils dry and the large cracks that form in clay soils allows light to penetrate. However, it is also important to note that production of active vitamin D_3 is possible in a number of species without exposure to sunlight and there is some debate about alternative routes for the catalytic production of vitamin D_3 from 7-DHC (Norman and Norman 1993). It is also known that thermal conversion of vitamin D_3 from previtamin D_3 is over a thousand times more efficient in cellular membranes than in solution. This increase in efficiency is consistent with a microenvironment, or possible enzymatic partner, that greatly facilitates conversion and thus suggests possible routes for UV-independent conversion of vitamin D_3 in *A. limnaeus*. Thus, while there are still many details to be worked out concerning the production of active vitamin D_3 and the molecular consequences of vitamin D signaling in *A. limnaeus*, we hypothesize that vitamin D signaling regulates developmental trajectory associated with entrance into diapause in *A. limnaeus*.

Vitamin D precursor treatments

To test the vitamin D hypothesis, we exposed embryos to various vitamin D analogs during development at 20 and 25°C to see if exogenous incubation could lead to escape embryos even under conditions that should favor development along the diapause trajectory. When exposed to 7-DHC (10 nM), cholecalciferol (1 µM), ergocalciferol (1 µM), calcifidiol (100 nM), and the synthetic vitamin D analog alfacalcidol (10 nM), 100% of *A. limnaeus* embryos (n = 12), developed along the escape trajectory at 25°C. Similarly, when incubated at 20°C in the presence of vitamin D analogs, embryos

developed strictly along the escape trajectory in all cases expect for 7-DHC (Figure 4.9A). These data are strong experimental evidence that active vitamin D signaling induces embryos to develop along the escape trajectory. Interestingly, 7-DHC was able to induce the escape trajectory at 25°C, but not at 20°C. This may indicate that a temperature of 20°C is below the threshold of thermal activity required to covert 7-DHC to vitamin D_3. Alternatively, it may be linked to the temperature-dependent expression of a novel enzyme that can catalyze the conversion. Additional studies are required to work out this part of the vitamin D synthetic pathway in *A. limnaeus*.

Downstream changes in gene expression as a result of vitamin D signaling

There were seven modules that displayed gene expression patterns that were positively correlated with multiple stages of development, whether or not they were determined as stage-specific (modules magenta, red, brown, grey, blue, black, and purple; Figure 4.2B). This is not surprising considering transcriptomic diversity also increases throughout developmental time requiring gene programs for newly differentiating tissue types (Peshkin, Wühr et al. 2015). However, among them, only the brown module was not preserved in co-expression patterns between diapause and escape networks. A lack of preservation likely reflects diapause-specific gene networks that co-express in preparation for entry into dormancy. The brown module consists of 634 genes that are enriched along the diapause trajectory starting after the cell dispersion stage APPENDIX C10-30. With module hub genes such as fibroblast growth factor receptor 2 (*fgfr2*) and homeobox D3 (*hoxd3*), this module may reveal some interesting contribution to the

phenotype regulation that supports entry into diapause downstream of the vitamin D signaling cascade and requires future investigation.

Conclusion

WGCNA co-expression analysis provided a powerful perspective on a large transcriptomic dataset and lead to the discovery of a critical role for vitamin D signaling in the regulation of developmental trajectory. Without this systems level approach, it is highly unlikely that this line of investigation would have been initiated. Importantly, the vertebrate vitamin D receptor is homologous to daf-12, a nuclear hormone receptor in *C. elegans* known to regulate dauer arrest, a state similar to diapause (Fielenbach and Antebi 2008). The involvement of vitamin D signaling in life history decisions and metabolic arrest in a vertebrate model is a novel finding beyond its canonical role in calcium homeostasis. In addition, this is the first description, in a vertebrate, of a molecular signaling pathway that regulates developmental trajectory and identifies a direct molecular mechanism for the integration of environmental cues in the developmental program. However, the most significant implication of this work is the discovery of a potential deep homology between nuclear hormone receptors and their conserved functions to regulate major life history decisions in a diverse group of animals. Given the importance of metabolic dormancy across all domains of life, the conservation of the daf-12/vitamin D receptor control of dormancy in *C. elegans* and *A. limnaeus* suggests a deep homology for this receptor in the regulation of life-history and plasticity during animal

development. The regulation of life history *A. limnaeus* by signaling cascades induced by the VDR will provide a critical tool for developmental biologists to explore the link between inheritance and phenotype across vertebrate lineages.

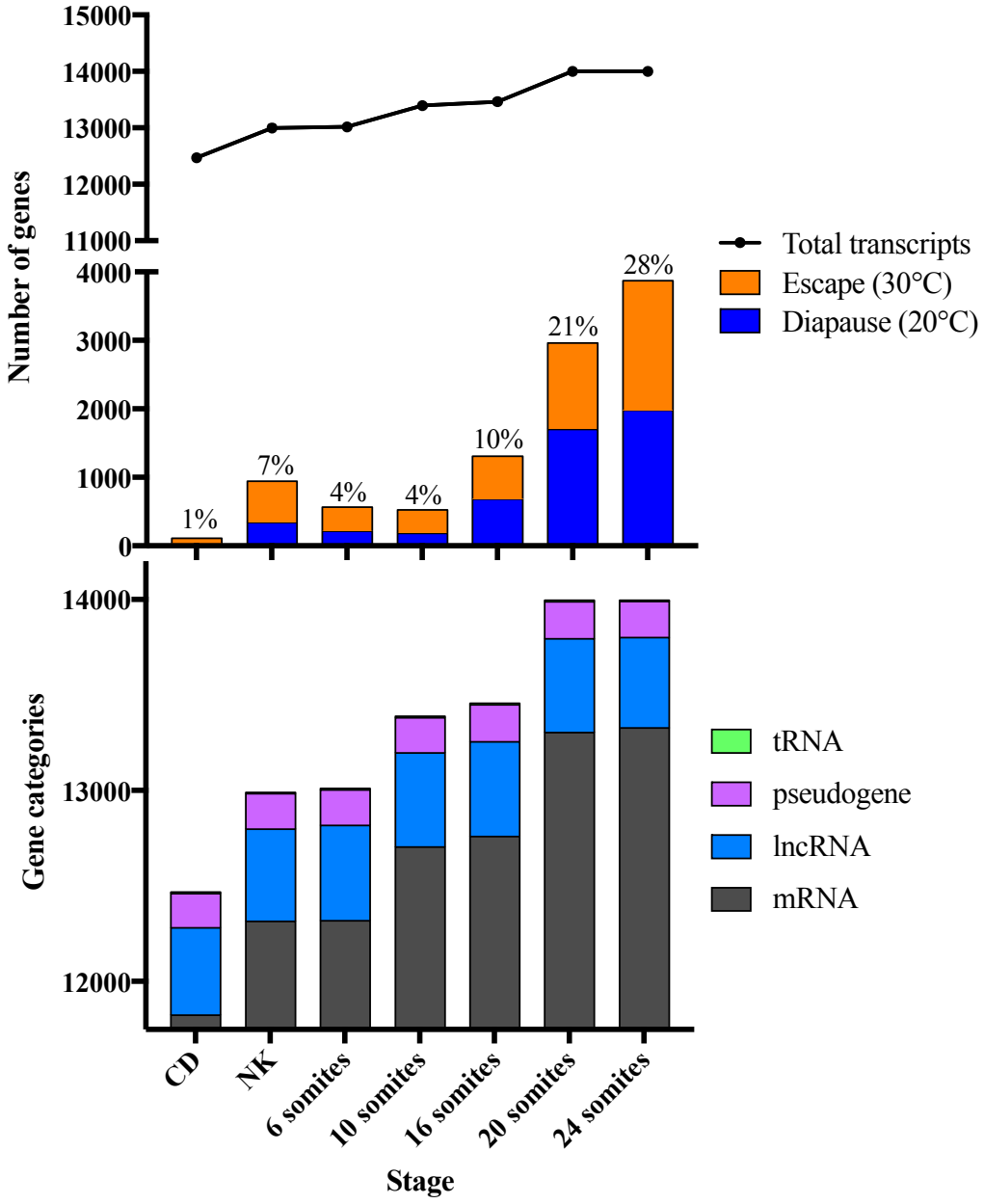


Figure 4.1: Transcriptome analysis of genes throughout development of *A. limnaeus*. (A) Total number of gene transcriptomes detected at each stage of development (line) and the number of genes differentially expressed genes (DEG; bars) at each stage of development in the escape (orange) and diapause (blue) phenotypes. The percentage of DEGs is identified above each bar. (B) Abundance of gene biological class types (biotypes) at each stage of development determined by NCBI genome annotation (see Methods).

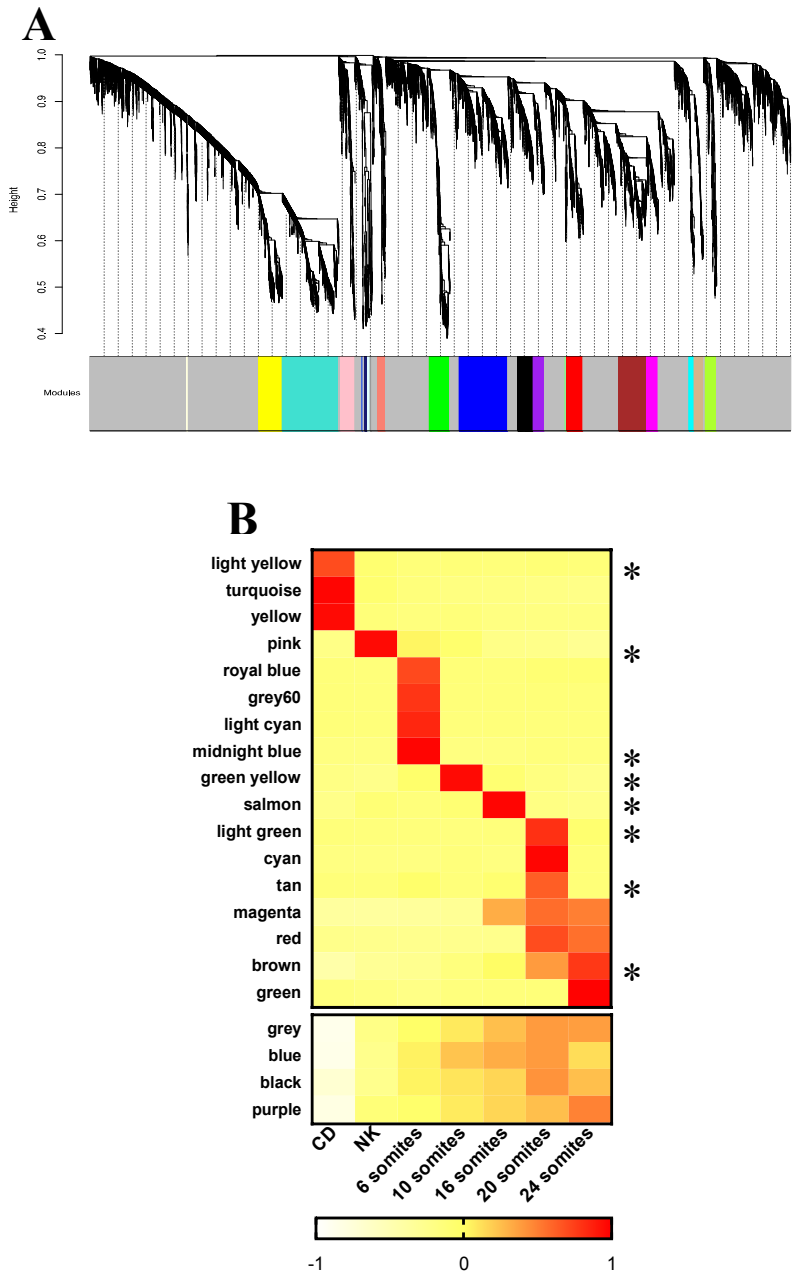


Figure 4.2: Network analysis of gene co-expression along the diapause trajectory. (A) Hierarchical cluster tree where genes form branches based on a measure of similarity (y axis; 'height') in gene expression profiles. The data presented here correspond to a similarity matrix table. Gene with correlations to each other are highly connected, or similar, and are grouped into gene clusters (modules) identified in color along the x-axis. (B) Heat map of module eigengene correlations to each stage of development along the diapause trajectory (positive is red and negative is white). Modules within the upper panel are significantly associated ($r > 0.50$; p value $< 10^{-2}$) and with any stage; modules within the lower panel are not. Modules with asterisks on the right side of panel are not preserved between diapause and escape networks.

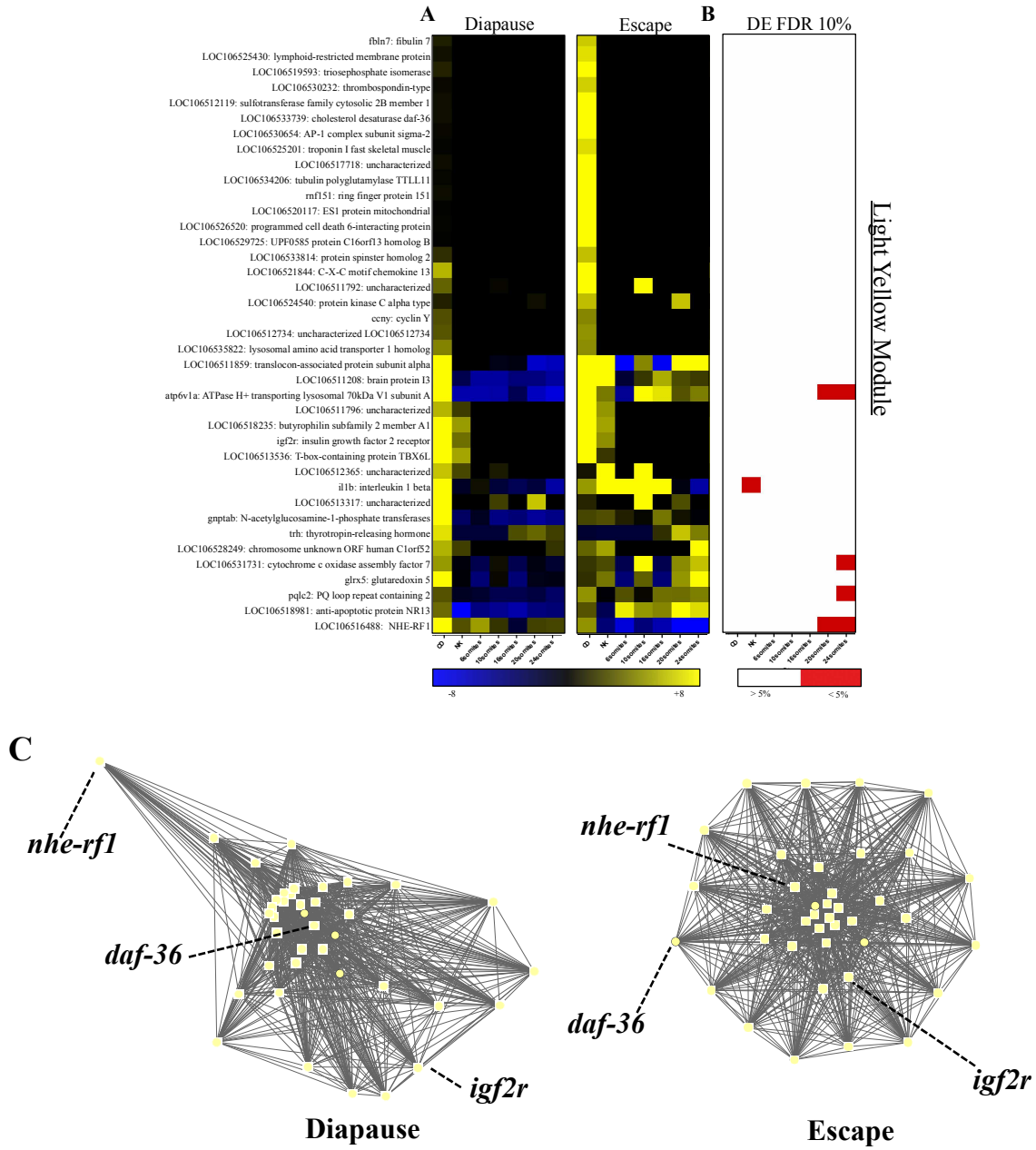


Figure 4.3: The light yellow module is not preserved between networks at the cell dispersion stage. (A) Heat maps of relative fold changes of normalized expression values (FPKM) across development and phenotypic trajectory of the light yellow module (39 genes). Each row represents genes that are up-regulated in yellow and down-regulated genes in blue across developmental stages. (B) Heat map showing significant *p* values < FDR 10% (red) of differential expression (DE) between phenotypes for each gene and stage of development. (C) A visualization of the light yellow module with yellow circles representing the genes and the connecting lines represent similarity strengths (thickness). The left module represents the diapause trajectory and the right module is from the escape trajectory. The gene LOC106533739– Cholesterol-desaturase *daf-36*, is identified by its position within each, along with *nhe-rf1* and *igf2r*.

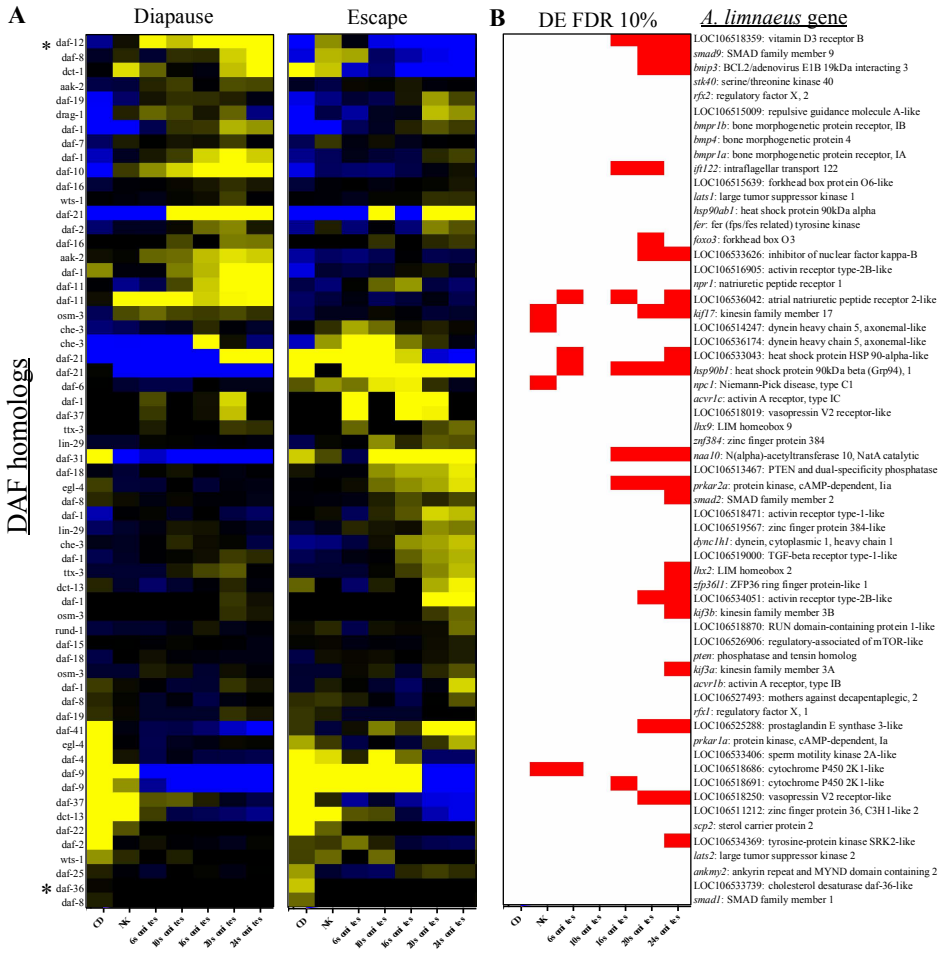


Figure 4.4: DAF gene homologs in *A. limnaeus*. (A) Heat maps of relative FPKM values of *A. limnaeus* genes for each phenotype (identified by labels on the right-side) that are homologous to the *C. elegans* DAF gene family (identified by gene symbols on the left) form the dauer arrest pathway. Daf-36 and the NHR Daf-12 have an asterisk. (B) Heat map showing significant *p* values < FDR 10% (red) of DE between phenotypes for each gene.

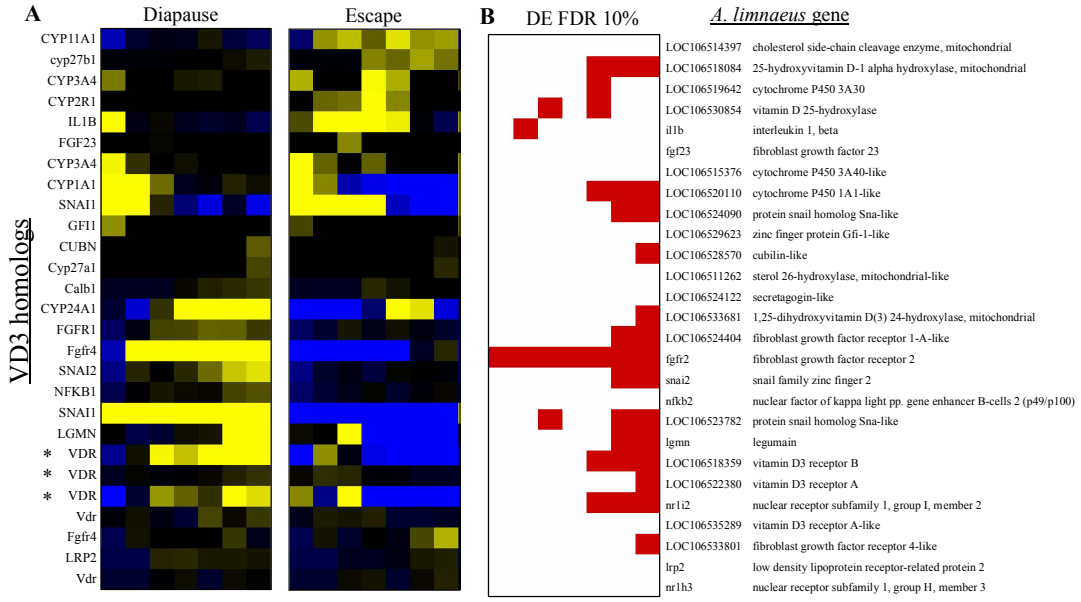


Figure 4.5: Vitamin D3 (VD3) gene homologs in *A. limnaeus*. (A) Heat maps of relative FPKM values of *A. limnaeus* genes for each phenotype (identified by labels on the right-side) that are homologous to vertebrate spp. VD3 gene family (identified by gene symbols on the left). Vitamin D receptors (VDR) have an asterisk. (B) Heat map showing significant *p* values < FDR 10% (red) of DE between phenotypes for each gene.

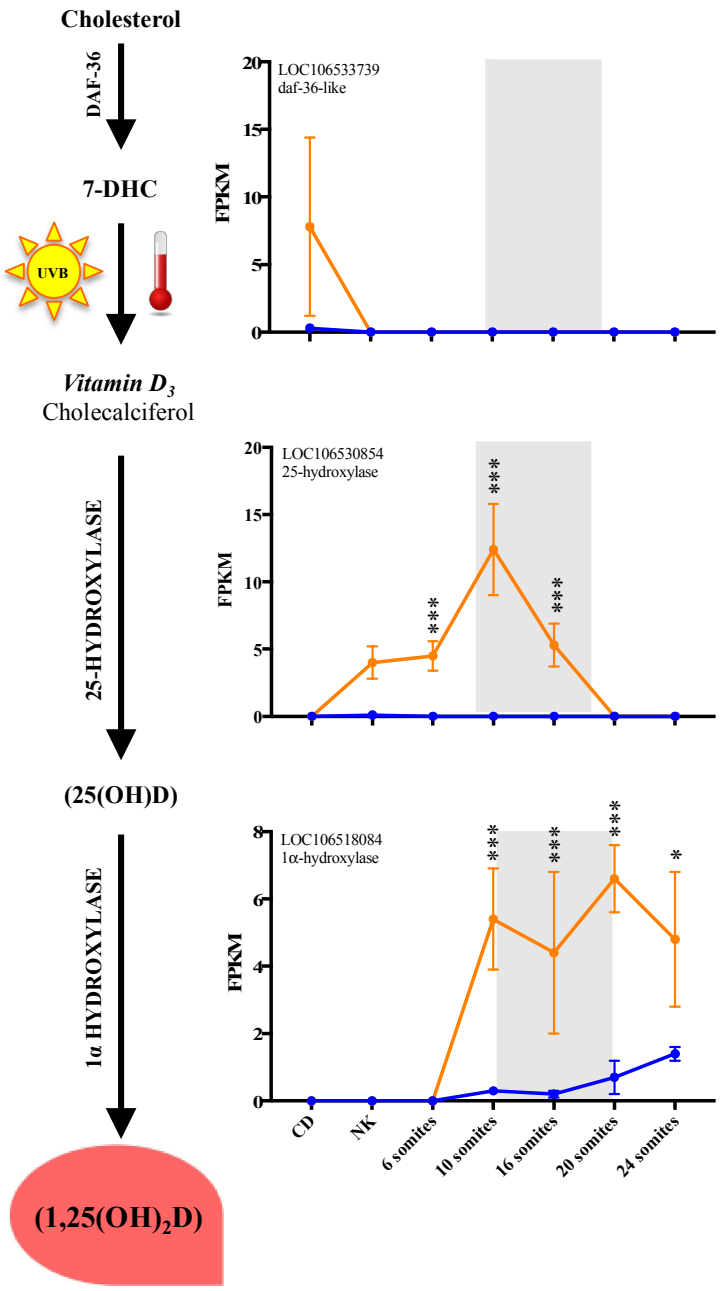


Figure 4.6: Vitamin D Synthesis. Cholesterol is converted from 7-dehydrocholesterol (pro-vitamin D3; DHC) by the enzyme daf-36 and and further into cholecalciferol (vitamin D3) in response to UV light exposure and warm temperatures. Cholecalciferol is converted into 25-hydroxyvitamin D3 ($25(OH)D_3$) by the enzyme 35-hydroxylase. Conversion by the enzyme 1α-hydroxylase makes the biologically active form of vitamin D3, ($1,25(OH)_2D$). Homologous genes to synthesis enzymes show mean FPKM values ± SEM across development along escape (orange) and diapause (blue) phenotypes. Asterisks identify DE genes with significant *p* values < FDR 10% (one), 5% (two), and 1% (three) thresholds.

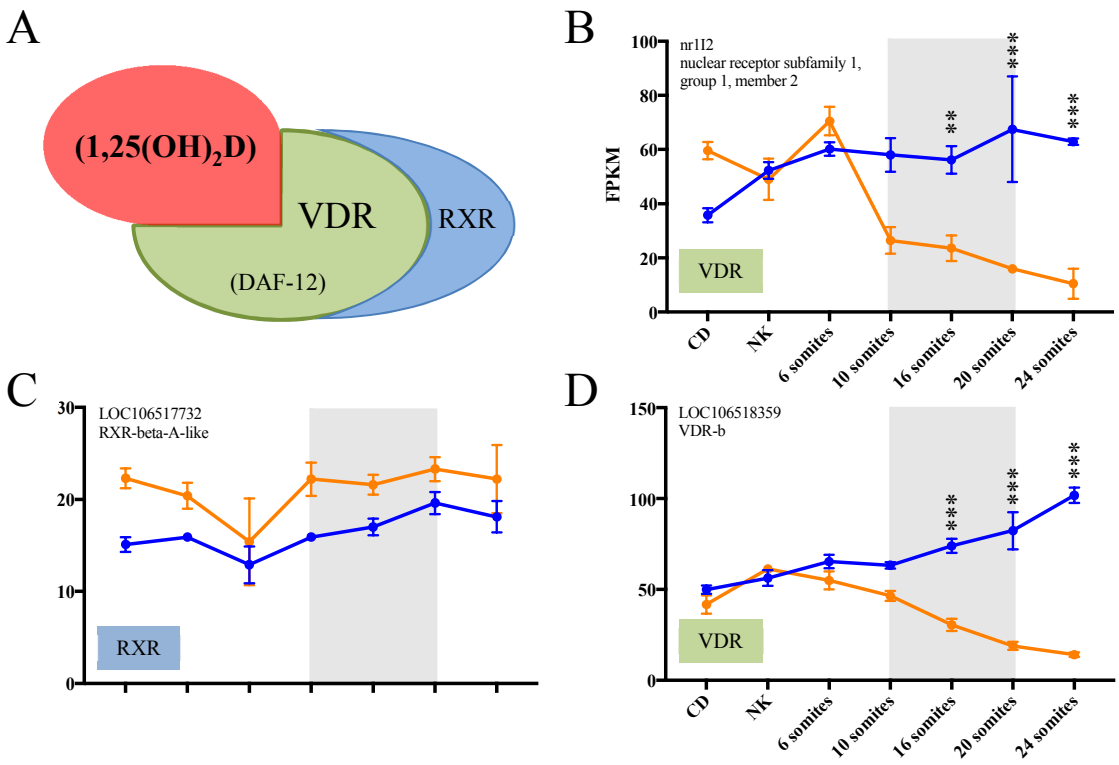


Figure 4.7: Vitamin D Receptors. (A) Schematic of ligand binding (red) to the Vitamin D Receptors (VDR; green). Upon binding, the VDR heterodimerizes with other nuclear hormone receptors in the nucleus, most commonly, members of the Retinoic X receptors (RXR; blue). (B) In A. limnaeus, VDR beta, (LOC106518359), and the nuclear receptor, *nr1I2* (a VDR homolog) demonstrate divergent expression profiles (FPKM, ±SEM) prior to the commitment to developmental phenotype (escape is orange, and diapause is blue. Asterisks identify DE genes with significant *p* values < FDR 10% (one), 5% (two), and 1% (3) thresholds.

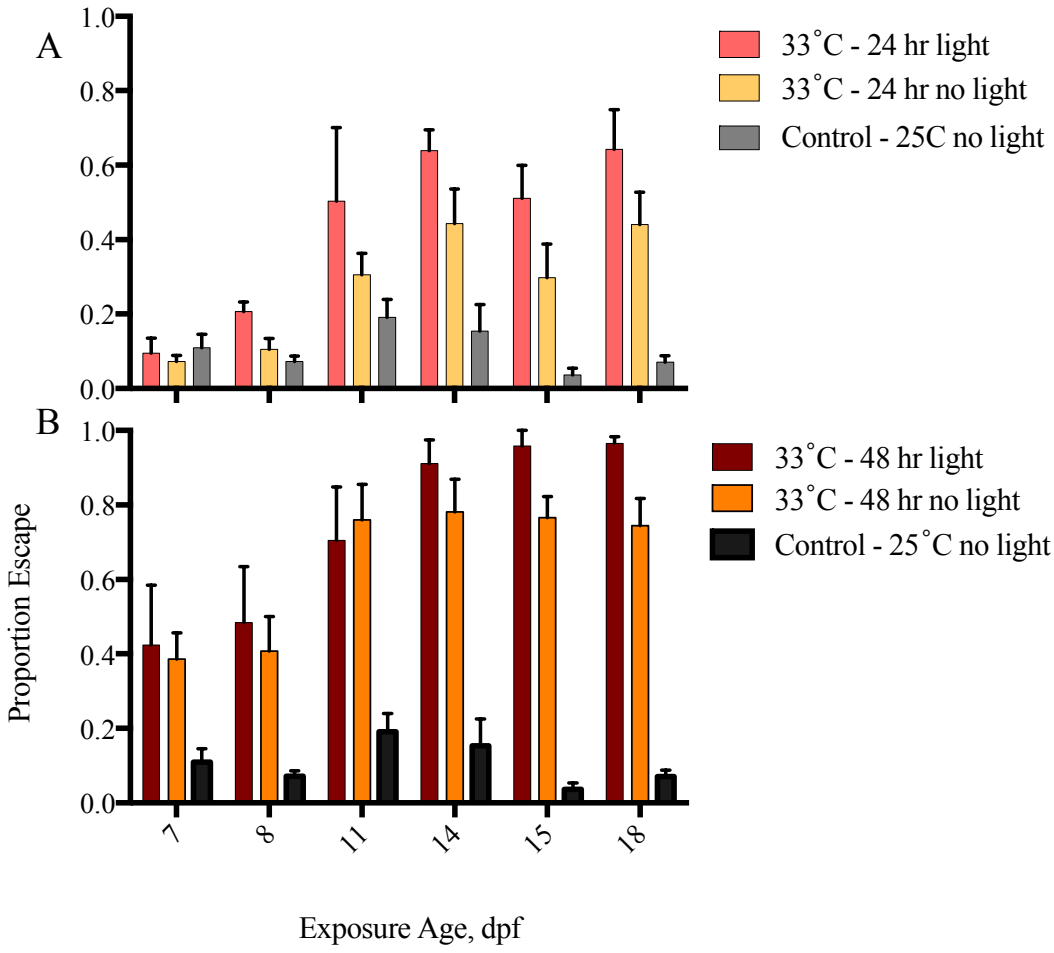


Figure 4.8: Light and temperature effects on developmental phenotype. Increased proportions of embryos develop along the escape trajectory when incubated at 33 C and exposed to short (24 and 48 hr) durations of UV light exposure.

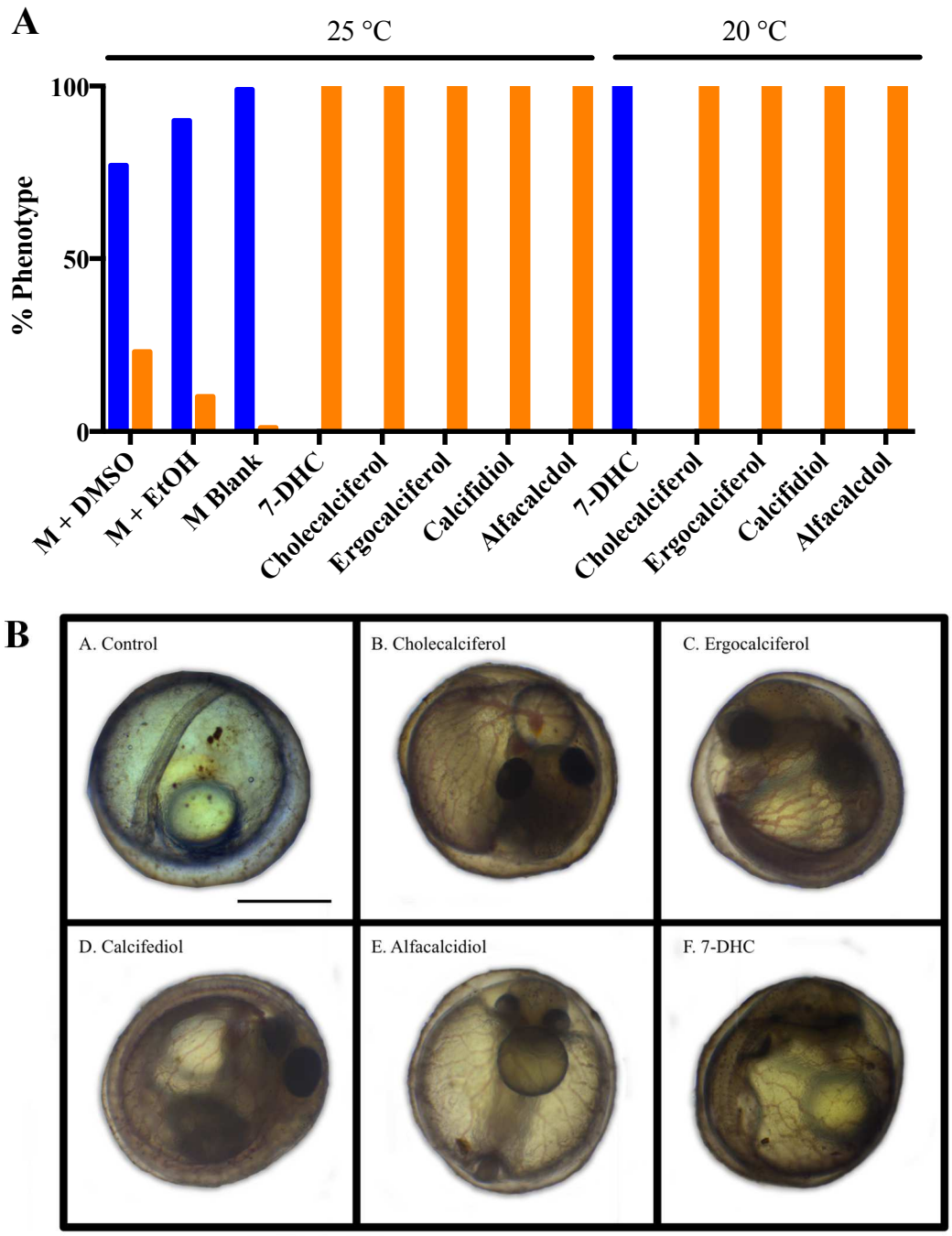


Figure 4.9: Embryonic exposures to Vitamin D precursors. (A) Embryos (n = 12) were exposed to embryo media (M) supplemented with vitamin D precursors and allowed to develop at 25 and 20 °C and scored by phenotype which was determined morphologically as escape (orange) and diapause (blue). (B) Representative images of embryos from each treatment at 22dpf; Control (M + blank), is a diapause embryo where all others are escape trajectory embryos. Scale bar = 1 mm.

Table 4.1: Summary of modules and hub genes. Each module can be described by what stage of development it's expression pattern is highly correlated with. Each module has an abundance of genes with high intramodular connectivity (kME) at high correlations ($r > 0.90$). Modules with the diapause network can be tested for preservation in the escape trajectory. Hubs can be determined in the escape trajectory whether they maintain high kME values or not.

Stage	Module	Hubs > 0.90	Total genes	$Z_{summary}$ Score	Hubs remain in test network
Epiboly	Yellow	468	540	25.4	105
Epiboly	Turquoise	1046	1305	24.9	236
Epiboly	Light yellow	30	39	**6.2**	16
Neural Keel	Pink	58	336	13.6	16
6 somites	Royal Blue	21	30	**9.1**	17
6 somites	Grey60	33	41	11.0	29
6 somites	Light cyan	37	49	13.1	23
6 somites	Midnight Blue	58	65	**9.5**	16
10 somites	Green yellow	42	251	**8.5**	13
16 somites	Salmon	30	179	**8.9**	16
20 somites	Light green	33	40	**9.2**	18
20 somites	Cyan	77	118	16.8	50
20 somites	Tan	77	205	**5.5**	27
20 somites	Red	197	375	31.0	129
20 somites	Magenta	202	275	29.0	87
24 somites	Green	209	475	21.2	135
24 somites	Brown	416	634	**4.5**	15
Later development	Black*	141	356	11.2	10
Later development	Blue*	167	1126	18.0	46
Later development	Grey*	113	9474	**9.6**	9
Later development	Purple*	107	261	12.4	17

CHAPTER 5:

Connecting gene expression associated with alternative developmental trajectories across life-history stages

Introduction

The regulation of gene expression in an early embryo is crucial in determining cellular differentiation and generation of phenotype. Maternal effects and epigenetic inheritance are known to be important in the development of phenotype in a variety of animal systems, yet these topics are seldom investigated in species capable of phenotypic plasticity. In this thesis, I characterize the poly-A and small RNA transcriptomes across multiple developmental stages, and two life stages of *A. limnaeus*. This work has identified gene regulatory signatures from maternal programming as well as environmentally responsive networks in embryos developing along escape and diapause trajectories. This work has identified unique trajectory-specific mRNA splice variants as a potential mechanism that supports maternal programming of developmental phenotype. Transcriptomic analysis of embryos at later stages of development suggests that *miR-430* family miRNAs exhibit differential expression in response to temperature. The known role for these miRNAs in the clearing of maternally provisioned transcripts further suggests a mechanism by which the developmental environment can override maternal programming. Most importantly, developmental trajectory in *A. limnaeus* appears to be predominantly regulated by vitamin D receptor (VDR) signaling during the 6-10 somite pair stage that directs signaling cascades for growth and differentiation in support of the escape trajectory. As the evidence of vitamin D signaling was only recently discovered, it is necessary to revisit earlier findings, especially with respect to maternal programming,

to explore the potential of these regulatory mechanisms to alter vitamin D signaling and thus affect developmental trajectory. I hypothesize that signatures of gene expression associated with regulation of vitamin D synthesis or VDR signaling may be present at fertilization that support developmental trajectory through maternal programming. Additionally, patterns of miRNA expression during early development may support commitment to development phenotype by targeting maternal transcripts, or other mRNA targets at later stages of development that can regulate vitamin D metabolism and signaling. Specifically, my aims are to 1) determine if the splice variants at fertilization may alter vitamin D metabolism and signaling at later stages of development 2) determine if zygotically transcribed miRNAs have the potential to shape the vitamin D signaling cascade to influence the development of alternative phenotypes.

Contributions to Vitamin D signaling by maternal mRNA splice variants at fertilization

Maternal control over developmental phenotype could occur through provisioning of gene products (RNA, proteins) into the cytoplasm of the egg during oogenesis. Maternally packaged gene products are known to occur in all vertebrates (Li, Zheng et al. 2010). In *A. limnaeus*, direct maternal programming through differential gene expression does not occur. However, splice variants of maternally packaged transcripts are differentially packaged into eggs that will develop along either trajectory. Of the 57 genes that were determined to have trajectory-specific splice variants, 10 were found in the literature to be functionally related to vitamin D signaling pathways and cholesterol metabolism (Table 5.1).

prolactin-releasing hormone, prlh

The gene, prolactin-releasing hormone (*prlh*), is possibly the most promising splice variant related to vitamin D signaling to pursue because prolactin is understood to increase production of vitamin D (Christakos, Ajibade et al. 2010). In diapausing embryos at fertilization, there is an increased abundance of the third exon whereas escape embryos omitted this exon almost completely (Figure 5.1A). Known also as prolactin-releasing peptide (PrRP), vertebrate *prlh* generates two peptide isoforms that differ in length and have been investigated in the Japanese crucian carp (Fujimoto, Takeshita et al. 1998, Tachibana and Sakamoto 2014). Early work described that the protein isoforms of this gene result from post-translational modification; however, in a study by Anderson et al (2003), it was suggested that the two peptide isoforms could result from alternative mRNA splicing. The longer mRNA isoform in the present study likely corresponds to the longer variant identified by Anderson et al. The alternative biological activities of *prlh* mRNA isoforms have not been described, and cannot currently be linked as independent origins of each *prlh* peptide in *A. limnaeus*. However, there may be some alternative functional contribution of the insertion of exon 3 present in the diapause embryos. While *prlh* has been initially described as stimulator of prolactin (PRL) release in teleosts, it has been more recently been described as a negative regulator of growth hormone release (Moriyama, Ito et al. 2002, Seale, Itoh et al. 2002). Furthermore, Hinuma et al (1998) demonstrated that *prlh*-induced PRL secretion increases arachidonic acid metabolism, a fatty acid known to inhibit dauer formation in *Caenorhabditis elegans* (Becker 2014).

Since insulin-like growth factor signaling is known to be critical for the regulation of developmental trajectory in *A. limnaeus* (Woll and Podrabsky 2017), reduction of growth hormone release could be a potential mechanism to promote the diapause trajectory by blocking the activity of insulin-like growth factor signaling independent of other signaling pathways.

N-myristoyltransferase 2, nmt2

Another gene that was discovered with differential exon usage at fertilization was N-myristoyltransferase 2, (*nmt2*). In *A. limnaeus*, embryos that are destined to follow the diapause trajectory were packaged with a greater abundance of exon 8 when compared with escape embryos (Figure 5.1B). Transcript variants are known to exist in the literature for this gene and have been detected with differences in length that mimic the variants observed in *A. limnaeus* (109 nt; (McIlhinney, Young et al. 1998). Tissue-specific splice patterns have been reported in adult rats, but nothing is known about differences in biological activity of the variants. However, enzymatic activity is coded in the c-terminus of the translated product which corresponds to the 3'-end of the transcribed mRNA sequence and may perhaps the shorter variant has altered for absent catalytic activity (Weston, Camble et al. 1998).

nmt2, also known in the literature as myristoyl-CoA: protein N-myristoyltransferase-2, performs post-translational modifications by attaching myristic acid, a 14-carbon saturated fatty acid, to proteins. Known proteins that are myristoylated by *nmt2* are ADP-ribosylation factors (ARFs), myristoylated alanine rich protein kinase

C substrate (MARCKS), prokineticin receptor 1 (PKR1), cytochrome B5 reductase 3 (CYB5R3), and nitric oxide synthase (NOS) (Hentschel, Zahedi et al. 2016). Two gene paralogs have been described in humans and some other mammals, N-myristoyltransferase 1 (*nmt1*) and *nmt*2 with overlapping roles and substrate specificity in adult tissues (Rioux, Beauchamp et al. 2006). Compared to *nmt1*, *nmt2* is less enzymatically active in mouse stages of early cell division, and is (Yang, Shrivastav et al. 2005). Their role in global translational control (reviewed in (Curtis, Lehmann et al. 1995) provides convincing evidence for a contribution to the regulation of protein synthesis and thus metabolism and growth rate in early embryos. Myristoylation of a protein often serve to target the protein to membranes. As vitamin D metabolism likely occurs within the membrane compartments in cells, an analysis of differentially myristoylated proteins in escape- and diapause-bound embryos may help to identify proteins involved in the regulation of vitamin D metabolism.

Ephrin receptor B2, ephb2

Ephrin receptors belong to a family of receptor tyrosine kinase transmembrane glycoproteins that are located within membranes of most cell types. Ephrin receptors and their membrane-bound ligands, ephrins, are both capable of transducing signaling cascades, a molecular property that is less common among cellular receptors, termed bidirectionality (Pasquale 2010). This form of signaling provides a mechanism that leads to cell repulsion, a feature incredibly important during early development that is associated with cell morphology, adhesion, and migration (Comer 2007). Cell-cell

recognition via ephrin-ephb2 binding events provide critical cues for axon pathfinding and formation of cardiovascular networks in early embryos (Flanagan and Vanderhaeghen 1998, Frisén, Holmberg et al. 1999). While many reports document the critical role of morphological development in embryos by ephrin-ephb2-signaling, very little on the functional differences of mRNA splice variants has been described.

At fertilization, only embryos of *A. limnaeus* that will develop along the diapause trajectory are packaged with an *ephb2* transcript that contains exon 4 (Figure 5.1C). Both variants retain the domain structure critical for ephrin receptor function (Thierry-Mieg and Thierry-Mieg 2006), but *in vivo* studies are lacking to demonstrate additional or alternative roles for these variants.

Importantly, *ephb2* expression is regulated by β-catenin/TCF signaling (Brückner and Klein 1998). In the nucleus, β-catenin binds to proteins of the T-cell factor/ lymphoid enhancer factor (TCF/LEF) and regulates the expression of genes for cell cycle progression and cellular differentiation, including *ephb2* (Van De Wetering, Sancho et al. 2002, Larriba, Valle et al. 2007). Importantly, vitamin D_3 (VD3) binding to VDR inhibits β-catenin signaling. Therefore, maternal packaging of *ephb2* may be a way to ensure the activity of this gene even in the presence of active VD3 signaling.

The genes discussed here were previously identified as possible maternally provisioned cues that could affect developmental trajectory. With a more detailed understanding of their functional roles, we can conclude that the isoforms have the potential to regulate or augment vitamin D signaling. Direct investigations of these

proteins and their effect on cellular processes will be needed to more fully understand their roles in the potential regulation of developmental phenotype in *A. limnaeus*.

Contributions to Vitamin D signaling by zygotic-miRNAs

The previous analysis of temperature-responsive sncRNAs was focused on the possible role of post-transcriptional modification of maternally packaged mRNAs (Chapter 3). However, activation of the VDR appears to occur at later stages of development during early somitogenesis. After this stage, the two developmental trajectories begin to diverge. Thus, a re-evaluation of possible miRNA targets involved in vitamin D signaling during later stages of development is warranted.

The miRNA, *miR-430*, was described in detail in Chapter 3 and predicted gene targets include binding sites within the 3' translated region (UTR). The *miR-430* gene family is known primarily for accelerating the deadenylation and decay of maternally loaded mRNAs, but other roles for this family have been described (Giraldez, Mishima et al. 2006, Bazzini, Lee et al. 2012). In zebrafish, microarray expression analysis of maternal–zygotic *dicer* mutants (MZ*dicer*) and subsequent *in vivo* experimentation identified 203 *miR-430* target mRNAs, a population that was enriched (~4-fold) for maternally provided mRNAs (Giraldez, Mishima et al. 2006). Giraldez, et al. (2005), first identified the role of *miR-430* in context of neurulation. His findings suggested that *miR-430* supported neural tube folding that occurs concomitantly with epiboly in zebrafish. His team later identified maternal mRNA transcripts that are directly targeted by *miR-430* (Giraldez, Mishima et al. 2006). They specifically profiled target genes at and around the

stages of epiboly to describe the importance of *miR-430* under the context of the maternal to zygotic transition and establishment of the embryonic axis (MTZ). Non-canonical roles for *miR-430* are not well understood.

A total of 2,270 genes were identified in the *A. limnaeus* genome with a 6-mer *miR-430* seed within its 3' UTR (Chapter 3). These predicted gene targets could potentially influence developmental phenotype in response to incubation temperature. Increased expression of mature *miR-430* sequences in escape trajectory embryos suggest that abundance of potential targets should have decreased in escape embryos and increased in diapause-bound embryos. Alternatively, *miR-430* expression in escape embryos may simply be required for precise expression of escape-specific genes. The subset of mRNAs that are enriched for putative *miR-430* binding sites in *A. limnaeus* are more common during early development in escape trajectory embryos. However, at the 20 somite stage of development, *miR-430* targets become much more abundant in diapause-bound embryos (Figure 5.2). This shift in available targets may result from clearing of *miR-430* targets from the escape-bound embryos. *miR-430* may therefore be contributing to gene expression in both phenotypes, but targeting different transcripts. Further, it is important to note that miRNA action most often blocks translation of proteins rather than resulting in degradation of mRNA target transcripts. Thus, we may not expect to see a change in transcript abundance even in the face of active translational blockage. Below, I discuss mRNAs that are potential targets of *miR-430* that may also play a role in regulation of developmental phenotype due to their action in the dauer pathway of *C. elegans* or the vertebrate vitamin D signaling cascade.

Activin receptor type-2B

In *C. elegans*, DAF-1 is a type II serine/threonine kinase transforming growth factor-beta (TGF- β) receptor expressed in neurons. Under favorable conditions, this receptor binds TGF- β and transduces signals intracellularly that induce hormone binding to DAF-12 in *C. elegans* that promote growth and development (Fielenbach and Antebi 2008). In *A. limnaeus*, LOC106516905, activin receptor type-2B, is homologous to DAF-1 and contains a *miR-430* binding site (7-mer). Activins are growth and differentiation factors that belong to the TGF- β superfamily of structurally related signaling proteins. Type IIB receptors are constitutively active and have higher affinity for the growth factor ligand than type A (Sako, Grinberg et al. 2010). Throughout development, LOC106516905 is almost 2-fold more abundant in diapause trajectory embryos; however, this difference is not statistically significant (Figure 5.3A). This expression pattern is nonetheless consistent with *miR-430* targeting, but seems counter to the known role for DAF-1 in the promotion of developmental dormancy in *C. elegans*. Alternatively, increased transcript abundance in diapause-bound embryos may represent an increased sensitivity to escape-promoting signals that must be detected by this receptor. Future studies at the protein level will be needed to clarify the potential role of this gene in the regulation of diapause in *A. limnaeus*.

forkhead box O3

In *C. elegans*, DAF-16 has also been extensively described regarding its role in dauer formation. The forkhead transcription factor, DAF-16/FOXO, is the key nuclear mediator of longevity and stress resistance associated with entrance into dauer dormancy. In unfavorable environments, insulin/IGF-1 signaling is inactivated, which allows DAF-16/FOXO to enter the nucleus, where it turns on genes for stress resistance, dauer formation, and longevity (Fielenbach and Antebi 2008). In *A. limnaeus*, the DAF-16 homolog, *foxo3*, forkhead box O3, has two separate *miR-430* binding sites (7-mer and 6-mer). While expression was undetected in the earliest stages of development, *foxo3* increases at later stages and is statistically higher in diapause embryos at 20 pairs of somites (Figure 5.3B). If *foxo3* acts in *A. limnaeus* as it does in other animals, then *miR-430* blockage of expression would result in the lack of expression of diapause-specific proteins.

Repulsive guidance molecule A-like

Another dauer-related signaling molecule is DRAG-1, the solitary member of the repulsive guidance molecule (RGM) family in *C. elegans* (Gumienny and Savage-Dunn 2013). DRAG-1 promotes the non-dauer phenotypes of normal body size and mesoderm development through the positive regulation of bone morphogenetic protein (BMP) signaling (Tian, Sen et al. 2010, Tian, Shi et al. 2013). In *A. limnaeus*, the homolog for DRAG-1, LOC106515009 repulsive guidance molecule A-like, carries a 7-mer *miR-430* binding site within its 3' UTR. Expression is detected in both trajectories at the neural

keel stage (Figure 5.3C). Abundance of this transcript is roughly equivalent along the two developmental trajectories, suggesting that differential expression due to *miR-430* binding would have to be dictated by blockage of translation.

Vitamin D receptor A

Surprisingly, one of the VDRs of *A. limnaeus*, LOC106522380 (VDRα) is a putative target for *miR-430* binding. Of the two forms (-α and -β), VDRα is expressed at substantially lower levels than its paralog VDRβ (Chapter 4), but with strikingly similar abundance trends between both genes (Figure 5.3D). The two paralogs are conserved among teleosts, and in medaka, it has been demonstrated that both are functional nuclear hormone receptors capable of DNA binding, ligand binding, heterodimerization with retinoid X receptor, and coactivator recruitment (Howarth, Law et al. 2008). However, this study also demonstrated significant functional divergence between the paralogs with VDRβ exhibiting a higher sensitivity to ligand binding. The study concluded that in medaka, paralogous copies of VDRs may demonstrate an evolutionary transitional state for this gene, and that VDRα may exhibit novel functions that are yet to be characterized (Howarth, Law et al. 2008). Thus, the role of VDRα and its potential role in the regulation of developmental trajectory is uncertain and will require a functional comparison of the two possible CDV paralogs in *A. limnaeus*.

Key components of the DAF and vitamin D signaling pathways that have the potential to regulate developmental trajectory in *A. limnaeus* are potential targets of *miR-430*. Repression of gene expression via *miR-430* has the potential to modify the

transcriptomic profiles of escape-promoting genes such as LOC106516905, (activin receptor type-2B), LOC106515009, (repulsive guidance molecule A-like), and LOC106522380 (VDRα). Such mechanisms likely aid to sharpen the transcriptomic transitions associated with the escape trajectory and the heterokairic shift between developmental phenotypes.

Alternatively, *miR-430* may serve to directly silence diapause signals and cue trajectory along the escape pathway under conditions that favor growth and development. The expression profile for *foxo3*, forkhead box O3, is consistent with this hypothesis. The degree to which *miR-430* may regulate development could be better understood with further investigation and the validation of the *in silico* predicted targets discussed here. Nonetheless, the likelihood that *miR-430* participates in complicated patterns of dampening and balancing of antagonist/agonist relationships of mRNA transcripts along either trajectory very likely exist in the regulation of phenotype in this species. It is also possible that *miR-430* targeting plays both roles in promoting the diapause and escape trajectories by targeting a wide variety of transcripts involved in numerous cell signaling pathways.

The transcriptomic programs that regulate alternative developmental trajectory are active across two life stages of *A. limnaeus.* It is possible that these two transcriptomic profiles are integrated into a system that allows for maternal programming and environmental modification. The analysis detailed here characterizes the significance of maternally packaged splice variants at fertilization and the role of zygotically-transcribed miRNAs after the MTZ in light of the vitamin D signaling cascade that promotes

developmental trajectory in this species. As a conserved regulator of vertebrate development, ageing, and cancer, VDR may operate under a broad spectrum of genetic and epigenetic processes that have yet to be described. Further investigations on the functional properties of mRNA isoforms and miRNA gene targets may elucidate novel gene regulatory pathways of vitamin D signaling that support the adaptive phenotypic plasticity of *A. limnaeus*.

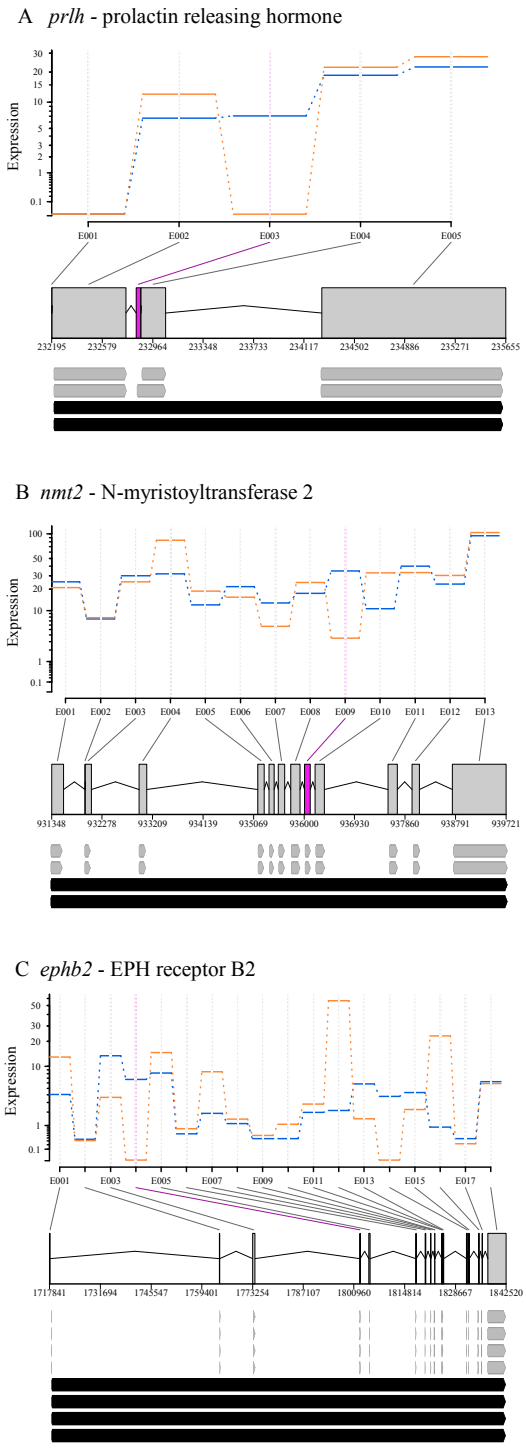


Figure 5.1: Intron and exon structures of genes and their annotated mRNA isoforms. Figures A-C represent genes with differential expression of exons between phenotypes in embryos at fertilization. The black segments at the bottom are unique mRNA transcript annotations for each gene and the grey segments above represent the exon pattern specific to each transcript. The exon highlighted in pink was statistically determined as differentially expressed (FDR 10%).

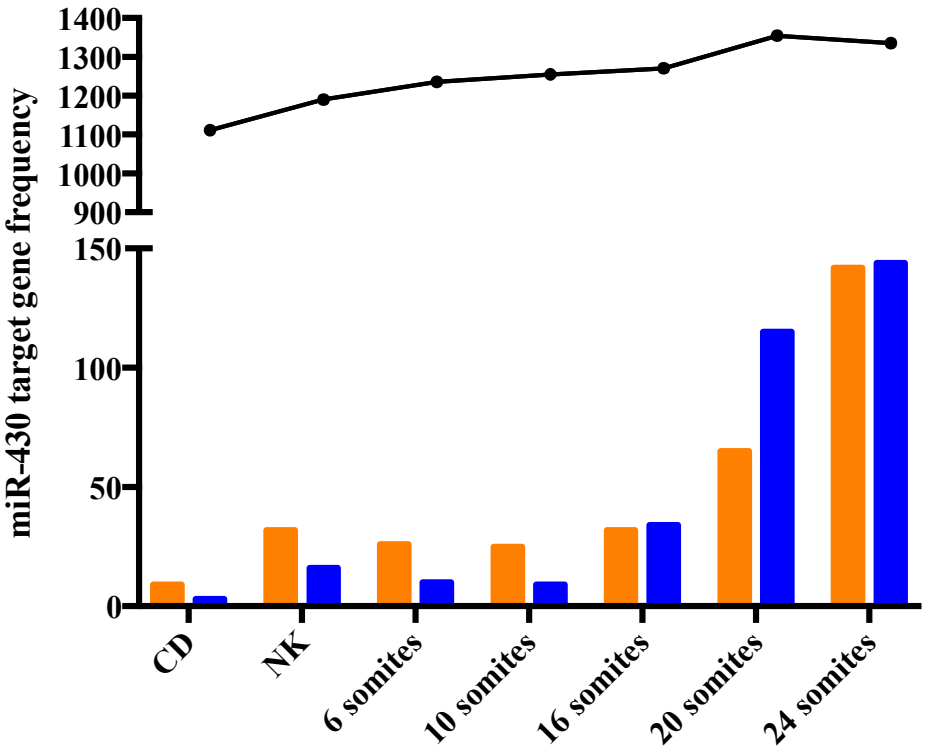


Figure 5.2: Frequencies of putative *miR-430* targets throughout development. *miR-430* binding sites of mRNAs in *A. limnaeus* genome were profiled and counted throughout development (black line). Genes that were either statistically upregulated along the escape trajectory (orange) and the diapause (trajectory) show unique abundances across developmental stages.

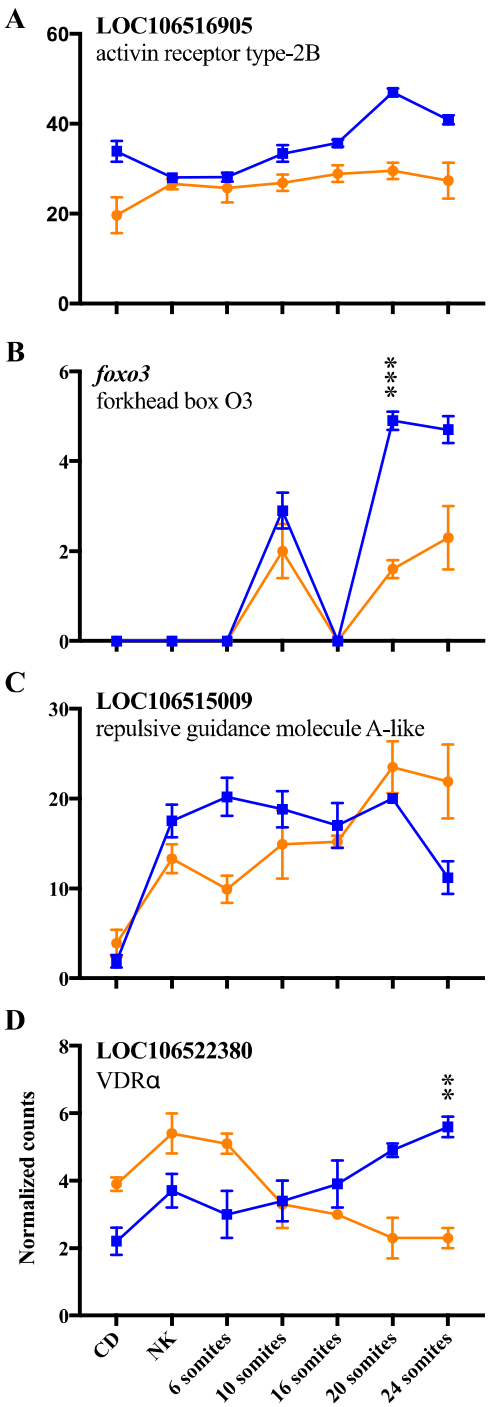


Figure 5.3: Transcriptomic profiles of putative *miR-430* targets throughout development. Figures A-D are protein-coding mRNAs in *A. limnaeus* that demonstrated homology with dauer signaling in *C. elegans* and vitamin D signaling pathway conserved among vertebrates. Profiles of FPKM expression are specific to the escape trajectory (orange) and the diapause (trajectory). Differential expression between phenotypes are identified by 1-3 asterisks (FDR 10%, 5%, and 1%, respectively).

Table 5.1: Differentially expressed exons in 1-2 cell stage embryos of *A. limnaeus.* Originally identified in chapter 2, these genes demonstrate relationships with steroid metabolism and vitamin D signaling.

Gene Symbol	**Description**	**Exon**	**Phenotype upregulated**
LOC106536223	Ankycorbin-like	5	Diapause
prlh	Prolactin releasing hormone	3	Diapause
LOC106535645	Filamin-C-like	6	Escape
nmt2	N-myristoyltransferase 2	9	Diapause
malrd1	MAM and LDL receptor class A domain containing 1	14	Diapause
snrnp200	Small nuclear ribonucleoprotein 200kDa (U5)	3	Escape
LOC106533628	Disintegrin and metalloproteinase domain-containing protein 8-like	14	Diapause
ephb2	EPH receptor B2	4	Diapause
kiaa0825	KIAA0825 ortholog	13	Diapause
dgkz	Diacylglycerol kinase, zeta	2	Diapause

CHAPTER 6:

Ecologically relevant developmental programs

This investigation of the development of *A. limnaeus* used the power of deep sequencing technology (RNA-seq) to identify gene regulatory mechanisms for phenotypic plasticity in a vertebrate. The use of network analysis uncovered vitamin D signaling as an environmentally responsive pathway that regulates a major life history decision in *A. limnaeus*. These results impact our understanding of the genetic mechanisms that regulate entrance into diapause in *A. limnaeus*, and provide insight into a potential conserved role for nuclear hormone receptors (NHRs) and environmentally responsive hormones in the regulation of major life history transitions across all animals.

Perhaps the most exciting discovery of my thesis research is the identification of vitamin D regulation of developmental phenotype. This mechanism can explain the importance of light and temperature cues for determining developmental trajectory. Furthermore, the homology of vitamin D signaling to the dauer regulating DAF pathway in *C. elegans* points to the potential for a conserved mechanism for the regulation of developmental arrest across all animals. Organismal development requires a degree of robustness despite perturbations from environmental conditions or random mutation to the genome. This idea, named canalization, provides critical constraints on ontogenetic programs, and is most important during the earliest stages of development. Canalization allows important molecular, cellular and physiological processes to persist throughout evolutionary time as a complex and adaptive system.

A systems level approach permitted the discovery of low abundance transcription factors that induce developmental trajectory in *A. limnaeus*. Such an approach can be divided into two methods (O'malley and Dupré 2005, Gilbert and Epel 2009). One, a pragmatic perspective, explores genetic networks in an attempt to understand outcomes of gene action; the other, a systems theory perspective, explores how new properties that arise from these gene networks possibly function in development. The discovery that the vitamin D receptor, a non-steroid nuclear hormone receptor, regulates developmental phenotype in *A. limnaeus* provides a conceptual framework to explore these critical evolutionary questions and perspectives in a vertebrate model. This study and a mountain of evidence in *C. elegans* and *Drosophila* suggest a conserved role of NHR regulation of life history decisions across vertebrates and non-vertebrates alike. In all of these cases, the harsh and often unpredictable environment experienced by the organisms has presumably driven the evolution of embryonic diapause as an adaptation for survival. The ability to arrest development likely utilized mechanisms that already existed in metazoans and were primed for integrating environmental stimuli into complex patterns of gene expression.

All transcription factors, in essence, convert signaling factors or cues into coordinated networks of gene expression. The key importance of NHRs, especially the VDR, is that they can very directly translate environmental cues without the need for a complex signaling cascade. The mechanisms that regulate dormancy among nematodes, arthropods, and now fishes, all rely upon cholesterol-derived hormones and ligand-activation of NHRs. In the nematode, *C. elegans*, environmentally cued synthesis of

dafachronic acids, which bind to the NHR, *daf-12*, illicit bypass of the diapause-like dauer stage (Fielenbach and Antebi 2008, Bethke, Fielenbach et al. 2009). In many arthropods, the binding of the ecdysone steroid to the ecdysone receptor is known to coordinate with over-wintering diapause (Denlinger 2002, Denlinger 2013). Nuclear hormone receptors are juxtaposed between endocrine signaling and gene expression regulation within all animal taxa and operate with potentially deeply conserved chemistries and biological function (Markov, Gutierrez-Mazariegos et al. 2017). Importantly, the synthesis of dafachronic acids, ecdysones, and vitamin D all share a common initial synthesis pathway that starts with the production of 7-dehydrocholesterol. Thus, it is possible that this 7-DHC pathway represents an ancient environmental hormone signaling pathway that has evolved to integrate environmental cues into the developmental program across animal phyla. Nuclear receptors are shared by all metazoans and are poised as critical regulators of metabolism, development, and physiological homeostasis, and thereby affect broad life history traits such as sex determination, molting, developmental timing, diapause, and life span (reviewed in (Antebi 2006).

Previous work and our findings presented here support a mechanistic model for the regulation of developmental trajectory and entrance into diapause in embryos of *A. limnaeus* (Figure 6.1). Under conditions of high temperature and perhaps exposure to light, vitamin D is synthesized in embryos and binding of vitamin D to the VDR induces transcriptional pathways that drive development along the escape trajectory. The absence of light and low temperatures prevents synthesis of vitamin D and thus the VDR absent

of ligand binding activates alternative transcriptional pathways associated with diapause. In my model, this basic mechanism can be altered by inputs from a variety of other signaling pathways, allowing for maternal and environmental mechanisms that control developmental outcomes to be integrated through vitamin D synthesis and signaling.

Earlier studies in *A. limnaeus* hinted at transcriptional regulation via NHRs in the regulation of diapause. Microarray analysis of gene expression in ovarian tissue from females producing exclusively escape embryos identified upregulation of a nuclear-receptor coactivator, activating signal cointegrator-1 (ASC-1; Pri-Tal 2010). This gene acts as a transcriptional coactivator of nuclear hormone receptors and a variety of other co-activators such as the steroid receptor coactivator, SRC-1. This link suggests a possible role for hormonal regulation of maternal programming of diapause that could integrate the female's nutritional environment, steroid hormone status, and other environmental cues (Kim, Yi et al. 1999, Lee, Lee et al. 2001).

Phenotypic plasticity during early development of *A. limnaeus* is known to be regulated across multiple life stages, with maternal programming playing a key role. Thus, any model for the regulation of entrance into diapause must be able to accommodate mechanisms for maternal input. Specifically, in *A. limnaeus*, there must exist interactions that connect maternally provisioned cues for developmental trajectory that can be altered by environmental triggers at later developmental stages. Careful reevaluation of maternally provisioned transcriptome under the lens of vitamin D signaling may help to identify these mechanistic links. At this point, the most likely candidate for altering the maternally provisioned transcriptome to influence

developmental trajectory is the temperature-specific expression of *miR-43*0 family members. While we lack the evidence to experimentally support this aspect of the model at this time, the gene expression data and known function of *miR-430* in clearing maternal transcripts suggests this as a viable hypothesis that should be explored in future studies.

Future experimentation into the regulation of developmental trajectory and diapause should be focused on regulation of the vitamin D signaling pathway. First, endogenous levels of vitamin D analogs during development and in response to appropriate environmental cues must be determined. Using a sampling regimen that mimics the present datasets would afford the ability to integrate metabolite abundance with RNA expression. Particularly interesting avenues to follow would be: 1) detect the endogenous levels of vitamin D precursors in diapause and escape embryos. This would include 7-dehydrocholesterol, cholecalciferol, calcifidiol, and calcitriol. 2) Profile changes in protein expression for the key enzymes important in synthesis of vitamin D, and VDR signaling. 3) Search for the downstream targets of VDR signaling. 4) Identify the specific location within the embryo where vitamin D signaling is acting to regulate developmental trajectory. 4) Finally, exposing *A. limnaeus* embryos to media supplemented with nematode dafachronic acid to explore the level of biochemical conservation of function in these two highly derived lineages.

It is also important to understand the possible regulatory contributions of *miR-430* to developmental phenotype in this species, and the potential for *miR-430* family members to alter vitamin D signaling. Predicted gene targets for *miR-430* should be

validated by experimental strategies that extract miRNA-mRNA duplexes. This is important because it is well understood that miRNA-mRNA seed pairing does not strictly rely upon perfect base pairing. *In silico* target predictions should therefore be experimentally determined. Highly efficient protocols for this exist, and an excellent place to start would be target RNA pull down with biotinylated microRNA mimics. In result, physical evidence of the interaction between specific miRNAs of interest and target mRNAs can be identified. Following these results, microinjection of oligonucleotides to knock down expression of mRNA targets could elucidate the effect of transcriptional modification by miRNAs on phenotypic development in *A. limnaeus*.

While a great deal of evidence has been collected that supports the roles of environmental hormones in the regulation of diapause throughout vertebrate and non-vertebrate systems, less attention is given to the nuclear receptors to which these ligands bind. We predict that there is strong evidence demonstrated here and shown in other species that developmental arrest in the form of diapause relies upon environmental hormone signaling and NHR activation that can translate environmental cues into genomic programs to alter major life history transitions.

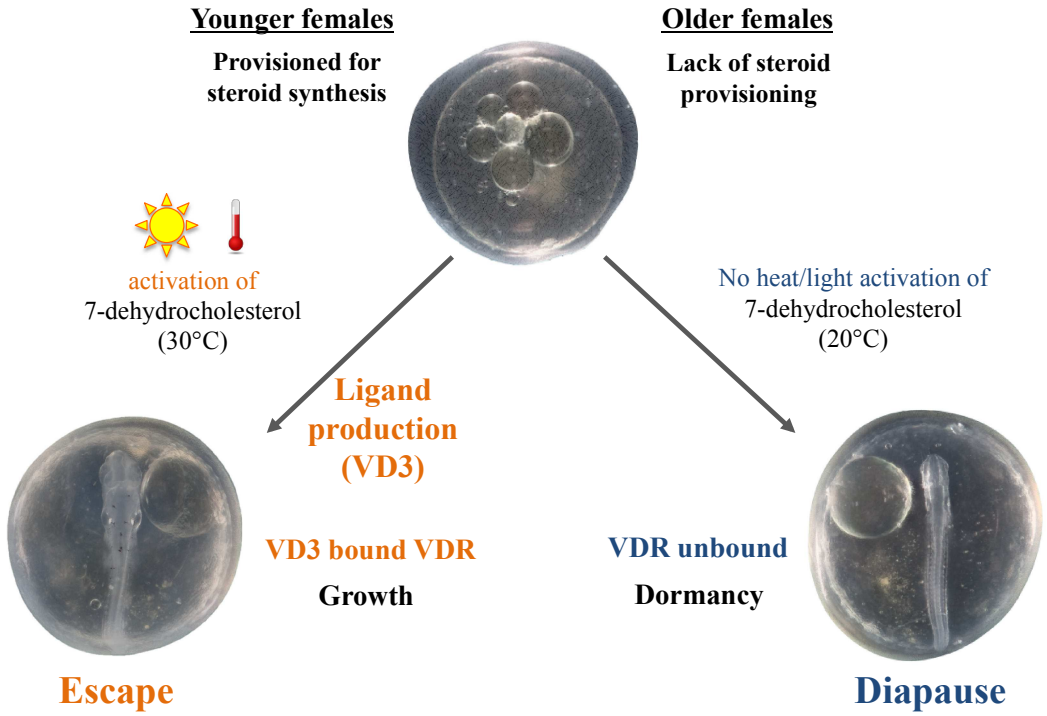


Figure 6.1: A model for developmental plasticity in *A. limnaeus*. Development in *A. limnaeus* spans two life stages whereby embryos are provisioned with steroidal cues for trajectory. Embryos respond to environmental temperatures and UV light that support the conversion of pre-vitamin D into the active form of vitamin D3 (VD3) that can bind to the vitamin D receptor (VDR). When bound, the receptor regulates the expression of genes involved in diverse biological functions that promote growth and development. In its unbound state, VDR activates gene expression patterns that promote dormancy and metabolic arrest associated with diapause.

REFERENCES

Aanes, H., C. L. Winata, C. H. Lin, J. P. Chen, K. G. Srinivasan, S. G. Lee, A. Y. Lim, H. S. Hajan, P. Collas and G. Bourque (2011). "Zebrafish mRNA sequencing deciphers novelties in transcriptome dynamics during maternal to zygotic transition." Genome research **21**(8): 1328-1338.

Agarwal, V., G. W. Bell, J.-W. Nam and D. P. Bartel (2015). "Predicting effective microRNA target sites in mammalian mRNAs." eLIFE **4**: e05005.

Allard, J., H. Kamei and C. Duan (2013). "Inducible transgenic expression in the short‐lived fish Nothobranchius furzeri." Journal of fish biology.

Altschul, S. F., W. Gish, W. Miller, E. W. Myers and D. J. Lipman (1990). "Basic local alignment search tool." Journal of molecular biology **215**(3): 403-410.

Anders, S., A. Reyes and W. Huber (2012). "Detecting differential usage of exons from RNA-seq data." Genome research **22**(10): 2008-2017.

Anderson, S., I. Kokay, T. Lang, D. Grattan and J. Curlewis (2003). "Quantification of prolactin-releasing peptide (PrRP) mRNA expression in specific brain regions of the rat during the oestrous cycle and in lactation." Brain research **973**(1): 64-73.

Anderson, S. N. and J. E. Podrabsky (2014). "The effects of hypoxia and temperature on metabolic aspects of embryonic development in the annual killifish *Austrofundulus limnaeus*." Journal of Comparative Physiology B **184**: 355-370.

Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data.

Antebi, A. (2006). "Nuclear hormone receptors in C. elegans." WormBook **3**: 1-13.

Antebi, A., W.-H. Yeh, D. Tait, E. M. Hedgecock and D. L. Riddle (2000). "Daf-12 encodes a nuclear receptor that regulates the dauer diapause and developmental age in *C. elegans*." Genes and Development **14**: 1512-1527.

Bai, J., C. Solberg, J. M. Fernandes and I. A. Johnston (2007). "Profiling of maternal and developmental-stage specific mRNA transcripts in Atlantic halibut *Hippoglossus hippoglossus*." Gene **386**(1-2): 202-210.

Baroux, C., D. Autran, C. Gillmor, D. Grimanelli and U. Grossniklaus (2008). "The maternal to zygotic transition in animals and plants." Cold Spring Harbor Sympia on Quantitative Biology: sqb-2008.

Barrey, E., G. Saint-Auret, B. Bonnamy, D. Damas, O. Boyer and X. Gidrol (2011). "Pre-microRNA and mature microRNA in human mitochondria." PLoS ONE **6**: e20220.

Bartel, D. P. (2004). "MicroRNAs: Genomics, biogenesis, mechanism, and function." Cell **116**(2): 281-297.

Bartel, D. P. (2009). "MicroRNAs: Target recognition and regulatory functions." Cell **136**(2): 215-233.

Baumeister, R., E. Schaffitzel and M. Hertweck (2006). "Endocrine signaling in Caenorhabditis elegans controls stress response and longevity." Journal of Endocrinology **190**(2): 191-202.

Bazzini, A. A., M. T. Lee and A. J. Giraldez (2012). "Ribosome profiling shows that miR-430 reduces translation before causing mRNA decay in zebrafish." Science **336**: 233-237.

Becker, B. (2014). Small Molecule Modulators of Dauer Formation and Longevity in Caenorhabditis Elegans, Universität zu Köln.

Beildeck, M. E., E. P. Gelmann and W. B. Stephen (2010). "Cross-regulation of signaling pathways: an example of nuclear hormone receptors and the canonical Wnt pathway." Experimental cell research **316**(11): 1763-1772.

Benvenuti, S. (1995). "Soil light penetration and dormancy of jimsonweed (Datura stramonium) seeds." Weed Science: 389-393.

Bethke, A., N. Fielenbach, Z. Wang, D. J. Mangelsdorf and A. Antebi (2009). "Nuclear hormone receptor regulation of microRNAs controls developmental progression." Science **324**(5923): 95-98.

Bikle, D. D. (2014). "Vitamin D metabolism, mechanism of action, and clinical applications." Chemistry & biology **21**(3): 319-329.

Bolger, A. A., M. Lohse and B. Usadel (2014). "Trimmomatic: A flexible trimmer for Illumina Sequence Data." Bioinformatics **30**(15): 2114-2120.

Bourc'his, D. and O. Voinnet (2010). "A small-RNA perspective on gametogenesis, fertilization, and early zygotic development." Science **330**(6004): 617-622.

Bowerman, B. (1998). "Maternal control of pattern formation in early *Caenorhabditis elegans* embryos." Current topics in developmental biology **39**: 73.

Boyer, L. A., M. R. Langer, K. A. Crowley, S. Tan, J. M. Denu and C. L. Peterson (2002). "Essential role for the SANT domain in the functioning of multiple chromatin remodeling enzymes." Molecular cell **10**(4): 935-942.

Brückner, K. and R. Klein (1998). "Signaling by Eph receptors and their ephrin ligands." Current opinion in neurobiology **8**(3): 375-382.

Camp, E., A. V. Sánchez‐Sánchez, A. García‐España, R. DeSalle, L. Odqvist, J. Enrique O'Connor and J. L. Mullor (2009). "Nanog regulates proliferation during early fish development." Stem Cells **27**(9): 2081-2091.

Carbon, S., A. Ireland, C. J. Mungall, S. Shu, B. Marshall and S. Lewis (2009). "AmiGO: online access to ontology and annotation data." Bioinformatics **25**(2): 288-289.

Carroll, S. B., J. K. Grenier and S. D. Weatherbee (2013). From DNA to diversity: molecular genetics and the evolution of animal design, John Wiley & Sons.

Chebotareva, N. A., V. F. Makeeva, S. G. Bazhina, T. B. Eronina, N. B. Gusev and B. I. Kurganov (2010). "Interaction of Hsp27 with native phosphorylase kinase under crowding conditions." Macromolecular bioscience **10**(7): 783-789.

Chen, J., D. L.-W. Kwong, C.-L. Zhu, L.-L. Chen, S.-S. Dong, L.-Y. Zhang, J. Tian, C.-B. Qi, T.-T. Cao and A. M. G. Wong (2012). "RBMS3 at 3p24 inhibits nasopharyngeal carcinoma development via inhibiting cell proliferation, angiogenesis, and inducing apoptosis." PloS one **7**(9): e44636.

Chen, P. Y., H. Manninga, K. Slanchev, M. Chien, J. J. Russo, J. Ju, R. Sheridan, B. John, D. S. Marks and D. Gaidatzis (2005). "The developmental miRNA profiles of zebrafish as determined by small RNA cloning." Genes & development **19**(11): 1288-1293.

Chennault, T. and J. E. Podrabsky (2010). "Aerobic and anaerobic capacities differ in embryos of the annual killifish *Austrofundulus limnaeus* that develop on alternate developmental trajectories." Journal of Experimental Zoology Part A: Ecological Genetics and Physiology **313A**(9): 587-596.

Cheong, A. W., R. T. Pang, W.-M. Liu, K. S. A. Kottawatta, K.-F. Lee and W. S. Yeung (2014). "MicroRNA Let-7a and dicer are important in the activation and implantation of delayed implanting mouse embryos." Human Reproduction: det462.

Chippindale, A. K., J. A. Alipaz, H.-W. Chen and M. R. Rose (1997). "Experimental evolution of accelerated development in *Drosophila*. 1. Developmental speed and larval survival." Evolution: 1536-1551.

Christakos, S., D. V. Ajibade, P. Dhawan, A. J. Fechner and L. J. Mady (2010). "Vitamin D: metabolism." Endocrinology and metabolism clinics of North America **39**(2): 243.
Comer, S. B. (2007). Cholecalciferol protects against deoxycholic acid-induced loss of EphB2 in human colorectal cancer cells, The University of Arizona.

Cotter, K. A., D. Nacci, D. Champlin, A. T. Yeo, T. D. Gilmore and G. V. Callard (2016). "Adaptive significance of Erα splice variants in killifish (*Fundulus heteroclitus*) resident in an estrogenic environment." Endocrinology **157**(6): 2294-2308.

Curtis, D., R. Lehmann and P. D. Zamore (1995). "Translational regulation in development." Cell **81**(2): 171-178.

Danzmann, R. G., M. M. Ferguson, F. W. Allendorf and K. L. Knudsen (1986). "Heterozygosity and developmental rate in a strain of rainbow trout (*Salmo gairdneri*)." Evolution: 86-93.

Denlinger, D. (2013). "Hormonal Control of Diapause." Endocrinology II **8**: 353.

Denlinger, D. L. (2002). "Regulation of Diapause." Annual Review of Entomology **47**(1): 93-122.

DiMichele, L. and D. A. Powers (1991). "Allozyme variation, developmental rate, and differential mortality in the teleost *Fundulus heteroclitus*." Physiological Zoology **64**(6): 1426-1443.

DiMichele, L. and M. E. Westerman (1997). "Geographic variation in developmental rate between populations of the teleost *Fundulus heteroclitus*." Mar.Biol. **128**(1): 1-7.

Dolfi, L., R. Ripa and A. Cellerino (2014). "Transition to annual life history coincides with reduction in cell cycle speed during early cleavage in three independent clades of annual killifish." EvoDevo **5**(1): 32.

Eisen, M., M. de Hoon, S. Imoto and S. Miyano (1998). Cluster. C Clustering Library version 1.52. Stanford, CA, Stanford University.

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

Emerson, K. J., W. E. Bradshaw and C. M. Holzapfel (2009). "Complications of complexity: integrating environmental, genetic and hormonal control of insect diapause." Trends in Genetics **25**(5): 217-225.

Feder, M. E., N. V. Cartaño, L. Milos, R. A. Krebs and S. L. Lindquist (1996). "Effect of engineering Hsp70 copy number on Hsp70 expression and tolerance of ecologically relevant heat shock in larvae and pupae of *Drosophila melanogaster*." Journal of experimental biology **199**(8): 1837-1844.

Fielenbach, N. and A. Antebi (2008). "*C. elegans* dauer formation and the molecular basis of plasticity." Genes and Development **22**(16): 2149-2165.

Flanagan, J. G. and P. Vanderhaeghen (1998). "The ephrins and Eph receptors in neural development." Annual review of neuroscience **21**(1): 309-345.

Frisén, J., J. Holmberg and M. Barbacid (1999). "Ephrins and their Eph receptors: multitalented directors of embryonic development." The EMBO journal **18**(19): 5159-5165.

Fujimoto, M., K.-i. Takeshita, X. Wang, I. Takabatake, Y. Fujisawa, H. Teranishi, M. Ohtani, Y. Muneoka and S. Ohta (1998). "Isolation and characterization of a novel bioactive peptide, Carassius Rfamide (C-Rfa), from the brain of the Japanese crucian carp." Biochemical and biophysical research communications **242**(2): 436-440.

Furness, A. I., K. Lee and D. N. Reznick (2015). "Adaptation in a variable environment: Phenotypic plasticity and bet‐hedging during egg diapause and hatching in an annual killifish." Evolution **69**: 1461-1475.

Galloway, A., A. Saveliev, S. Łukasiak, D. J. Hodson, D. Bolland, K. Balmanno, H. Ahlfors, E. Monzón-Casanova, S. C. Mannurita and L. S. Bell (2016). "RNA-binding proteins ZFP36L1 and ZFP36L2 promote cell quiescence." Science **352**(6284): 453-459.

García-López, J., L. Alonso, D. B. Cárdenas, H. Artaza-Alvarez, J. de Dios Hourcade, S. Martínez, M. A. Brieño-Enríquez and J. del Mazo (2015). "Diversity and functional convergence of small noncoding RNAs in male germ cell differentiation and fertilization." RNA.

Gentleman, R. C., V. J. Carey, D. M. Bates, B. Bolstad, M. Dettling, S. Dudoit, B. Ellis, L. Gautier, Y. Ge, J. Gentry, K. Hornik, T. Hothorn, W. Huber, S. Iacus, R. Irizarry, F. Leisch, C. Li, M. Maechler, A. J. Rossini, G. Sawitzki, C. Smith, G. Smyth, L. Tierney, J. Y. Yang and J. Zhang (2004). "Bioconductor: open software development for computational biology and bioinformatics." Genome Biol **5**(10): R80.

Gilbert, S. F. and D. Epel (2009). "Ecological developmental biology." Sinauer, Sunderland: 480.

Giraldez, A. J., R. M. Cinalli, M. E. Glasner, A. J. Enright, J. M. Thomson, S. Baskerville, S. M. Hammond, D. P. Bartel and A. F. Schier (2005). "MicroRNAs regulate brain morphogenesis in zebrafish." Science **308**(5723): 833-838.

Giraldez, A. J., Y. Mishima, J. Rihel, R. J. Grocock, S. Van Dongen, K. Inoue, A. J. Enright and A. F. Schier (2006). "Zebrafish MiR-430 Promotes Deadenylation and Clearance of Maternal mRNAs." Science **312**(5770): 75-79.

Gould, S. J. (1992). "Ontogeny and phylogeny--revisited and reunited." Bioessays **14**(4): 275-279.

Griffiths-Jones, S., R. J. Grocock, S. van Dongen, A. Bateman and A. J. Enright (2006). "miRBase: microRNA sequences, targets and gene nomenclature." Nucleic Acids Research **34**(suppl 1): D140-D144.

Guilgur, L. G., P. Prudêncio, D. Sobral, D. Liszekova, A. Rosa and R. G. Martinho (2014). "Requirement for highly efficient pre-mRNA splicing during *Drosophila* early embryonic development." Elife **3**: e02181.

Gumienny, T. L. and C. Savage-Dunn (2013). "TGF-β signaling in C. elegans." WormBook: the online review of C. elegans biology: 1.

Hammell, C. M., X. Karp and V. Ambros (2009). "A feedback circuit involving let-7-family miRNAs and DAF-12 integrates environmental signals and developmental timing in Caenorhabditis elegans." Proceedings of the National Academy of Sciences **106**(44): 18668-18673.

Hannon, G. J. (2002). "RNA interference." Nature **418**(6894): 244-251.
Harding, J. L., S. Horswell, C. Heliot, J. Armisen, N. M. Luscombe, L. B. Zimmerman, E. A. Miska and C. S. Hill (2013). "Small RNA profiling of *Xenopus* embryos reveals novel miRNAs and a new class of small RNAs derived from intronic transposable elements." Genome Research.

Harfe, B. D., M. T. McManus, J. H. Mansfield, E. Hornstein and C. J. Tabin (2005). "The RNaseIII enzyme Dicer is required for morphogenesis but not patterning of the vertebrate limb." Proceedings of the National Academy of Sciences of the United States of America **102**(31): 10898-10903.

Hartmann, N. and C. Englert (2012). "A microinjection protocol for the generation of transgenic killifish (Species: Nothobranchius furzeri)." Developmental Dynamics **241**(6): 1133-1141.

Harvey, S. A., I. Sealy, R. Kettleborough, F. Fenyes, R. White, D. Stemple and J. C. Smith (2013). "Identification of the zebrafish maternal and paternal transcriptomes." Development **140**(13): 2703-2710.

Haslbeck, M., T. Franzmann, D. Weinfurtner and J. Buchner (2005). "Some like it hot: the structure and function of small heat-shock proteins." Nature Structural & Molecular Biology **12**(10): 842-846.

Haverty, P. M., E. Lin, J. Tan, Y. Yu, B. Lam, S. Lianoglou, R. M. Neve, S. Martin, J. Settleman and R. L. Yauch (2016). "Reproducible pharmacogenomic profiling of cancer cell line panels." Nature **533**(7603): 333-337.

Hentschel, A., R. P. Zahedi and R. Ahrends (2016). "Protein lipid modifications—More than just a greasy ballast." Proteomics **16**(5): 759-782.

Heyn, P., M. Kircher, A. Dahl, J. Kelso, P. Tomancak, A. T. Kalinka and K. M. Neugebauer (2014). "The earliest transcribed zygotic genes are short, newly evolved, and different across species." Cell reports **6**(2): 285-292.

Hightower, L. E., C. E. Norris, P. J. Di Iorio and E. Fielding (1999). "Heat shock responses of closely related species of tropical and desert fish." American Zoologist **39**(6): 877-888.

Hinuma, S., Y. Habata, R. Fujii, Y. Kawamata, M. Hosoya, S. Fukusumi, C. Kitada, Y. Masuo, T. Asano and H. Matsumoto (1998). "A prolactin-releasing peptide in the brain." Nature **393**(6682): 272.

Holick, M. F., J. MacLaughlin, M. Clark, S. Holick, J. Potts, R. Anderson, I. Blank, J. Parrish and P. Elias (1980). "Photosynthesis of previtamin D3 in human skin and the physiologic consequences." Science **210**(4466): 203-205.

Holoch, D. and D. Moazed (2015). "RNA-mediated epigenetic regulation of gene expression." Nature Reviews Genetics.

Howarth, D. L., S. H. Law, B. Barnes, J. M. Hall, D. E. Hinton, L. Moore, J. M. Maglich, J. T. Moore and S. W. Kullman (2008). "Paralogous vitamin D receptors in teleosts: transition of nuclear receptor function." Endocrinology **149**(5): 2411-2422.

Hrbek, T., D. C. Taphorn and J. E. Thomerson (2005). "Molecular phylogeny of *Austrofundulus* Myers (Cyprinodontiformes: Rivulidae), with revision of the genus and the description of four new species." Zootaxa **825**: 1-39.

Hu, Y.-C., C.-K. Kang, C.-H. Tang and T.-H. Lee (2015). "Transcriptomic analysis of metabolic pathways in milkfish that respond to salinity and temperature changes." PloS ONE **10**(8): e0134959.

Hubbs, C. L. (1926). "The structural consequences of modifications of the developmental rate in fishes, considered in reference to certain problems of evolution." American Naturalist **60**(666): 57-81.

Irie, N. and S. Kuratani (2011). "Comparative transcriptome analysis reveals vertebrate phylotypic period during organogenesis." Nat Commun **2**: 248.

Jiajie, T., Y. Yanzhou, A. C. Hoi-Hung, C. Zi-Jiang and C. Wai-Yee (2017). "Conserved miR-10 family represses proliferation and induces apoptosis in ovarian granulosa cells." Scientific reports **7**.

Kane, D. A. and C. B. Kimmel (1993). "The zebrafish midblastula transition." Development **119**(2): 447-456.

Kanehisa, M., Y. Sato, M. Kawashima, M. Furumichi and M. Tanabe (2016). "KEGG as a reference resource for gene and protein annotation." Nucleic Acids Research **44**(D1): D457-462.

Karp, X., M. Hammell, M. C. Ow and V. Ambros (2011). "Effect of life history on microRNA expression during C. elegans development." RNA **17**(4): 639-651.

Kim, H.-J., J.-Y. Yi, H.-S. Sung, D. D. Moore, B. H. Jhun, Y. C. Lee and J. W. Lee (1999). "Activating signal cointegrator 1, a novel transcription coactivator of nuclear receptors, and its cytosolic localization under conditions of serum deprivation." Molecular and Cellular Biology **19**(9): 6323-6332.

Kimura, K. D., H. A. Tissenbaum, Y. Liu and G. Ruvkun (1997). "*daf-2*, an insulin receptor-like gene that regulates longevity and diapause in *Caenorhabditis elegans*." Science **277**: 942-946.

King, A. M. and T. H. MacRae (2012). "The small heat shock protein p26 aids development of encysting *Artemia* embryos, prevents spontaneous diapause termination and protects against stress." PloS One **7**(8): e43723.

King, A. M., J. Toxopeus and T. H. MacRae (2014). "Artemin, a diapause-specific chaperone, contributes to the stress tolerance of *Artemia franciscana* cysts and influences their release from females." Journal of Experimental Biology **217**(10): 1719-1724.

Krebs, R. A. and M. E. Feder (1998). "Hsp70 and larval thermotolerance in *Drosophila melanogaster*: how much is enough and when is more too much?" Journal of insect physiology **44**(11): 1091-1101.

Krone, P. H., Z. Lele and J. B. Sass (1997). "Heat shock genes and the heat shock response in zebrafish embryos." Biochemistry and Cell Biology **75**(5): 487-497.

Kultz, D. (2004). "Molecular and evolutionary basis of the cellular stress response." Annual Review of Physiology **67**: 225-257.

Kwon, H.-J. (2016). "Vitamin D receptor signaling is required for heart development in zebrafish embryo." Biochemical and biophysical research communications **470**(3): 575-578.

Lanes, C. F., T. T. Bizuayehu, J. M. de Oliveira Fernandes, V. Kiron and I. Babiak (2013). "Transcriptome of Atlantic cod (*Gadus morhua* L.) early embryos from farmed and wild broodstocks." Mar Biotechnol (NY) **15**(6): 677-694.

Langfelder, P. and S. Horvath (2008). "WGCNA: an R package for weighted correlation network analysis." BMC bioinformatics **9**(1): 559.

Langfelder, P., R. Luo, M. C. Oldham and S. Horvath (2011). "Is my network module preserved and reproducible?" PloS Comput Biol **7**(1): e1001057.

Langmead, B. and S. L. Salzberg (2012). "Fast gapped-read alignment with Bowtie 2." Nature methods **9**(4): 357-359.

Larriba, M. J., N. Valle, H. G. Pálmer, P. Ordóñez-Morán, S. Álvarez-Díaz, K.-F. Becker, C. Gamallo, A. G. de Herreros, J. M. González-Sancho and A. Muñoz (2007). "The inhibition of Wnt/β-catenin signalling by 1α, 25-dihydroxyvitamin D3 is abrogated by Snail1 in human colon cancer cells." Endocrine-related cancer **14**(1): 141-151.

Lee, J., Y. Lee, S.-Y. Na, D.-J. Jung and S.-K. Lee (2001). "Transcriptional coregulators of the nuclear receptor superfamily: coactivators and corepressors." Cellular and Molecular Life Sciences CMLS **58**(2): 289-297.

Lee, R. C. and V. Ambros (2001). "An extensive class of small RNAs in *Caenorhabditis elegans*." Science **294**(5543): 862-864.

Levels, P. J. and J. M. Denuce (1988). "Intrinsic variability in the frequency of embryonic diapauses of the annual fish *Nothobranchius korthausae*, regulated by light: dark cycle and temperature." Environmental Biology of Fishes **22**(3): 211-223.

Levraut, J., H. Iwase, Z. H. Shao, T. L. Vanden Hoek and P. T. Schumacker (2003). "Cell death during ischemia: relationship to mitochondrial depolarization and ROS generation." American Journal of Physiology - Heart and Circulatory Physiology **284**(2): H549-H558.

Li, L., P. Zheng and J. Dean (2010). "Maternal control of early mouse development." Development **137**(6): 859-870.

Li, X., D. Han, R. Kin Ting Kam, X. Guo, M. Chen, Y. Yang, H. Zhao and Y. Chen (2010). "Developmental expression of sideroflexin family genes in Xenopus embryos." Developmental Dynamics **239**(10): 2742-2747.

Liang, Y.-n., Y. Liu, Q. Meng, X. Li, F. Wang, G. Yao, L. Wang, S. Fu and D. Tong (2015). "RBMS3 is a tumor suppressor gene that acts as a favorable prognostic marker in lung squamous cell carcinoma." Medical Oncology **32**(2): 1-9.

Lin, C.-T., W.-C. Tseng, N.-W. Hsiao, H.-H. Chang and C.-F. Ken (2009). "Characterization, molecular modelling and developmental expression of zebrafish manganese superoxide dismutase." Fish & Shellfish Immunology **27**(2): 318-324.

Liu, Q., S. Sun, W. Yu, J. Jiang, F. Zhuo, G. Qiu, S. Xu and X. Jiang (2015). "Altered expression of long non-coding RNAs during genotoxic stress-induced cell death in human glioma cells." Journal of Neuro-oncology **122**(2): 283-292.

Liu, W.-M., R. T. Pang, A. W. Cheong, E. H. Ng, K. Lao, K.-F. Lee and W. S. Yeung (2012). "Involvement of microRNA lethal-7a in the regulation of embryo implantation in mice." PloS One **7**(5): e37039.

Love, M. I., W. Huber and S. Anders (2014). "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2." *Genome Biology* **15**: 550.

Love, M. I., W. Huber and S. Anders (2014). "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2." *Genome Biology* **15**.

Lund, A. H. (2010). "miR-10 in development and cancer." Cell Death and Differentiation **17**: 209-214.

Machado, B. E. and J. E. Podrabsky (2007). "Salinity tolerance in diapausing embryos of the annual killifish *Austrofundulus limnaeus* is supported by exceptionally low water and ion permeability." Journal of Comparative Physiology B: Biochemical, Systemic, and Environmental Physiology **177**(7): 809-820.

Maia, M., A. DeVriese, T. Janssens, M. Moons, R. J. Lories, J. Tavernier and E. M. Conway (2011). "CD248 facilitates tumor growth via its cytoplasmic domain." BMC cancer **11**(1): 162.

Mao, L. and E. Shelden (2006). "Developmentally regulated gene expression of the small heat shock protein Hsp27 in zebrafish embryos." Gene expression patterns **6**(2): 127-133.

Marien, D. (1958). "Selection for developmental rate in *Drosophila pseudoobscura*." Genetics **43**(1): 3.

Marion, R. M., A. Regev, E. Segal, Y. Barash, D. Koller, N. Friedman and E. K. O'Shea (2004). "Sfp1 is a stress-and nutrient-sensitive regulator of ribosomal protein gene expression." Proceedings of the National Academy of Sciences of the United States of America **101**(40): 14315-14322.

Markofsky, J. and J. R. Matias (1977). "Effects of light: dark cycles and temperature on embryonic diapause in the East African annual fish, *Nothobranchius guentheri*." Chronobiologia **4**: 130-131.

Markofsky, J. and J. R. Matias (1977). "The effects of temperature and season of collection on the onset and duration of diapause in embryos of the annual fish *Nothobranchius guentheri*." Journal of Experimental Zoology **202**: 49-56.

Markofsky, J., J. R. Matias, K. Inglima, J. H. Vogelman and N. Orentreich (1979). "The variable effects of ambient and artificial light: dark cycles on embryonic diapause in a laboratory population of the annual fish *Nothobranchius guentheri*." Journal of Experimental Biology **83**: 203-215.

Markov, G. V., J. Gutierrez-Mazariegos, D. Pitrat, I. M. Billas, F. Bonneton, D. Moras, J. Hasserodt, G. Lecointre and V. Laudet (2017). "Origin of an ancient hormone/receptor couple revealed by resurrection of an ancestral estrogen." Science Advances **3**(3): e1601778.

Matson, C. W., B. W. Clark, M. J. Jenny, C. R. Fleming, M. E. Hahn and R. T. Di Giulio (2008). "Development of morpholino gene knockdown technique in *Fundulus heteroclitus*: a tool for studying molecular mechanisms in an established environmental model." Aquatic Toxicology **87**: 289-295.

Mattick, J. S. and M. J. Gagen (2001). "The evolution of controlled multitasked gene networks: The role of introns and other noncoding RNAs in the development of complex organisms." Mol.Biol.Evol. **18**(9): 1611-1630.

McDonnell, D. P., D. J. Mangelsdorf, J. W. Pike, M. R. Haussler and B. W. O'Malley (1987). "Molecular cloning of complementary DNA encoding the avian receptor for vitamin D." Science **235**(4793): 1214-1217.

McIlhinney, R. J., K. Young, M. Egerton, R. Camble, A. White and M. Soloviev (1998). "Characterization of human and rat brain myristoyl-CoA: protein N-myristoyltransferase:

evidence for an alternative splice variant of the enzyme." Biochemical Journal **333**(3): 491-495.

Mi, H., S. Poudel, A. Muruganujan, J. T. Casagrande and P. D. Thomas (2016). "PANTHER version 10: expanded protein families and functions, and analysis tools." Nucleic Acids Research **44**(D1): D336-D342.

Mishima, Y., A. J. Giraldez, Y. Takeda, T. Fujiwara, H. Sakamoto, A. F. Schier and K. Inoue (2006). "Differential regulation of germline mRNAs in soma and germ cells by zebrafish miR-430." Curr Biol **16**: 2135-2142.

Mo, M., K. Yokawa, Y. Wan and F. Baluška (2015). "How and why do root apices sense light under the soil surface?" Frontiers in plant science **6**: 775.

Mo, X., E. Kowenz-Leutz, Y. Laumonnier, H. Xu and A. Leutz (2005). "Histone H3 tail positioning and acetylation by the c-Myb but not the v-Myb DNA-binding SANT domain." Genes & development **19**(20): 2447-2457.

Mohr, A. M. and J. L. Mott (2015). Overview of MicroRNA Biology. Seminars in liver disease.

Moriyama, S., T. Ito, A. Takahashi, M. Amano, S. A. Sower, T. Hirano, K. Yamamori and H. Kawauchi (2002). "A homolog of mammalian PRL-releasing peptide (fish arginyl-phenylalanyl-amide peptide) is a major hypothalamic peptide of PRL release in teleost fish." Endocrinology **143**(6): 2071-2079.

Morris, K. V. and J. S. Mattick (2014). "The rise of regulatory RNA." Nature Reviews Genetics **15**(6): 423-437.

Murphy, W. J. and G. E. Collier (1997). "A molecular phylogeny for Aplocheiloid fishes (Atherinomorpha, Cyprinodontiformes): the role of vicariance and the origins of annualism." Molecular Biology and Evolution **14**(8): 790-799.

Nawrocki, E. P., S. W. Burge, A. Bateman, J. Daub, R. Y. Eberhardt, S. R. Eddy, E. W. Floden, P. P. Gardner, T. A. Jones, J. Tate and R. D. Finn (2014). "RFAM 12.0: updates to the RNA familiies database." Nucleic Acids Research.

Norman, T. C. and A. W. Norman (1993). "Consideration of chemical mechanisms for the nonphotochemical production of vitamin D 3 in biological systems." Bioorganic & Medicinal Chemistry Letters **3**(9): 1785-1788.

Norris, C. E., M. A. Brown, E. Hickey, L. A. Weber and L. E. Hightower (1997). "Low-molecular-weight heat shock proteins in a desert fish (*Poeciliopsis lucida*): Homologs of human Hsp27 and *Xenopus* Hsp30." Mol.Biol.Evol. **14**(10): 1050-1061.

O'malley, M. A. and J. Dupré (2005). "Fundamental issues in systems biology." BioEssays **27**(12): 1270-1276.

Pang, K. C., S. Stephen, P. G. Engström, K. Tajul-Arifin, W. Chen, C. Wahlestedt, B. Lenhard, Y. Hayashizaki and J. S. Mattick (2005). "RNAdb--a comprehensive mammalian noncoding RNA database." Nucleic Acids Research **33**(Database issue): D125-130.

Paranjpe, S. S., U. G. Jacobi, S. J. van Heeringen and G. J. Veenstra (2013). "A genome-wide survey of maternal and embryonic transcripts during *Xenopus tropicalis* development." BMC Genomics **14**: 762.

Parenti, L. R. (1981). "A phylogenetic and biogeographic analysis of cyprinodontiform fishes (teleostei, Atherinomorpha)." Bull. Amer. Mus. Nat. Hist. **168**(4): 335-557.

Pasquale, E. B. (2010). "Eph receptors and ephrins in cancer: bidirectional signaling and beyond." Nature reviews. Cancer **10**(3): 165.

Pauli, A., E. Valen, M. F. Lin, M. Garber, N. L. Vastenhouw, J. Z. Levin, L. Fan, A. Sandelin, J. L. Rinn and A. Regev (2012). "Systematic identification of long noncoding RNAs expressed during zebrafish embryogenesis." Genome research **22**(3): 577-591.

Pelegri, F. (2003). "Maternal factors in zebrafish development." Developmental Dynamics **228**(3): 535-554.

Pereiro, L., F. Loosli, J. Fernández, S. Härtel, J. Wittbrodt and M. L. Concha (2017). "Gastrulation in an annual killifish: Molecular and cellular events during germ layer formation in Austrolebias." Developmental Dynamics.

Peshkin, L., M. Wühr, E. Pearl, W. Haas, R. M. Freeman, J. C. Gerhart, A. M. Klein, M. Horb, S. P. Gygi and M. W. Kirschner (2015). "On the relationship of protein and mRNA dynamics in vertebrate embryonic development." Developmental cell **35**(3): 383-394.

Podrabsky, J., C. Riggs, A. Romney, S. Woll, J. Wagner, K. Culpepper and T. Cleaver (2017). "Embryonic development of the annual killifish *Austrofundulus limnaeus*: An emerging model for ecological and evolutionary developmental biology research and instruction." Developmental Dynamics.

Podrabsky, J., C. Riggs and J. Wagner (2016). Tolerance of Environmental Stress. Annual Fishes. Life History Strategy, Diversity, and Evolution. N. Berois, G. García and R. De Sá. Boca Raton, FL USA, CRC Press, Taylor & Francis**:** 159-184.

Podrabsky, J., A. Romney and K. Culpepper (2016). Alternative Developmental Pathways. Annual Fishes. Life History Strategy, Diversity, and Evolution. N. Berois, G. García and R. De Sá. Boca Raton, FL USA, CRC Press, Taylor & Francis**:** 63-73.

Podrabsky, J. E. (1999). "Husbandry of the annual killifish *Austrofundulus limnaeus* with special emphasis on the collection and rearing of embryos." Environmental Biology of Fishes **54**(4): 421-431.

Podrabsky, J. E., I. D. F. Garrett and Z. F. Kohl (2010). "Alternative developmental pathways associated with diapause regulated by temperature and maternal influences in embryos of the annual killifish *Austrofundulus limnaeus*." Journal of Experimental Biology **213**(19): 3280-3288.

Podrabsky, J. E. and S. C. Hand (1999). "The bioenergetics of embryonic diapause in an annual killifish, *Austrofundulus limnaeus*." Journal of Experimental Biology **202**(19): 2567-2580.

Podrabsky, J. E. and S. C. Hand (2000). "Depression of protein synthesis during diapause in embryos of the annual killifish *Austrofundulus limnaeus*." Physiological and Biochemical Zoology **73**(6): 799-808.

Podrabsky, J. E., T. Hrbek and S. C. Hand (1998). "Physical and chemical characteristics of ephemeral pond habitats in the Maracaibo basin and Llanos region of Venezuela." Hydrobiologia **362**(1-3): 67-78.

Podrabsky, J. E., J. P. Lopez, T. W. M. Fan, R. Higashi and G. N. Somero (2007). "Extreme anoxia tolerance in embryos of the annual killifish *Austrofundulus limnaeus*: Insights from a metabolomics analysis." Journal of Experimental Biology **210**(13): 2253-2266.

Podrabsky, J. E., C. L. Riggs and J. M. Duerr (2012). Anoxia Tolerance During Vertebrate Development - Insights from Studies on the Annual Killifish *Austrofundulus limnaeus*. Anoxia. P. Padilla, InTech**:** 3-24.

Podrabsky, J. E. and G. N. Somero (2003). "Changes in gene expression associated associated with acclimation to constant temperatures and fluctuating daily temperatures in an annual killifish *Austrofundulus limnaeus*." Journal of Experimental Biology **207**(12): 2237-2254.

Podrabsky, J. E., A. Tingaud-Sequeira and J. Cerdà (2010). Metabolic dormancy and responses to environmental desiccation in fish embryos. Dormancy and resistance in harsh environments. E. Lubzens, J. Cerda and M. Clark, Springer**:** 203-226.

Pompella, A., A. Visvikis, A. Paolicchi, V. De Tata and A. F. Casini (2003). "The changing faces of glutathione, a cellular protagonist." Biochemical pharmacology **66**(8): 1499-1503.

Pons, T. L. and M. Fenner (2000). "Seed responses to light." Seeds: the ecology of regeneration in plant communities **2**: 237-260.

Prasad, N., M. Shakarad, V. Gohil, V. Sheeba, M. Rajamani and A. Joshi (2000). "Evolution of reduced pre-adult viability and larval growth rate in laboratory populations of *Drosophila melanogaster* selected for shorter development time." Genetical Research **76**(03): 249-259.

Pri-Tal, B. M. (2010). Genomic and Hormonal Components of Altered Developmental Pathways in the Annual Killifish, Austrofundulus limnaeus. M.S., Portland State University.

Pri-Tal, B. M., S. Blue, F. K. Y. Pau and J. E. Podrabsky (2011). "Hormonal components of altered developmental pathways in the annual killifish, *Austrofundulus limnaeus*." General and Comparative Endocrinology **174**(2): 166-174.

Queitsch, C., T. A. Sangster and S. Lindquist (2002). "Hsp90 as a capacitor of phenotypic variation." Nature **417**: 618-624.

Ramagopalan, S. V., A. Heger, A. J. Berlanga, N. J. Maugeri, M. R. Lincoln, A. Burrell, L. Handunnetthi, A. E. Handel, G. Disanto and S.-M. Orton (2010). "A ChIP-seq defined genome-wide map of vitamin D receptor binding: associations with disease and evolution." Genome research **20**(10): 1352-1360.

Reynolds, J. A., J. T. Peyton and D. L. Denlinger (2017). "Changes in microRNA abundance may regulate diapause in the flesh fly, Sarcophaga bullata." Insect Biochemistry and Molecular Biology.

Reynolds, J. A., J. T. Peyton and D. L. Denlinger (2017). "Changes in microRNA abundance may regulate diapause in the flesh fly, Sarcophaga bullata." Insect Biochemistry and Molecular Biology **84**: 1-14.

Rioux, V., E. Beauchamp, F. Pedrono, S. Daval, D. Molle, D. Catheline and P. Legrand (2006). "Identification and characterization of recombinant and native rat myristoyl-CoA: protein N-myristoyltransferases." Molecular and cellular biochemistry **286**(1): 161-170.

Rissone, A., M. Monopoli, M. Beltrame, F. Bussolino, F. Cotelli and M. Arese (2007). "Comparative genome analysis of the neurexin gene family in *Danio rerio*: insights into their functions and evolution." Molecular biology and evolution **24**(1): 236-252.

Romney, A. and J. Podrabsky (2017). "Transcriptomic analysis of maternally provisioned cues for phenotypic plasticity in the annual killifish, *Austrofundulus limnaeus*." EvoDevo **8**: 6.

Rottiers, V., D. L. Motola, B. Gerisch, C. L. Cummins, K. Nishiwaki, D. J. Mangelsdorf and A. Antebi (2006). "Hormonal control of C. elegans dauer formation and life span by a Rieske-like oxygenase." Developmental cell **10**(4): 473-482.

Rougvie, A. E. and E. G. Moss (2013). "Developmental transitions in *C. elegans* larval stages." Current Topics in Developmental Biology **105**: 153-180.

Rupik, W., K. Jasik, J. Bembenek and W. Widłak (2011). "The expression patterns of heat shock genes and proteins and their role during vertebrate's development." Comparative Biochemistry and Physiology Part A **159**(4): 349-366.

Rutherford, S. and S. Lindquist (1998). "Hsp90 as a capacitor for morphological evolution." Nature **396**(6709): 336-342.

Sako, D., A. V. Grinberg, J. Liu, M. V. Davies, R. Castonguay, S. Maniatis, A. J. Andreucci, E. G. Pobre, K. N. Tomkinson and T. E. Monnell (2010). "Characterization of the ligand binding functionality of the extracellular domain of activin receptor type Iib." Journal of Biological Chemistry **285**(27): 21037-21048.

Saldanha, A. (2003). Java TreeView, Source Forage.

Saunders, D., D. Richard, S. Applebaum, M. Ma and L. Gilbert (1990). "Photoperiodic diapause in Drosophila melanogaster involves a block to the juvenile hormone regulation of ovarian maturation." General and comparative endocrinology **79**(2): 174-184.

Schier, A. F. and M. M. Shen (2000). "Nodal signalling in vertebrate development." Nature **403**(6768): 385.

Seale, A., T. Itoh, S. Moriyama, A. Takahashi, H. Kawauchi, T. Sakamoto, M. Fujimoto, L. Riley, T. Hirano and E. Grau (2002). "Isolation and characterization of a homologue of mammalian prolactin-releasing peptide from the tilapia brain and its effect on prolactin release from the tilapia pituitary." General and comparative endocrinology **125**(3): 328-339.

Shannon, P., A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski and T. Ideker (2003). "Cytoscape: a software environment for integrated models of biomolecular interaction networks." Genome research **13**(11): 2498-2504.

Shen-Orr, S. S., Y. Pilpel and C. P. Hunter (2010). "Composition and regulation of maternal and zygotic transcriptomes reflects species-specific reproductive mode." Genome biology **11**(6): 1.

Sim, C. and D. Denlinger (2013). "Insulin signaling and the regulation of insect diapause." Frontiers in Physiology **4**(189): 1-10.

Sim, C., D. S. Kang, S. Kim, X. Bai and D. L. Denlinger (2015). "Identification of FOXO targets that generate diverse features of the diapause phenotype in the mosquito Culex pipiens." Proceedings of the National Academy of Sciences of the United States of America **112**(12): 3811-3816.

Strathmann, R. R., J. M. Staver and J. R. Hoffman (2002). "Risk and the evolution of cell-cycle durations of embryos." Evolution **56**(4): 708-720.

Tachibana, T. and T. Sakamoto (2014). "Functions of two distinct "prolactin-releasing peptides" evolved from a common ancestral gene." Frontiers in endocrinology **5**.

Tadros, W. and H. D. Lipshitz (2009). "The maternal-to-zygotic transition: a play in two acts." Development **136**(18): 3033-3042.

Tanzer, A., C. T. Amemiya, C. B. Kim and P. F. Stadler (2005). "Evolution of microRNAs located within Hox gene clusters." Journal of Experimental Zoology Part B: Molecular and Developmental Evolution **304**(1): 75-85.

Taphorn, D. C. and J. E. Thomerson (1978). "A revision of the South American cyprinodont fishes of the genera *Rachovia* and *Austrofundulus,* with the description of a new genus." Acta Biologica Venezuelica **9**(4): 377-452.

Tester, M. and C. Morris (1987). "The penetration of light through soil." Plant, Cell Environ. **10**(4): 281-286.

Thatcher, E. J., A. S. Flynt, N. Li, J. R. Patton and J. G. Patton (2007). "miRNA expression analysis during normal zebrafish development and following inhibition of the Hedgehog and Notch signaling pathways." Developmental Dynamics **236**(8): 2172-2180.

Thierry-Mieg, D. and J. Thierry-Mieg (2006). "AceView: a comprehensive cDNA-supported gene and transcripts annotation." Genome biology **7**(1): S12.

Tian, C., D. Sen, H. Shi, M. L. Foehr, Y. Plavskin, O. K. Vatamaniuk and J. Liu (2010). "The RGM protein DRAG-1 positively regulates a BMP-like signaling pathway in Caenorhabditis elegans." Development **137**(14): 2375-2384.

Tian, C., H. Shi, S. Xiong, F. Hu, W.-C. Xiong and J. Liu (2013). "The neogenin/DCC homolog UNC-40 promotes BMP signaling via the RGM protein DRAG-1 in C. elegans." Development **140**(19): 4070-4080.

Tingaud-Sequeira, A., J.-J. Lozano, C. Zapater, D. Otero, M. Kube, R. Reinhardt and J. Cerdà (2013). "A rapid transcriptome response is associated with desiccation resistance in aerially-exposed killifish embryos." PloS one **8**(5): e64410.

Tomkowicz, B., K. Rybinski, D. Sebeck, P. Sass, N. C. Nicolaides, L. Grasso and Y. Zhou (2010). "Endosialin/TEM-1/CD248 regulates pericyte proliferation through PDGF receptor signaling." Cancer biology & therapy **9**(11): 908-915.

Van De Wetering, M., E. Sancho, C. Verweij, W. De Lau, I. Oving, A. Hurlstone, K. Van Der Horn, E. Batlle, D. Coudreuse and A.-P. Haramis (2002). "The β-catenin/TCF-4 complex imposes a crypt progenitor phenotype on colorectal cancer cells." Cell **111**(2): 241-250.

Vanden Hoek, T. L., L. B. Becker, Z. Shao, C. Li and P. T. Schumacker (1998). "Reactive oxygen species released from mitochondria during brief hypoxia induce preconditioning in cardiomyocytes." Journal of Biological Chemistry **273**(29): 18092-18098.

Vesterlund, L., H. Jiao, P. Unneberg, O. Hovatta and J. Kere (2011). "The zebrafish transcriptome during early development." BMC Developmental Biology **11**(1): 1.

Vincent, J. B., E. Petek, S. Thevarkunnel, D. Kolozsvari, J. Cheung, M. Patel and S. W. Scherer (2002). "The RAY1/ST7 tumor-suppressor locus on chromosome 7q31 represents a complex multi-transcript system." Genomics **80**(3): 283-294.

Wagner, J., W. Warren, P. Minx and J. Podrabsky (2015). *Austrofundulus limnaeus* 1.0 draft genome assembly with annotation. National Center for Biotechnology Information.

Wagner, J. T. and J. E. Podrabsky (2015). "Extreme tolerance and developmental buffering of UV-C induced DNA damage in embryos of the annual killifish *Austrofundulus limnaeus*." Journal of Experimental Zoology **323A**: 10-30.

Wagner, J. T. and J. E. Podrabsky (2015). "Gene expression patterns that support novel developmental stress buffering in embryos of the annual killifish *Austrofundulus limnaeus*." EvoDevo **6**: 2.

Wang, J. and S. Kim (2003). "Global analysis of dauer gene expression in *Caenorhabditis elegans*." Development **130**: 1621-1634.

Wei, C., L. Salichos, C. M. Wittgrove, A. Rokas and J. G. Patton (2012). "Transcriptome-wide analysis of small RNA expression in early zebrafish development." RNA **18**: 915-929.

Weitzman, S. H. and J. P. Wourms (1967). "South American cyprinodont fishes allied to *Cynolebias* with the description of a new species of *Austrofundulus* from Venezuela." Copeia: 89-100.

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 1): 6543-6549.

Weston, S. A., R. Camble, J. Colls, G. Rosenbrock, I. Taylor, M. Egerton, A. D. Tucker, A. Tunnicliffe, A. Mistry and F. Mancia (1998). "Crystal structure of the anti-fungal target N-myristoyl transferase." Nature Structural & Molecular Biology **5**(3): 213-221.

White, J. A. and J. Heasman (2008). "Maternal control of pattern formation in *Xenopus laevis*." Journal of Experimental Zoology Part B: Molecular and Developmental Evolution **310**(1): 73-84.

Wienholds, E., W. P. Kloosterman, E. Miska, E. Alvarez-Saavedra, E. Berezikov, E. de Bruijn, H. R. Horvitz, S. Kauppinen and R. H. A. Plasterk (2005). "MicroRNA expression in zebrafish embryonic development." Science **309**(5732): 310-311.

Woll, S. and J. Podrabsky (2017). "Insulin-like growth factor signaling regulates developmental trajectory associated with diapause in embryos of the annual killifish *Austrofundulus limnaeus*." Journal of Experimental Biology.

Woll, S. C. (2016). Insulin-like Growth Factor Pathway Described in *Austrofundulus limnaeus* Diapause and Escape Embryos, Portland State University.

Wolter, J. M., H. H. T. Le, A. Linse, V. A. Godlove, T.-D. Nguyen, K. Kotagama, A. Lynch, A. Rawls and M. Mangone (2017). "Evolutionary patterns of metazoan microRNAs reveal targeting principles in the let-7 and miR-10 families." Genome research **27**(1): 53-63.

Woltering, J. M. and A. J. Durston (2008). "MiR-10 represses HoxB1a and HoxB3a in zebrafish." PloS One **3**(1): e1396.

Wourms, J. P. (1972). "The developmental biology of annual fish II. Naturally occuring dispersion and reaggregation of blastomeres during the development of annual fish eggs." Journal of Experimental Zoology **182**: 169-200.

Wourms, J. P. (1972). "The developmental biology of annual fishes I. Stages in the normal development of *Austrofundulus myersi* Dahl." Journal of Experimental Zoology **182**(2): 143-168.

Wourms, J. P. (1972). "The developmental biology of annual fishes III. Pre-embryonic and embryonic diapause of variable duration in the eggs of annual fishes." Journal of Experimental Zoology **182**(3): 389-414.

Xu, Q. (1999). "Microinjection into zebrafish embryos." Molecular Methods in Developmental Biology: Xenopus and Zebrafish: 125-132.

Xu, Z. and D. Norris (1998). "The SFP1 gene product of *Saccharomyces cerevisiae* regulates G2/M transitions during the mitotic cell cycle and DNA-damage response." Genetics **150**(4): 1419-1428.

Yanai, I., L. Peshkin, P. Jorgensen and M. W. Kirschner (2011). "Mapping gene expression in two Xenopus species: evolutionary constraints and developmental flexibility." Developmental cell **20**(4): 483-496.

Yang, S. H., A. Shrivastav, C. Kosinski, R. K. Sharma, M.-H. Chen, L. G. Berthiaume, L. L. Peters, P.-T. Chuang, S. G. Young and M. O. Bergo (2005). "N-myristoyltransferase 1 is essential in early mouse development." Journal of Biological Chemistry **280**(19): 18990-18995.

Yeyati, P. L., R. M. Bancewicz, J. Maule and V. van Heyningen (2007). "Hsp90 selectively modulates phenotype in vertebrate development." PloS Genetics **3**(3): 0431-0447.

Zeng, Z., C. R. Sharpe, J. P. Simons and D. C. Górecki (2003). "The expression and alternative splicing of alpha-neurexins during *Xenopus* development." International Journal of Developmental Biology **50**(1): 39-46.

Zhang, L., L. Prak, V. Rayon-Estrada, P. Thiru, J. Flygare, B. Lim and H. F. Lodish (2013). "ZFP36L2 is required for self-renewal of early burst-forming unit erythroid progenitors." Nature **499**(7456): 92-96.

Zhao, Y. and D. Srivastava (2007). "A developmental view of microRNA function." Trends in biochemical sciences **32**(4): 189-197.

Ørngreen, M., H. Schelhaas, T. Jeppesen, H. Akman, R. Wevers, S. Andersen, H. Ter Laak, O. van Diggelen, S. DiMauro and J. Vissing (2008). "Is muscle glycogenolysis impaired in X-linked phosphorylase b kinase deficiency?" Neurology **70**(20): 1876-1882.

APPENDIX A:

Tissue sampling and RNA-sequencing from diapause- and escape-bound embryos

A.1: Biological information for embryos and RNA extractions

Sample ID[1]	NCBI Biosample	Date of spawn[2]	# monitored[3]	Diapause[4]	Escape[4]	# embryos for RNA[5]	RNA (ng/uL)	A_{260}/A_{280}
Escape-1 (A5)	SAMN03610089	5/14/12	11	0%	100%	14	101.6	2.04
		5/10/12	10	0%	100%	18		
		5/7/12	14	0%	100%	20		
Escape-2 (A6)	SAMN03610090	5/7/12	13	0%	100%	47	160.1	1.99
Escape-3 (A11)	SAMN03610086	5/28/12	12	0%	100%	25	122.8	2.11
		5/24/12	8	0%	100%	11		
Escape-4 (A12)	SAMN03610087	5/7/12	12	0%	100%	65	77	1.99
Escape-5 (A14)	SAMN03610088	5/21/12	19	0%	100%	50	128.6	2.07
Escape-6 (B6)	SAMN03610091	2/20/12	6	0%	100%	50	112.6	2.02
Diapause-1 (B17)	SAMN03610096	5/28/12	11	100%	0%	40	82.9	2.16
		5/24/12	9	100%	0%	20		
Diapause-2 (B5)	SAMN03610099	5/7/12	13	100%	0%	36	274.6	2.16
Diapause-3 (A16)	SAMN03610093	5/14/12	24	100%	0%	65	132.6	1.94
		5/7/12	10	100%	0%	63		
Diapause-4 (A9)	SAMN03610095	5/14/12	19	100%	0%	31	87.2	1.8
Diapause-5 (B3)	SAMN03610097	2/23/12	10	100%	0%	45	163.4	2.03
Diapause-6 (B5)	SAMN03610098	2/23/12	9	100%	0%	50	140.2	2.04

1. Each sample represents embryonic output from a single female designated by the letter A/B (fish rack system) and a number that corresponds to their location in the rack system.
2. In some cases multiple spawning events from a single pair of fish were used to accumulate enough embryos for RNA extraction. Each sample replicate represents contains RNA from 30 or more embryos from the same female + male pairing.
3. Number of embryos allowed to develop at 25°C and monitored for developmental trajectory
4. Percent of the embryos listed in in third column that developed along the diapause-II or escape developmental trajectories.
5. The number of embryos that were immediately frozen in the 1-2 cell stage and later used for RNA extraction.

A.2: Poly-A Sequencing library trimmings statistics.

Sample ID	NCBI Accession ID	Total Read pairs	Both surviving, %[1]		Forward only surviving, %[2]		Reverse only surviving, %[2]		Dropped, %[3]	
Escape-1 (A5)	SRX1032652	52,146,021	39,468,715	76	1,399,106	3	1,810,775	4	9,467,425	18
Escape-2 (A6)	SRX1032653	49,400,021	40,786,557	83	2,314,544	5	531,788	1	5,767,132	12
Escape-3 (A11)	SRX1032647	32,741,329	27,448,685	84	334,534	1	184,505	1	4,773,605	15
Escape-4 (A12)	SRX1032648	13,890,504	12,355,081	89	215,134	2	180,802	1	1,139,487	8
Escape-5 (A14)	SRX1032649	40,230,917	38,629,999	96	511,463	1	238,057	1	851,398	2
Escape-6 (B6)	SRX1032662	18,053,153	17,139,665	95	140,256	1	78,367	0	694,865	4
Diapause-1 (B17)	SRX1032656	116,817,021	100,919,239	86	1,925,383	2	396,354	0	13,576,045	12
Diapause-2 (B5)	SRX1032661	33,229,842	31,767,178	96	464,805	1	87,517	0	910,342	3
Diapause-3 (A16)	SRX1032650	84,897,334	83,277,026	98	1,141,882	1	179,666	0	298,760	0
Diapause-4 (A9)	SRX1032654	30,408,053	27,922,285	93	480,919	2	81,349	0	1,868,000	6
Diapause-5 (B3)	SRX1032659	31,566,724	19,697,874	62	423,626	1	128,451	0	11,316,773	36
Diapause-6 (B5)	SRX1032660	34,363,642	33,289,793	97	507,344	2	129,694	0	436,811	1

1. The number and percentage of paired reads for each library.
2. The number and percentage of non-paired, or orphaned, reads specific to orientation that survived the trimming criteria for each library.
3. Total number and percentage of reads (paired and non-paired, or orphaned) that did not survive trimming criteria.

A.3: Alignment statistics for poly-A sequence libraries.

Sample ID	Paired[1]		Single[2]	
Escape-1 (A5)	39,468,715	90.08%	3,209,881	84.87%
Escape-2 (A6)	40,786,557	91.71%	2,846,332	84.82%
Escape-3 (A11)	27,448,685	88.80%	519,039	84.05%
Escape-4 (A12)	12,355,081	94.64%	395,936	88.70%
Escape-5 (A14)	38,629,999	96.55%	749,520	89.98%
Escape-6 (B6)	17,139,665	89.07%	218,623	87.48%
Diapause-1 (B17)	100,919,239	94.67%	2,321,737	87.50%
Diapause-2 (B5)	31,767,178	89.61%	552,322	82.41%
Diapause-3 (A16)	83,277,026	95.77%	1,321,548	88.89%
Diapause-4 (A9)	27,977,785	94.12%	562,268	87.36%
Diapause-5 (B3)	19,697,874	88.26%	552,077	84.88%
Diapause-6 (B5)	33,289,793	93.95%	637,038	87.36%

1. The number of paired reads for each library that aligned to the *A. limnaeus* genome.
2. The number of non-paired, or orphaned, reads specific to orientation that aligned to the *A. limnaeus* genome.

A.4: Small RNA library trimming and mapping statistics

Sample ID	NCBI accession ID	Sequenced reads[1]	Trimmed and filtered reads	Reads aligned to genome[2]	Unique sequence tags[3]	Aligned to genome[4]	Tags shared among replicates
Escape-1 (A5)	SRR5256769	963,329	420,854	333,623	23,469	12,232	
Escape-2 (A6)	SRR5256768	1,293,243	681,962	561,067	27,068	15,925	
Escape-3 (A11)	SRR5256772	1,668,791	764,311	650,372	27,267	13,217	
Escape-4 (A12)	SRR5256771	4,553,952	1,227,159	688,596	60,263	18,523	
Escape-5 (A14)	SRR5256770	1,199,063	498,329	353,536	27,068	11,860	
Escape-6 (B6)	SRR5256767	1,174,827	471,680	360,933	31,485	11,977	1,465
Diapause-1 (B17)	SRR5256764	1,409,374	680,588	579,077	32,534	19,386	
Diapause-2 (B5)	SRR5256761	460,003	256,053	221,309	12,707	5,828	
Diapause-3 (A16)	SRR5256766	1,374,883	706,717	606,802	23,439	10,537	
Diapause-4 (A9)	SRR5256765	2,794,962	1,527,871	1,350,663	37,967	20,512	
Diapause-5 (B3)	SRR5256763	512,292	266,166	224,567	16,458	9,042	
Diapause-6 (B5)	SRR5256762	2,210,931	961,165	821,462	32,098	16,984	1,185

1. The number of total reads and percentage of the total reads for each library that aligned to the *A. limnaeus* genome with 0 mismatches.
2. The total number of reads that aligned to the *A. limnaeus* genome with 0 mismatches.
3. Total number of unique small RNA sequences identified in the library
4. The total number of unique RNA sequences that aligned to the *A. limnaeus* genome with 0 mismatches.

A.5: GO analysis of transcriptome of *A. limnaeus* in 1-2 cell stage embryos.

Analysis Type: PANTHER Overrepresentation Test (release 20160715)
Annotation Version and Release Date: PANTHER version 11.1 Released 2016-10-24
Analyzed List: *A. limnaeus* top 500 most abundant transcripts with homologs to *D. rerio*
Reference List: *D. rerio* (all genes in database)
Bonferroni correction: TRUE
Displaying only results with $p < 0.05$

PANTHER GO-Slim Molecular Function	Danio rerio (REF) #	Alim_Top500 #	expected	Fold Enrichment	+/-	*p* value
structural constituent of ribosome	201	42	2.98	14.1	+	8.22E-32
structural molecule activity	1,197	66	17.74	3.72	+	1.05E-17
RNA binding	380	28	5.63	4.97	+	1.44E-09
Unclassified	12,972	153	192.29	0.8	-	0.00E+00
receptor activity	1,894	4	28.08	< 0.2	-	1.58E-06

PANTHER GO-Slim Biological Process	Danio rerio (REF) #	Alim_Top500 #	expected	Fold Enrichment	+/-	*p* value
oxidative phosphorylation	53	8	0.79	10.18	+	4.00E-04
generation of precursor metabolites and energy	241	17	3.57	4.76	+	4.70E-05
metabolic process	8,627	173	127.88	1.35	+	3.70E-04
rRNA metabolic process	114	12	1.69	7.1	+	5.01E-05
primary metabolic process	6,999	151	103.75	1.46	+	3.51E-05
translation	348	34	5.16	6.59	+	3.20E-15
protein metabolic process	2,373	75	35.18	2.13	+	9.34E-08
nucleobase-containing compound transport	82	8	1.22	6.58	+	9.07E-03
cellular component biogenesis	517	45	7.66	5.87	+	1.18E-18
cellular component organization or biogenesis	1,936	74	28.7	2.58	+	1.78E-11
chromosome segregation	121	10	1.79	5.58	+	4.12E-03
cellular process	9,764	181	144.73	1.25	+	2.94E-02
respiratory electron transport chain	157	12	2.33	5.16	+	1.32E-03
mitosis	467	24	6.92	3.47	+	5.28E-05
cell cycle	1,071	45	15.88	2.83	+	1.37E-07
organelle organization	804	32	11.92	2.69	+	1.53E-04
cellular component organization	1,797	56	26.64	2.1	+	3.48E-05
biosynthetic process	1,573	49	23.32	2.1	+	2.31E-04
Unclassified	11,501	135	170.48	0.79	-	0.00E+00
signal transduction	3,057	24	45.31	0.53	-	4.75E-02
neurological system process	1,219	4	18.07	0.22	-	1.45E-02
single-multicellular organism process	2,029	10	30.08	0.33	-	2.91E-03
multicellular organismal process	2,029	10	30.08	0.33	-	2.91E-03

immune system process	1,929	6	28.59	0.21	-	4.36E-05
nervous system development	947	2	14.04	< 0.2	-	1.78E-02
system development	1,472	4	21.82	< 0.2	-	5.96E-04
developmental process	2,991	15	44.34	0.34	-	2.29E-05

PANTHER GO-Slim Cellular Component	**Danio rerio (REF) #**	**Alim_Top500 #**	**expected**	**Fold Enrichment**	**+/-**	***p* value**
ribosome	104	38	1.54	24.65	+	6.74E-38
ribonucleoprotein complex	325	47	4.82	9.76	+	2.67E-29
macromolecular complex	1,574	80	23.33	3.43	+	2.89E-20
tubulin complex	24	6	0.36	16.87	+	1.21E-04
protein complex	1,279	37	18.96	1.95	+	6.17E-03
cytosol	380	47	5.63	8.34	+	2.04E-26
cytoplasm	2,859	85	42.38	2.01	+	2.44E-08
intracellular	5,122	138	75.92	1.82	+	1.15E-11
cell part	5,373	140	79.65	1.76	+	9.21E-11
microtubule	146	10	2.16	4.62	+	4.95E-03
cytoskeleton	750	24	11.12	2.16	+	2.61E-02
organelle	2,947	101	43.68	2.31	+	4.56E-14
Unclassified	19,495	242	288.98	0.84	-	0.00E+00

PANTHER Protein Class	**Danio rerio (REF) #**	**Alim_Top500 #**	**expected**	**Fold Enrichment**	**+/-**	***p* value**
tubulin	24	6	0.36	16.87	+	4.08E-04
microtubule family cytoskeletal protein	259	16	3.84	4.17	+	5.02E-04
cytoskeletal protein	1,039	36	15.4	2.34	+	6.19E-04
ribosomal protein	221	42	3.28	12.82	+	3.92E-30
RNA binding protein	910	71	13.49	5.26	+	8.39E-28
nucleic acid binding	2,927	89	43.39	2.05	+	8.87E-09
ribonucleoprotein	53	6	0.79	7.64	+	3.30E-02
oxidoreductase	720	24	10.67	2.25	+	4.96E-02
Unclassified	12,481	137	185.01	0.74	-	0.00E+00
receptor	2291	7	33.96	0.21	-	1.77E-06

PANTHER Pathways	**Danio rerio (REF) #**	**Alim_Top500 #**	**expected**	**Fold Enrichment**	**+/-**	***p* value**
ATP synthesis	8	3	0.12	25.3	+	3.87E-02
Heterotrimeric G-protein signaling pathway-rod outer segment phototransduction	59	7	0.87	8	+	5.34E-03
Unclassified	24,075	336	356.87	0.94	-	0.00E+00

A.6: Differentially expressed exons in 1-2 cell stage embryos of *A. limnaeus* (FDR *p*-value < 0.10).

Gene Symbol	Gene	Exon	Length (bp)	Mean expression (FPKM)	*P* Value (FDR)	LOG_2 Fold change	Phenotype
LOC106533268	Uncharacterized	E004	532	46.8	0	-0.2	**Escape**
LOC106522342	SSUH2 homolog: Zinc finger domain of DnaJ and HSP40	E006	125	6.4	0	2.4	**Diapause**
svep1	Sushi, von Willebrand factor type A, EGF and pentraxin domain containing 1	E025	202	8.2	0	2.8	**Diapause**
LOC106524347	Titin-like	E007	264	3.9	0	-1.7	**Escape**
LOC106520353	Vinculin-like	E013	200	27.1	0.0004	1.1	**Diapause**
hspa14	Heat shock 70kDa protein 14	E003	83	13.1	0.001	1.2	**Diapause**
fam135a	Family with sequence similarity 135, member A	E009	155	11	0.0012	1.5	**Diapause**
arhgap39	Rho GTPase activating protein 39	E003	92	5.7	0.0037	2.5	**Diapause**
hectd4	HECT domain containing E3 ubiquitin protein ligase 4	E005	109	5.6	0.0037	-2	**Escape**
phkb	Phosphorylase kinase, beta	E019	130	2.9	0.0037	6.6	**Diapause**
LOC106536223	Ankycorbin-like	E005	65	4.5	0.0037	5.9	**Diapause**
LOC106524277	Sideroflexin-5-like	E003	86	25.5	0.0124	0.8	**Diapause**
LOC106531012	Sodium channel protein type 4 subunit alpha B-like	E015	155	4	0.013	7.4	**Diapause**
LOC106535987	Histone RNA hairpin-binding protein-like	E010	9	4.4	0.0133	5.5	**Diapause**
LOC106517127	DNA-dependent protein kinase catalytic subunit-like	E010	132	4.7	0.0144	6.9	**Diapause**
pign	Phosphatidylinositol glycan anchor biosynthesis, class N	E004	99	6.9	0.0189	1.5	**Diapause**
larp4	La ribonucleoprotein domain family, member 4	E004	82	6.1	0.0233	7	**Diapause**
LOC106511683	Ephrin type-B receptor 4-like	E006	69	5.8	0.0244	2.3	**Diapause**
prlh	Prolactin releasing hormone	E003	37	3.6	0.0266	5.1	**Diapause**
dync1h1	Dynein, cytoplasmic 1, heavy chain 1	E033	215	5.3	0.031	2.4	**Diapause**
tmem38a	Transmembrane protein 38A	E001	800	39.5	0.0337	-0.2	**Escape**
LOC106535019	Uncharacterized	E001	153	5.1	0.0337	1.6	**Diapause**
slc25a26	Solute carrier family 25 (S-adenosylmethionine carrier), member 26	E010	149	4.2	0.0353	2	**Diapause**
tmem67	Transmembrane protein 67	E024	95	2.9	0.0355	5	**Diapause**
kiaa0430	KIAA0430 ortholog	E013	187	16.6	0.0367	1.1	**Diapause**
LOC106535645	Filamin-C-like	E006	116	10.9	0.0367	-1.2	**Escape**
LOC106514991	Tropomyosin alpha-1 chain-like	E009	79	8.4	0.0367	1.9	**Diapause**
nmt2	N-myristoyltransferase 2	E009	109	18.9	0.0374	1.1	**Diapause**
LOC106528181	Protein NLRC3-like	E004	165	7.8	0.0476	2.3	**Diapause**
LOC106530352	Polycystic kidney disease protein 1-like 2	E001	350	3	0.0496	2	**Diapause**
LOC106514264	Death-associated protein	E006	49	4.7	0.0513	2.1	**Diapause**

	kinase 2-like						
LOC106530165	Round spermatid basic protein 1-like	E001	601	3.8	0.0532	2.1	**Diapause**
malrd1	MAM and LDL receptor class A domain containing 1	E014	246	3.5	0.0569	8	**Diapause**
LOC106518696	Uncharacterized	E012	54	3.8	0.0586	1.8	**Diapause**
LOC106521136	DNA-binding protein RFX7-like	E007	85	6.6	0.0586	2	**Diapause**
cnnm3	Cyclin and CBS domain divalent metal cation transport mediator 3	E003	97	3.9	0.0586	1.9	**Diapause**
f13a1	Coagulation factor XIII, A1 polypeptide	E006	139	5.5	0.0586	5.3	**Diapause**
LOC106536558	E3 ubiquitin-protein ligase UBR2-like	E007	139	5.6	0.0586	2.4	**Diapause**
ireb2	Iron-responsive element binding protein 2	E014	85	14.2	0.0605	1.1	**Diapause**
snrnp200	Small nuclear ribonucleoprotein 200kDa (U5)	E003	172	256.3	0.0771	-0.2	**Escape**
lrrk2	Leucine-rich repeat kinase 2	E020	192	4.9	0.0771	1.8	**Diapause**
atat1	Alpha tubulin acetyltransferase 1	E009	114	8.2	0.0824	1.5	**Diapause**
LOC106533628	Disintegrin and metalloproteinase domain-containing protein 8-like	E014	178	5.6	0.0824	5.3	**Diapause**
kif1bp	KIF1 binding protein	E001	390	12.7	0.0824	1.3	**Diapause**
ephb2	EPH receptor B2	E004	156	3.2	0.0865	7.4	**Diapause**
kiaa0825	KIAA0825 ortholog	E013	191	1.6	0.0865	22.6	**Diapause**
bpgm	2,3-bisphosphoglycerate mutase	E001	1078	1.3	0.0865	-8.3	**Escape**
dgkz	Diacylglycerol kinase, zeta	E002	100	9.3	0.0865	1.4	**Diapause**
rps6kb1	Ribosomal protein S6 kinase, 70kDa, polypeptide 1	E007	101	63.3	0.0876	-0.3	**Escape**
LOC106521120	Copine-3-like	E016	180	8.2	0.0876	1.8	**Diapause**
pole	Polymerase (DNA directed), epsilon, catalytic subunit	E019	147	16	0.0876	1	**Diapause**
ints1	Integrator complex subunit 1	E018	61	3.5	0.0876	2.4	**Diapause**
LOC106536133	TRAF2 and NCK-interacting protein kinase-like	E004	131	2.1	0.0876	5.4	**Diapause**
LOC106513372	Golgi reassembly-stacking protein 1-like	E002	123	19.6	0.0876	0.6	**Diapause**
spata6	spermatogenesis associated 6	E002	146	4.5	0.0896	2	**Diapause**
tmem141	Transmembrane protein 141	E004	145	2.7	0.0985	1.9	**Diapause**
LOC106514639	Rho guanine nucleotide exchange factor 12-like	E010	111	7.6	0.0985	1.7	**Diapause**

A.7: GO analysis of top 100 most abundant transcripts in the 1-2 cell stage transcriptomes of *A. limnaeus* and *D. rerio*.

Analysis Type: PANTHER Overrepresentation Test (release 20160715)
Annotation Version and Release Date: PANTHER version 11.1 Released 2016-10-24
Analyzed List: *A. limnaeus* 100 most abundant transcripts
Reference List: *D. rerio* 100 most abundant transcripts
Bonferroni correction: TRUE
Displaying only results with $p < 0.05$

PANTHER GO-Slim Molecular Function	**Drer_top100 (REF)**	**Alim_top100**	**expected**	**Fold Enrichment**	**+/-**	***p* value**
calcium ion binding	2	8	2.04	3.92	+	2.23E-02
Unclassified	44	45	44.85	1	+	0.00E+00

PANTHER GO-Slim Biological Process	**Drer_top100 (REF)**	**Alim_top100**	**expected**	**Fold Enrichment**	**+/-**	***p* value**
cellular component movement	1	11	1.02	10.79	+	3.19E-07
Unclassified	32	41	32.62	1.26	+	0.00E+00
nucleobase-containing compound metabolic process	30	13	30.58	0.43	-	1.62E-03
primary metabolic process	43	27	43.83	0.62	-	1.88E-02
metabolic process	52	35	53.01	0.66	-	1.17E-02
chromatin organization	11	2	11.21	< 0.2	-	2.66E-02

PANTHER GO-Slim Cellular Component	**Drer_top100 (REF)**	**Alim_top100**	**expected**	**Fold Enrichment**	**+/-**	***p* value**
cytoskeleton	2	8	2.04	3.92	+	1.59E-02
Unclassified	68	70	69.32	1.01	+	0.00E+00

PANTHER Protein Class	**Drer_top100 (REF)**	**Alim_top100**	**expected**	**Fold Enrichment**	**+/-**	***p* value**
microtubule family cytoskeletal protein	1	6	1.02	5.89	+	1.19E-02
cytoskeletal protein	3	10	3.06	3.27	+	2.04E-02
calmodulin	2	8	2.04	3.92	+	2.12E-02
intracellular calcium-sensing protein	2	8	2.04	3.92	+	2.12E-02
calcium-binding protein	2	8	2.04	3.92	+	2.12E-02
Unclassified	37	34	37.72	0.9	-	0.00E+00
nucleic acid binding	26	12	26.5	0.45	-	7.25E-03

PANTHER Pathways	**Drer_top100 (REF)**	**Alim_top100**	**expected**	**Fold Enrichment**	**+/-**	***p* value**
Unclassified	85	78	86.65	0.9	-	0.00E+00

A.8: Most abundant sncRNAs in *A. limnaeus* 1-2 cell stage transcriptome and the annotations of their genomic alignments.

Name	Mean	Align-ments	Leftmost Bp	antisense or Sense	Gene	Transcript	Annotation	Comp-onent	BLAST result
680*	168,769	4	1289	N/A	Intergenic				
680*	168,769	4	49	N/A	Scaffold region				18s rRNA gene - internal spacer// Zinc finger protein
680*	168,769	4	69503	antisense	LOC106531887	XR_001312002.1	ncRNA - unannotated	intron	18s rRNA gene - internal spacer// Zinc finger protein
680*	168,769	4	6811644	antisense	LOC106525779	XM_014020117.1	mRNA - synaptophysin-like protein 1	intron	
1236	79,972	5	69502	antisense	LOC106531887	XR_001312002.1	ncRNA - unannotated	intron	18s rRNA gene - internal spacer// Zinc finger protein
1236	79,972	5	413408	antisense	LOC106529747	XM_014025242.1	mRNA - Fibulin-1-like	intron	
1236	79,972	5	660884	antisense	LOC106526902	XM_014021623.1	mRNA - toll-like receptor-2	intron	
1236	79,972	5	85961	N/A	Intergenic				
1236	79,972	5	49	N/A	Scaffold region				18s rRNA gene - internal spacer// Zinc finger protein
1	5,280	4	11756	N/A	Intergenic				
1	5,280	4	49	N/A	Scaffold region				18s rRNA gene - internal spacer// Zinc finger protein
1	5,280	4	1289	N/A	Intergenic				
1	5,280	4	69504	antisense	LOC106531887	XR_001312002.1	ncRNA - unannotated	intron	18s rRNA gene - internal spacer// Zinc finger protein
2	3,806	4	50	N/A	Scaffold region				18s rRNA gene - internal spacer// Zinc finger protein
2	3,806	4	6811645	antisense	LOC106525779	XM_014020117.1	mRNA - synaptophysin-like protein 1	intron	
2	3,806	4	1290	N/A	Intergenic				
2	3,806	4	69503	antisense	LOC106531887	XR_001312002.1	ncRNA - unannotated	intron	18s rRNA gene - internal spacer// Zinc finger protein
1239	941	4	69503	antisense	LOC106531887	XR_001312002.1	ncRNA - unannotated	intron	18s rRNA gene - internal spacer// Zinc finger protein
1239	941	4	1288	N/A	Intergenic				
1239	941	4	48	N/A	Scaffold region				18s rRNA gene - internal spacer// Zinc finger protein
1239	941	4	6811643	antisense	LOC106525779	XM_014020117.1	mRNA - synaptophysin-like protein 1	intron	
769	20	30	going	to	do	this			
1371	9	3	53270	Sense	slc1a7	XM_014024684.1	mRNA - solute carrier family 1 (glutamate transporter), member 7	intron	
1371	9	3	1940545	Sense	LOC106515129	XR_001309964.1	ncRNA - unannotated	intron	rho GTPase-activating protein
1371	9	3	282442	Sense	N/A	Intergenic			
1409	7	3	1289	N/A	Intergenic				
1409	7	3	6811644	antisense	LOC106525779	XM_014020117.1	mRNA - synaptophysin-like protein 1	intron	
1409	7	3	191389	Sense// antisense	ralgps2// LOC106532039	XM_014028025.1// XM_014028026.1	mRNA - Ral GEF with PH domain and SH3 binding motif 2// mRNA - angiopoietin-related protein 1-like	intronic // exonic	

APPENDIX B:

Tissue sampling and small RNA-sequencing from temperature induced diapause- and escape-bound embryos

B1: Biological information for embryos and RNA extractions

ID	Tx	Developmental Stage	Temp.	# of embryos	RNA Conc. (uL/ng)	A_{260}	A_{280}	A_{260}/A_{280}	Vol (µL)	RNA Yield (ng)
A1	Escape	Cell Dispersion	30°C	40	346	0.433	0.201	2.15	20	6925
A2	Escape	Cell Dispersion	30°C	40	322	0.403	0.189	2.13	20	6446
A3	Escape	Cell Dispersion	30°C	40	429	0.536	0.251	2.14	20	8581
B1	Escape	Neural Keel	30°C	40	547	0.684	0.317	2.16	20	10942
B2	Escape	Neural Keel	30°C	40	380	0.476	0.224	2.13	20	7608
B3	Escape	Neural Keel	30°C	40	300	0.374	0.177	2.12	20	5990
C1	Escape	6 somites	30°C	40	232	0.291	0.137	2.11	20	4650
C2	Escape	6 somites	30°C	40	371	0.464	0.217	2.14	20	7429
C3	Escape	6 somites	30°C	40	364	0.456	0.217	2.1	20	7290
D1	Escape	10 somites	30°C	30	316	0.395	0.190	2.08	20	6315
D2	Escape	10 somites	30°C	30	428	0.536	0.267	2.01	20	8568
D3	Escape	10 somites	30°C	30	394	0.492	0.234	2.11	20	7877
E1	Escape	16 somites	30°C	30	357	0.446	0.226	1.97	20	7141
E2	Escape	16 somites	30°C	30	367	0.459	0.220	2.09	20	7339
E3	Escape	16 somites	30°C	30	413	0.516	0.241	2.14	20	8262
F1	Escape	20 somites	30°C	20	510	0.638	0.325	1.96	20	10200
F2	Escape	20 somites	30°C	20	355	0.444	0.230	1.93	20	7101
F3	Escape	20 somites	30°C	20	407	0.508	0.257	1.98	20	8131
G1	Escape	24 somites	30°C	20	747	0.933	0.429	2.17	20	14934
G2	Escape	24 somites	30°C	20	333	0.416	0.204	2.04	20	6656
G3	Escape	24 somites	30°C	20	529	0.661	0.316	2.1	20	10578
H1	Diapause	Cell Dispersion	20°C	40	139	0.173	0.089	1.95	20	2771
H2	Diapause	Cell Dispersion	20°C	40	231	0.288	0.136	2.12	20	4613
H3	Diapause	Cell Dispersion	20°C	40	253	0.317	0.150	2.11	20	5066
I1	Diapause	Neural Keel	20°C	40	365	0.457	0.212	2.16	20	7309
I2	Diapause	Neural Keel	20°C	40	391	0.489	0.232	2.1	20	7824
I3	Diapause	Neural Keel	20°C	40	631	0.789	0.369	2.14	20	12621
J1	Diapause	6 somites	20°C	40	375	0.469	0.217	2.16	20	7498
J2	Diapause	6 somites	20°C	40	385	0.482	0.237	2.04	20	7706

J3	Diapause	6 somites	20°C	40	390	0.487	0.253	1.93	20	7795
K1	Diapause	10 somites	20°C	30	702	0.877	0.452	1.94	20	14035
K2	Diapause	10 somites	20°C	30	399	0.499	0.242	2.06	20	7982
K3	Diapause	10 somites	20°C	30	353	0.442	0.209	2.11	20	7066
L1	Diapause	16 somites	20°C	30	372	0.465	0.230	2.02	20	7443
L2	Diapause	16 somites	20°C	30	397	0.497	0.240	2.07	20	7944
L3	Diapause	16 somites	20°C	30	582	0.728	0.347	2.1	20	11648
M1	Diapause	20 somites	20°C	20	263	0.328	0.153	2.14	20	5254
M2	Diapause	20 somites	20°C	20	192	0.240	0.115	2.09	20	3832
M3	Diapause	20 somites	20°C	20	106	0.152	0.072	2.12	20	2118
N1	Diapause	24 somites	20°C	20	160	0.200	0.094	2.12	20	3198
N2	Diapause	24 somites	20°C	20	194	0.242	0.121	2	20	3874
N3	Diapause	24 somites	20°C	20	89	0.067	0.032	2.08	20	1778

B2: Small RNA library trimming and mapping statistics

Library	NCBI accession	Sequenced Reads	Trimmomatic Trimmed and filtered reads	Trimmed, %	CLC Reads	CLC Unique Small RNAs	Reads Genome Annotated	Unique RNAs Genome Annotated
A1	SRR5256760	12,780,750	8,541,636	66.8	5,987,690	987,125	4,608,951	517,576
A2	SRR5256759	4,856,608	3,541,510	72.9	2,289,604	404,019	1,834,148	237,314
A3	SRR5256758	9,191,706	6,019,441	65.5	4,705,319	567,459	3,980,870	382,862
B1	SRR5256757	10,873,150	9,001,373	82.8	6,621,166	657,759	5,447,509	389,632
B2	SRR5256756	3,511,776	2,666,124	75.9	1,761,301	304,645	1,261,354	135,499
B3	SRR5256755	5,723,005	4,322,133	75.5	3,150,536	472,270	2,355,283	228,454
C1	SRR5256754	13,903,348	6,971,224	50.1	4,319,364	740,355	3,035,840	340,523
C2	SRR5256753	4,266,114	2,597,006	60.9	1,581,028	332,331	1,002,810	127,666
C3	SRR5256752	13,344,926	9,879,440	74.0	6,433,024	1,001,669	4,659,601	457,208
D1	SRR5256751	7,037,313	5,977,788	84.9	3,619,794	544,219	2,226,691	168,368
D2	SRR5256750	7,615,189	5,624,509	73.9	3,380,670	719,747	2,071,673	268,559
D3	SRR5256749	12,701,779	9,142,759	72.0	5,687,568	944,627	4,018,353	431,595
E1	SRR5256748	4,320,921	2,927,591	67.8	1,974,704	320,079	1,598,338	161,323
E2	SRR5256747	6,855,594	5,204,739	75.9	3,505,137	681,758	2,260,660	252,567
E3	SRR5256746	12,907,291	9,422,636	73.0	6,524,781	966,607	4,924,120	460,377
F1	SRR5256745	5,583,242	3,799,243	68.1	2,767,447	459,885	2,076,152	226,462
F2	SRR5256744	4,519,542	3,278,514	72.5	2,276,359	392,405	1,655,722	165,484
F3	SRR5256743	14,896,396	10,263,851	68.9	6,185,707	913,324	4,415,798	436,626
G1	SRR5256742	12,914,966	2,927,314	22.7	2,382,968	378,935	1,804,139	190,231
G2	SRR5256741	17,584,154	12,570,457	71.5	9,532,486	941,890	6,791,875	463,390
G3	SRR5256740	14,630,299	11,575,916	79.1	7,512,450	752,396	5,844,029	319,909
H1	SRR5256739	11,976,520	1,704,110	14.2	1,218,719	294,105	587,050	81,960
H2	SRR5256738	11,307,931	8,213,069	72.6	5,011,203	866,452	2,974,575	300,625
H3	SRR5256737	7,150,569	2,010,089	28.1	1,186,080	310,324	536,608	77,539
I1	SRR5256736	15,105,962	1,837,440	12.2	1,525,132	397,992	450,227	62,184
I2	SRR5256735	6,619,027	5,234,686	79.1	3,187,415	640,186	1,184,705	141,580
I3	SRR5256734	4,897,372	3,132,736	64.0	1,700,686	466,770	678,113	101,953
J1	SRR5256733	19,199,894	745,964	3.9	556,192	172,593	271,479	47,344
J2	SRR5256732	7,400,348	4,695,502	63.5	2,477,542	543,698	1,073,588	115,180
J3	SRR5256731	12,881,057	5,933,094	46.1	3,978,429	641,783	2,009,511	197,243
K1	SRR5256730	9,442,434	2,043,927	21.7	1,397,391	297,870	713,423	70,656
K2	SRR5256729	10,615,608	908,967	8.6	481,127	124,237	224,451	27,504
K3	SRR5256728	8,144,583	6,518,785	80.0	3,370,873	597,520	1,833,257	148,443
L1	SRR5256727	10,992,369	8,781,410	79.9	6,462,118	714,431	4,347,728	213,428
L2	SRR5256726	9,147,147	6,875,361	75.2	4,878,863	693,656	3,098,067	191,720
L3	SRR5256725	10,154,840	6,548,484	64.5	4,679,953	783,813	2,569,639	205,227

M1	SRR5256724	7,808,214	4,247,452	54.4	3,494,704	425,937	2,556,538	149,055
M2	SRR5256723	15,215,847	9,940,414	65.3	8,561,099	907,986	6,107,382	355,370
M3	SRR5256722	16,476,151	10,843,065	65.8	9,148,368	982,304	6,714,717	390,713
N1	SRR5256721	15,329,790	9,963,589	65.0	7,484,159	1,025,156	5,621,467	563,788
N2	SRR5256720	13,503,683	4,085,011	30.3	2,488,384	501,959	1,341,167	168,779
N3	SRR5256719	4,886,990	1,106,057	22.6	826,454	200,388	605,858	113,311

APPENDIX C:

WGCNA Analysis

C.1 Temperature induced sampling and poly-A sample library ID and total reads sequenced

Library	Sample Identity	Stage	Temp.	Phenotype	NCBI Accession ID	Total Read Pairs
A1	Cell dispersion (E)	Cell dispersion	30°C	Escape	SRX1032697	81,601,161
A2	Cell dispersion (E)	Cell dispersion	30°C	Escape	SRX1032698	99,643,025
A3	Cell dispersion (E)	Cell dispersion	30°C	Escape	SRX1032699	65,545,750
B1	Neural keel (E)	Neural keel	30°C	Escape	SRX1032655	29,700,353
B2	Neural keel (E)	Neural keel	30°C	Escape	SRX1032657	33,628,167
B3	Neural keel (E)	Neural keel	30°C	Escape	SRX1032658	31,764,350
C1	6 s.p. (E)	6 pairs of somites	30°C	Escape	SRX1032664	17,174,837
C2	6 s.p. (E)	6 pairs of somites	30°C	Escape	SRX1032665	14,326,493
C3	6 s.p. (E)	6 pairs of somites	30°C	Escape	SRX1032666	13,858,273
D1	10 s.p. (E)	10 pairs of somites	30°C	Escape	SRX1032700	72,206,490
D2	10 s.p. (E)	10 pairs of somites	30°C	Escape	SRX1032681	32,464,655
D3	10 s.p. (E)	10 pairs of somites	30°C	Escape	SRX1032682	42,889,474
E1	16 s.p. (E)	16 pairs of somites	30°C	Escape	SRX1032667	28,970,462
E2	16 s.p. (E)	16 pairs of somites	30°C	Escape	SRX1032668	10,672,937
E3	16 s.p. (E)	16 pairs of somites	30°C	Escape	SRX1032669	19,022,571
F1	20 s.p. (E)	20 pairs of somites	30°C	Escape	SRX1032683	34,668,279
F2	20 s.p. (E)	20 pairs of somites	30°C	Escape	SRX1032684	46,886,327
F3	20 s.p. (E)	20 pairs of somites	30°C	Escape	SRX1032670	14,584,197
G1	24 s.p. (E)	24 pairs of somites	30°C	Escape	SRX1032685	47,671,680
G2	24 s.p. (E)	24 pairs of somites	30°C	Escape	SRX1032686	45,291,488
G3	24 s.p. (E)	24 pairs of somites	30°C	Escape	SRX1032687	60,652,844
H1	Cell dispersion (D)	Cell dispersion	20°C	Diapause	SRX1032688	42,527,726
H2	Cell dispersion (D)	Cell dispersion	20°C	Diapause	SRX1032689	53,727,130
H3	Cell dispersion (D)	Cell dispersion	20°C	Diapause	SRX1032690	34,009,793
I1	Neural keel (D)	Neural keel	20°C	Diapause	SRX1032671	30,202,190
I2	Neural keel (D)	Neural keel	20°C	Diapause	SRX1032672	30,650,268
I3	Neural keel (D)	Neural keel	20°C	Diapause	SRX1032673	29,624,680
J1	6 s.p. (D)	6 pairs of somites	20°C	Diapause	SRX1032674	18,511,610
J2	6 s.p. (D)	6 pairs of somites	20°C	Diapause	SRX1032675	14,738,622

J3	6 s.p. (D)	6 pairs of somites	20°C	Diapause	SRX1032676	10,945,746
K1	10 s.p. (D)	10 pairs of somites	20°C	Diapause	SRX1032691	54,741,661
K2	10 s.p. (D)	10 pairs of somites	20°C	Diapause	SRX1032692	57,868,256
K3	10 s.p. (D)	10 pairs of somites	20°C	Diapause	SRX1032693	40,003,497
L1	16 s.p. (D)	16 pairs of somites	20°C	Diapause	SRX1032677	46,199,465
L2	16 s.p. (D)	16 pairs of somites	20°C	Diapause	SRX1032678	9,927,116
L3	16 s.p. (D)	16 pairs of somites	20°C	Diapause	SRX1032679	8,092,672
M1	20 s.p. (D)	20 pairs of somites	20°C	Diapause	SRX1032701	56,136,088
M2	20 s.p. (D)	20 pairs of somites	20°C	Diapause	SRX1032694	50,709,319
M3	20 s.p. (D)	20 pairs of somites	20°C	Diapause	SRX1032680	21,146,825
N1	24 s.p. (D)	24 pairs of somites	20°C	Diapause	SRX1032702	46,007,419
N2	24 s.p. (D)	24 pairs of somites	20°C	Diapause	SRX1032695	65,288,546
N3	24 s.p. (D)	24 pairs of somites	20°C	Diapause	SRX1032696	45,014,206

C.2 Poly-A sequence library trimming statistics and alignments.

ID	Total Read Pairs	Both Surviving, %[1]		Forward-only[2]		Reverse-only[2]		Dropped[3]		Paired alignments[4]		Single alignments[5]	
A1	81,601,161	69,882,536	86	2,319,107	2.8	2,268,101	2.8	7,131,417	8.7	69,882,536	88	4,587,208	83
A2	99,643,025	88,703,787	89	6,000,686	6.0	1,275,951	1.3	3,662,601	3.7	88,703,787	94	7,276,637	88
A3	65,545,750	57,355,447	88	3,832,398	5.8	1,151,517	1.8	3,206,388	4.9	57,355,447	94	4,983,915	86
B1	29,700,353	28,141,786	95	653,470	2.2	59,689	0.2	845,408	2.8	28,141,786	82	713,159	81
B2	33,628,167	30,378,030	90	642,461	1.9	184,563	0.5	2,423,113	7.2	30,378,030	91	827,024	86
B3	31,764,350	29,083,502	92	483,824	1.5	195,120	0.6	2,001,904	6.3	29,083,502	94	678,944	87
C1	17,174,837	16,728,102	97	144,611	0.8	108,751	0.6	193,373	1.1	16,728,102	90	253,362	86
C2	14,326,493	6,135,660	43	77,137	0.5	42,342	0.3	8,071,354	56.3	6,135,660	71	119,479	73
C3	13,858,273	11,404,320	82	204,269	1.5	147,349	1.1	2,102,335	15.2	11,404,320	77	351,618	82
D1	72,206,490	65,235,665	90	3,697,277	5.1	916,464	1.3	2,357,084	3.3	65,235,665	92	4,613,741	85
D2	32,464,655	30,850,676	95	582,700	1.8	105,562	0.3	925,717	2.9	30,850,676	83	688,262	76
D3	42,889,474	41,486,377	97	952,764	2.2	165,582	0.4	284,751	0.7	41,486,377	90	1,118,346	81
E1	28,970,462	23,630,608	82	362,962	1.3	208,302	0.7	4,768,590	16.5	23,630,608	80	571,264	80
E2	10,672,937	9,140,547	86	187,629	1.8	166,686	1.6	1,178,075	11.0	9,140,547	89	354,315	84
E3	19,022,571	3,770,952	20	121,901	0.6	75,989	0.4	15,053,729	79.1	3,770,952	76	197,890	81
F1	34,668,279	33,260,821	96	994,312	2.9	141,410	0.4	271,736	0.8	33,260,821	96	1,135,722	91
F2	46,886,327	44,982,383	96	1,521,932	3.2	195,134	0.4	186,878	0.4	44,982,383	95	1,717,066	90
F3	14,584,197	6,377,102	44	176,615	1.2	95,970	0.7	7,934,510	54.4	6,377,102	80	272,585	84
G1	47,671,680	45,422,970	95	1,675,777	3.5	201,615	0.4	371,318	0.8	45,422,970	95	1,877,392	92
G2	45,291,488	43,881,398	97	920,269	2.0	203,503	0.4	286,318	0.6	43,881,398	93	1,123,772	86
G3	60,652,844	59,163,163	98	1,154,155	1.9	211,316	0.3	124,210	0.2	59,163,163	96	1,365,471	89
H1	42,527,726	41,514,567	98	697,901	1.6	146,379	0.3	168,879	0.4	41,514,567	93	844,280	86
H2	53,727,130	52,424,131	98	954,456	1.8	157,564	0.3	190,979	0.4	52,424,131	96	1,112,020	89
H3	34,009,793	33,055,318	97	581,630	1.7	117,839	0.3	255,006	0.7	33,055,318	95	699,469	87
I1	30,202,190	16,750,564	55	397,513	1.3	127,606	0.4	12,926,507	42.8	16,750,564	91	525,119	82
I2	30,650,268	29,706,379	97	480,396	1.6	69,635	0.2	393,858	1.3	29,706,379	89	550,031	84
I3	29,624,680	28,215,083	95	532,803	1.8	137,798	0.5	738,996	2.5	28,215,083	91	670,601	84
J1	18,511,610	18,073,972	98	156,180	0.8	122,018	0.7	159,440	0.9	18,073,972	89	278,198	86
J2	14,738,622	12,993,211	88	113,807	0.8	65,007	0.4	1,566,597	10.6	12,993,211	77	178,814	76
J3	10,945,746	9,092,222	83	100,138	0.9	83,075	0.8	1,670,311	15.3	9,092,222	89	183,213	83
K1	54,741,661	53,297,909	97	980,388	1.8	139,983	0.3	323,381	0.6	53,297,909	89	1,120,371	82
K2	57,868,256	56,161,993	97	1,416,126	2.4	164,332	0.3	125,805	0.2	56,161,993	93	1,580,458	85
K3	40,003,497	38,861,476	97	920,020	2.3	123,220	0.3	98,781	0.2	38,861,476	90	1,043,240	82
L1	46,199,465	40,630,879	88	752,585	1.6	451,255	1.0	4,364,746	9.4	40,630,879	90	1,203,840	85
L2	9,927,116	9,463,720	95	94,410	1.0	46,942	0.5	322,044	3.2	9,463,720	87	141,352	83
L3	8,092,672	7,869,136	97	81,790	1.0	33,913	0.4	107,833	1.3	7,869,136	87	115,703	83
M1	56,136,088	53,669,883	96	1,444,925	2.6	739,478	1.3	281,802	0.5	53,669,883	95	2,184,403	89
M2	50,709,319	48,492,639	96	1,312,301	2.6	208,860	0.4	695,519	1.4	48,492,639	95	1,521,161	89
M3	21,146,825	14,430,643	68	262,714	1.2	158,745	0.8	6,294,723	29.8	14,430,643	75	421,459	78
N1	46,007,419	43,163,365	94	1,370,070	3.0	524,129	1.1	949,855	2.1	43,163,365	92	1,894,199	87
N2	65,288,546	63,386,858	97	1,580,355	2.4	180,669	0.3	140,664	0.2	63,386,858	96	1,761,024	87
N3	45,014,206	43,627,085	97	1,135,538	2.5	121,927	0.3	129,656	0.3	43,627,085	94	1,257,465	86

1. The number of paired reads for each library.
2. The number of non-paired, or orphaned, reads specific to orientation that survived the trimming criteria for each library.
3. Total number of reads (paired and non-paired, or orphaned) that did not survive trimming criteria.
4. The number of paired reads for each library that aligned to the *A. limnaeus* genome.
5. The number of non-paired, or orphaned, reads specific to orientation that aligned to the *A. limnaeus* genome.

C.3: Top twenty differentially expressed gene transcripts at cell dispersion

Gene Symbol	Description	Gene Type	Tx[1]	30 °C Mean[2]	SEM[3]	20 °C Mean[2]	SEM[3]
LOC106531617	metalloproteinase inhibitor 2-like	mRNA	E	259	58	9	1
med13	mediator complex subunit 13	mRNA	E	21	3	3	0
LOC106523459	vitronectin-like	mRNA	E	118	17	17	1
naaladl1	N-acetylated alpha-linked acidic dipeptidase-like 1	mRNA	E	69	8	10	1
LOC106513218	ribonuclease-like 3	mRNA	E	442	100	52	6
fam13c	family with sequence similarity 13 member C	mRNA	E	32	2	6	1
LOC106532471	cyclic nucleotide-gated cation channel alpha-3-like	mRNA	D	0	0	8	2
LOC106525651	dimethylaniline monooxygenase [N-oxide-forming] 5-like	mRNA	E	61	4	11	1
LOC106533477	cryptochrome-1-like	mRNA	E	25	1	3	1
inpp1	inositol polyphosphate-1-phosphatase	mRNA	E	44	8	6	1
prodh2	proline dehydrogenase (oxidase) 2	mRNA	E	63	7	10	1
LOC106535731	uncharacterized LOC106535731	mRNA	E	16	1	2	0
flt1	fms-related tyrosine kinase 1	mRNA	E	2	0	10	1
LOC106523775	eosinophil peroxidase-like	mRNA	E	0	0	15	9
LOC106514179	high mobility group protein B1-like	mRNA	D	98	24	616	31
LOC106526824	ubiquitin-conjugating enzyme E2 Q2-like	mRNA	D	23	4	112	9
abcc5	ATP-binding cassette sub-family C (CFTR/MRP) member 5	mRNA	E	7	0	1	0
smarce1	SWI/SNF related matrix associated actin dependent regulator of chromatin subfamily e member 1	mRNA	E	19	3	2	0
LOC106530073	acidic mammalian chitinase-like	mRNA	E	13	9	0	0
LOC106520443	uncharacterized	pseudogene	E	14	4	0	0
LOC106533842	mitogen-activated protein kinase-binding protein 1-like	mRNA	D	63	12	307	33

1. Tx= treatment group; each gene is identified by which treatment group has the highest expression. D = diapause (20 °C) and E = escape (30 °C).
2. Mean normalized expression values in FPKM of the genes specific to treatment group.
3. Standard error of the mean calculated for each average normalized expression value.

C.4: Top twenty differentially expressed gene transcripts at Neural Keel.

Gene Symbol	Description	Gene Type	Tx[1]	30 °C Mean[2]	SEM[3]	20 °C Mean[2]	SEM[3]
LOC106535370	pseudouridine-metabolizing bifunctional protein C1861.05-like	mRNA	E	34	2	1	0
LOC106536745	metallothionein	mRNA	E	1,404	68	93	6
LOC106513667	UDP-glucuronosyltransferase 3A1-like	mRNA	D	3	1	87	1
nexn	nexilin (F actin binding protein)	mRNA	E	33	2	1	0
LOC106518781	keratin type I cytoskeletal 47 kDa-like	mRNA	D	53	2	411	6
LOC106531617	metalloproteinase inhibitor 2-like	mRNA	E	240	12	8	2
LOC106525651	dimethylaniline monooxygenase [N-oxide-forming] 5-like	mRNA	E	116	25	5	0
LOC106527792	integrin beta-3-like	mRNA	D	3	0	31	1
LOC106523459	vitronectin-like	mRNA	E	201	29	18	1
LOC106520014	EH domain-containing protein 3	mRNA	E	44	5	4	0
LOC106527288	REST corepressor 1-like	mRNA	E	137	4	16	2

LOC106528104	aquaporin-1-like	mRNA	E	141	17	9	1
LOC106533398	cytosolic non-specific dipeptidase-like	mRNA	D	6	1	110	7
LOC106514105	large neutral amino acids transporter small subunit 4-like	mRNA	E	96	10	12	1
LOC106515427	sodium/potassium-transporting ATPase subunit gamma-like	mRNA	D	0	0	54	4
ventx	VENT homeobox	mRNA	D	77	2	450	32
scpep1	serine carboxypeptidase 1	mRNA	E	173	13	28	1
LOC106534701	protein-methionine sulfoxide oxidase mical1-like	mRNA	D	6	1	47	4
LOC106523775	eosinophil peroxidase-like	mRNA	E	0	0	8	1
LOC106521385	parathyroid hormone/parathyroid hormone-related peptide receptor-like	mRNA	E	110	9	13	2

1. Tx= treatment group; each gene is identified by which treatment group has the highest expression. D = diapause (20 °C) and E = escape (30 °C).
2. Mean normalized expression values in FPKM of the genes specific to treatment group.
3. Standard error of the mean calculated for each average normalized expression value.

C.5: Top twenty differentially expressed gene transcripts at 6 somites

Gene Symbol	Description	Gene Type	Tx[1]	30 °C Mean[2]	SEM[3]	20 °C Mean[2]	SEM[3]
LOC106523459	vitronectin-like	mRNA	E	148	4	9	1
nexn	nexilin (F actin binding protein)	mRNA	E	44	5	1	0
LOC106513667	UDP-glucuronosyltransferase 3A1-like	mRNA	D	4	0	93	5
LOC106535370	pseudouridine-metabolizing bifunctional protein C1861.05-like	mRNA	E	52	6	1	0
gamt	guanidinoacetate N-methyltransferase	mRNA	E	932	27	143	5
LOC106531617	metalloproteinase inhibitor 2-like	mRNA	E	189	9	5	0
LOC106513762	uncharacterized	ncRNA	E	10,341	1,463	1,086	22
LOC106521385	parathyroid hormone/parathyroid hormone-related peptide receptor-like	mRNA	E	145	7	16	2
LOC106514105	large neutral amino acids transporter small subunit 4-like	mRNA	E	102	5	9	1
LOC106520014	EH domain-containing protein 3	mRNA	E	57	7	5	0
LOC106525651	dimethylaniline monooxygenase [N-oxide-forming] 5-like	mRNA	E	123	25	2	1
LOC106528104	aquaporin-1-like	mRNA	E	84	14	1	1
LOC106529223	major facilitator superfamily domain-containing protein 4-A-like	mRNA	D	14	4	291	20
LOC106513404	cystatin-B-like	mRNA	E	597	50	91	6
LOC106513693	fatty acid-binding protein liver-type-like	mRNA	E	153	7	15	2
LOC106517651	spondin-2-like	mRNA	D	1	0	31	3
LOC106520168	beta-microseminoprotein-like	mRNA	E	3,941	801	348	28
krt222	keratin 222 type II	mRNA	E	1	0	29	1
LOC106512923	retinol dehydrogenase 3-like	mRNA	E	31	7	2	0
LOC106513763		ncRNA	Escape	429	10	52	9

1. Tx= treatment group; each gene is identified by which treatment group has the highest expression. D = diapause (20°C) and E = escape (30°C).
2. Mean normalized expression values in FPKM of the genes specific to treatment group.
3. Standard error of the mean calculated for each average normalized expression value.

C.6: Top twenty differentially expressed gene transcripts and 10 somites

Gene Symbol	Description	Gene Type	Tx[1]	30 °C Mean[2]	SEM[3]	20 °C Mean[2]	SEM[3]
LOC106525651	dimethylaniline monooxygenase [N-oxide-forming] 5-like	mRNA	E	119	22	2	0
LOC106520014	EH domain-containing protein 3	mRNA	E	62	6	5	0
LOC106513667	UDP-glucuronosyltransferase 3A1-like	mRNA	D	4	1	87	7
LOC106514105	large neutral amino acids transporter small subunit 4-like	mRNA	E	102	13	8	1
LOC106530854	vitamin D 25-hydroxylase	mRNA	E	12	3	0	0
LOC106529223	major facilitator superfamily domain-containing protein 4-A-like	mRNA	D	8	3	230	39
LOC106523459	vitronectin-like	mRNA	E	119	16	11	2
rbp4	retinol binding protein 4 plasma	mRNA	E	748	7	152	11
gatm	glycine amidinotransferase (L-arginine:glycine amidinotransferase)	mRNA	E	134	30	9	2
LOC106528695	major histocompatibility complex class I-related gene protein-like	mRNA	D	-	-	8	1
LOC106523096	uncharacterized LOC106523096	mRNA	E	89	33	3	1
sqrdl	sulfide quinone reductase-like (yeast)	mRNA	E	110	10	21	1
LOC106512392	ryanodine receptor 2-like	mRNA	D	0	0	5	1
LOC106532471	cyclic nucleotide-gated cation channel alpha-3-like	mRNA	D	0	0	7	2
LOC106515142	plasma membrane calcium-transporting ATPase 1-like	mRNA	E	67	14	9	0
LOC106532067	glutathione S-transferase A-like	mRNA	E	712	57	162	6
LOC106536378	elongation of very long chain fatty acids protein 6-like	mRNA	E	88	23	6	0
LOC106534701	protein-methionine sulfoxide oxidase mical1-like	mRNA	D	4	1	52	7
LOC106515427	sodium/potassium-transporting ATPase subunit gamma-like	mRNA	D	1	0	43	4
LOC106530864	calcitonin-1-like	mRNA	E	15	4	0	0

1. Tx= treatment group; each gene is identified by which treatment group has the highest expression. D = diapause (20°C) and E = escape (30°C).
2. Mean normalized expression values in FPKM of the genes specific to treatment group.
3. Standard error of the mean calculated for each average normalized expression value.

C.7: Top twenty differentially expressed gene transcripts and 16 somites

Gene Symbol	Description	Gene Type	Tx[1]	30 °C Mean[2]	SEM[3]	20 °C Mean[2]	SEM[3]
LOC106535370	pseudouridine-metabolizing bifunctional protein C1861.05-like	mRNA	E	61	7	1	0
LOC106525651	dimethylaniline monooxygenase [N-oxide-forming] 5-like	mRNA	E	137	6	4	1
gatm	glycine amidinotransferase (L-arginine:glycine amidinotransferase)	mRNA	E	200	18	9	1
kiaa0513	KIAA0513 ortholog	mRNA	E	347	2	31	1
LOC106526560	low choriolytic enzyme-like	mRNA	E	277	41	8	0
LOC106528894	hemoglobin subunit beta-like	mRNA	E	550	170	2	0
LOC106513667	UDP-glucuronosyltransferase 3A1-like	mRNA	D	3	1	91	12
LOC106534701	protein-methionine sulfoxide oxidase mical1-like	mRNA	D	4	1	57	6
LOC106526482	digestive cysteine proteinase 1-like	mRNA	E	46	4	3	0
LOC106529223	major facilitator superfamily domain-containing protein 4-A-like	mRNA	D	4	2	222	34
LOC106513693	fatty acid-binding protein liver-type-	mRNA	E	251	44	10	3

	like						
LOC106526230	putative defense protein 3	mRNA	E	84	26	0	0
jarid2	jumonji AT rich interactive domain 2	mRNA	D	13	3	207	19
LOC106514105	large neutral amino acids transporter small subunit 4-like	mRNA	E	74	3	10	1
hspd1	heat shock 60kDa protein 1 (chaperonin)	mRNA	E	211	6	41	2
LOC106520014	EH domain-containing protein 3	mRNA	E	62	5	7	1
acsl5	acyl-CoA synthetase long-chain family member 5	mRNA	E	27	5	1	0
LOC106518781	keratin type I cytoskeletal 47 kDa-like	mRNA	D	60	5	406	27
soat2	sterol O-acyltransferase 2	mRNA	E	14	2	0	0
LOC106523459	vitronectin-like	mRNA	E	69	3	8	1

1. Tx= treatment group; each gene is identified by which treatment group has the highest expression. D = diapause (20°C) and E = escape (30°C).
2. Mean normalized expression values in FPKM of the genes specific to treatment group.
3. Standard error of the mean calculated for each average normalized expression value.

C.8: Top twenty differentially expressed gene transcripts 20 somites

Gene Symbol	Description	Gene Type	Tx[1]	30 °C Mean[2]	SEM[3]	20 °C Mean[2]	SEM[3]
LOC106525651	dimethylaniline monooxygenase [N-oxide-forming] 5-like	mRNA	E	87	3	3	0
LOC106535370	pseudouridine-metabolizing bifunctional protein C1861.05-like	mRNA	E	34	4	0	0
LOC106527792	integrin beta-3-like	mRNA	D	0	0	20	1
LOC106523096	uncharacterized	mRNA	E	194	39	2	0
LOC106525311	vigilin-like	mRNA	D	10	1	110	1
trip12	thyroid hormone receptor interactor 12	mRNA	D	22	0	149	2
LOC106521262	alcohol dehydrogenase 1-like	mRNA	D	5	0	104	12
LOC106528904	hemoglobin embryonic subunit alpha-like	mRNA	E	503	210	1	0
LOC106521979	autism susceptibility gene 2 protein-like	mRNA	D	18	2	205	12
LOC106533922	cytochrome c	mRNA	E	129	6	14	1
LOC106517651	spondin-2-like	mRNA	D	0	0	41	11
LOC106517656	glycerol-3-phosphate dehydrogenase [NAD(+)] cytoplasmic-like	mRNA	E	35	5	1	0
LOC106526824	ubiquitin-conjugating enzyme E2 Q2-like	mRNA	D	32	4	498	84
pcna	proliferating cell nuclear antigen	mRNA	E	399	27	56	3
LOC106528695	major histocompatibility complex class I-related gene protein-like	mRNA	D	0	0	16	1
LOC106528087	serine/threonine-protein kinase WNK4-like	mRNA	D	0	0	10	2
pah	phenylalanine hydroxylase	mRNA	D	8	1	89	2
LOC106513667	UDP-glucuronosyltransferase 3A1-like	mRNA	D	4	1	110	22
LOC106513404	cystatin-B-like	mRNA	E	616	74	68	2
LOC106534703	calcium/calmodulin-dependent protein kinase type 1D-like	mRNA	D	3	0	43	6

1. Tx= treatment group; each gene is identified by which treatment group has the highest expression. D = diapause (20°C) and E = escape (30°C).
2. Mean normalized expression values in FPKM of the genes specific to treatment group.
3. Standard error of the mean calculated for each average normalized expression value.

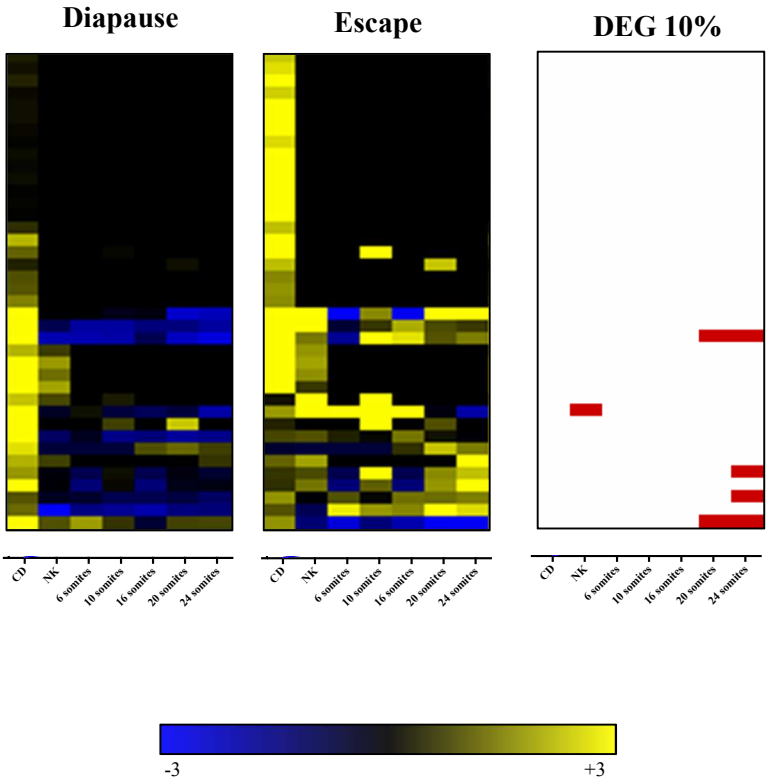
C.9: Top twenty differentially expressed gene transcripts 24 somites

Gene Symbol	Description	Gene Type	Tx[1]	30 °C Mean[2]	SEM[3]	20 °C Mean[2]	SEM[3]
LOC106523096	uncharacterized	mRNA	E	147	17	1	0
LOC106528894	hemoglobin subunit beta-like	mRNA	E	1,134	205	4	1
LOC106515188	immunoglobulin superfamily DCC subclass member 3-like	mRNA	E	55	19	0	0
LOC106518781	keratin type I cytoskeletal 47 kDa-like	mRNA	D	19	3	567	31
LOC106526824	ubiquitin-conjugating enzyme E2 Q2-like	mRNA	D	32	4	690	7
LOC106516979	dnaJ homolog subfamily C member 7-like	mRNA	D	18	1	255	6
pkd1l1	polycystic kidney disease 1 like 1	mRNA	D	0	0	5	0
ppp4r4	protein phosphatase 4 regulatory subunit 4	mRNA	D	2	0	74	6
LOC106533398	cytosolic non-specific dipeptidase-like	mRNA	D	1	0	33	3
LOC106534701	protein-methionine sulfoxide oxidase mical1-like	mRNA	D	2	0	54	4
LOC106528963	hemoglobin embryonic subunit alpha	mRNA	E	1,877	565	38	3
LOC106517651	spondin-2-like	mRNA	D	1	0	42	5
LOC106522980	uncharacterized	mRNA	D	13	1	208	6
LOC106528695	major histocompatibility complex class I-related gene protein-like	mRNA	D	0	0	21	2
LOC106527792	integrin beta-3-like	mRNA	D	0	0	16	2
acsl5	acyl-CoA synthetase long-chain family member 5	mRNA	E	31	5	1	0
LOC106521886	solute carrier family 2 facilitated glucose transporter member 8-like	mRNA	D	10	0	127	9
LOC106528493	inositol polyphosphate 5-phosphatase K-like	mRNA	D	3	0	47	3
LOC106516384	ribonucleoside-diphosphate reductase large subunit-like	mRNA	E	28	3	1	0
LOC106526230	putative defense protein 3	mRNA	E	62	6	1	0

1. Tx= treatment group; each gene is identified by which treatment group has the highest expression. D = diapause (20°C) and E = escape (30°C).
2. Mean normalized expression values in FPKM of the genes specific to treatment group.
3. Standard error of the mean calculated for each average normalized expression value.

C.10: Composition statistics for Light yellow module. Heat map shows significant differential expression (DE) between phenotypes for each gene and stage of development *p* values < FDR 10% (red).

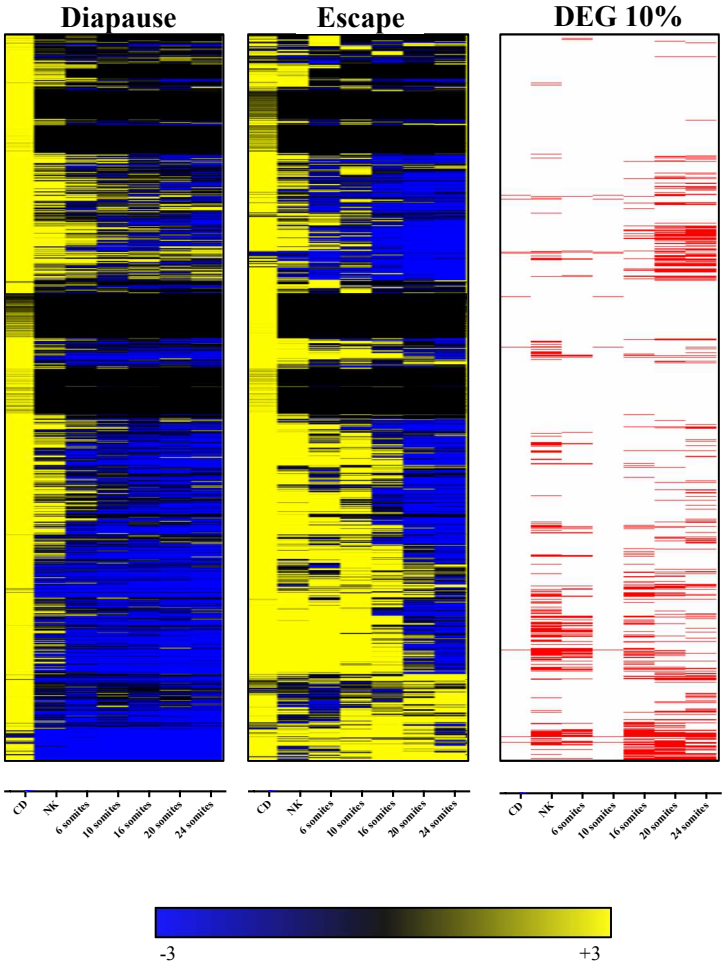
Light yellow (39 genes)		
Zsummary preservation statistic		6.2
Cluster count summary (FPKM):	**Across all stages and trajectories**	**Only 20 °C Cell dispersion**
mean	2.6	5.8
median	1	2.6
max	16.4	41.3
min	0.3	0
Hub genes	LOC106530232, LOC106525201, LOC106533814, LOC106525430, LOC106520117, LOC106534206, LOC106519593, LOC106533739, LOC106526520, LOC106521844	
Most abundant:	LOC106511859, LOC106516488, LOC106518981, *atp6v1a*, *il1b*, LOC106511208, *gnptab*, *pqlc2*, glrx5, LOC106531731	
Functional annotation:	No GO terms with an enrichment BH adjusted *p*-value <= 0.05	



C.11: Composition statistics for Turquoise module. Heat map shows significant differential expression (DE) between phenotypes for each gene and stage of development *p* values < FDR 10% (red).

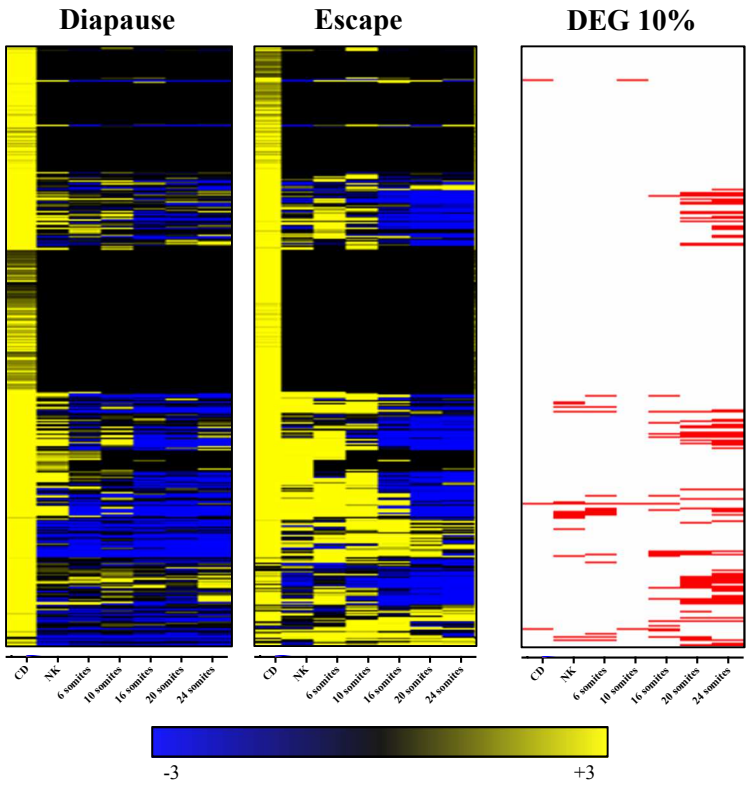
Turquoise (1,305 genes)		
Zsummary preservation statistic		24.9
Cluster count summary (FPKM):	**Across all stages and trajectories**	**Only 20 °C Cell dispersion**
mean	34.4	75.9
median	10.6	23.2
max	2,053.50	5,095.70
min	0.3	0.1
Hub genes	*acss3*, LOC106520833, LOC106521531, *tmem104*, LOC106521263, LOC106530205, LOC106534770, LOC106518686, LOC106521260, *npc2*	
Most abundant:	LOC106526003, LOC106520880, LOC106533257, LOC106523750, LOC106519310, LOC106525573, LOC106534545, LOC106526427, LOC106525414, LOC106528480	
Functional annotation:	phagosome acidification 1.04E-07 ER to Golgi vesicle-mediated transport 2.22E-07 transferrin transport 1.81E-06 ATP hydrolysis coupled proton transport 5.08E-06 positive regulation of TOR signaling 6.22E-06 carbohydrate metabolic process 1.05E-05 insulin receptor signaling pathway 1.45E-04 regulation of macroautophagy 1.43E-04 intracellular protein transport 3.61E-04 COPII vesicle coating 8.09E-04 platelet degranulation 9.81E-04 protein transport 1.13E-03 IRE1-mediated unfolded protein response 2.50E-03 lysosome organization 5.12E-03 macroautophagy 6.67E-03 proton transport 1.06E-02 ion transmembrane transport 1.50E-02 small GTPase mediated signal transduction 2.38E-02 cell cycle arrest 3.44E-02 proteolysis involved in cellular protein catabolic process 3.85E-02 cell-cell adhesion 3.78E-02 retrograde vesicle-mediated transport, Golgi to ER 4.12E-02 autophagy 4.39E-02	

C.11 Continued



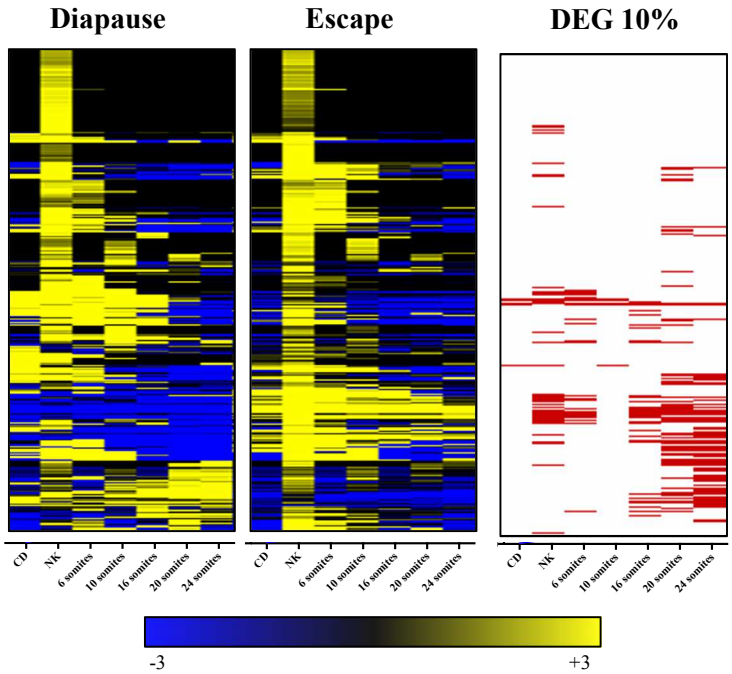
C.12: Composition statistics for Yellow module. Heat map shows significant differential expression (DE) between phenotypes for each gene and stage of development *p* values < FDR 10% (red).

Yellow (540 genes)		
Zsummary preservation statistic		25.4
Cluster count summary (FPKM):	**Across all stages and trajectories**	**Only 20 °C Cell dispersion**
mean	40.7	108.6
median	2.1	7.1
max	5,180.40	14,201.80
min	0.3	0.1
Hub genes	*mamdc4*, *dgka*, *pik3c2g*, LOC106522037, *dlg1*, LOC106516167, LOC106525246, LOC106516366, *capn9*, *nucb2*	
Most abundant:	LOC106513762, LOC106527934, LOC106520168, LOC106531475, LOC106519671, LOC106526064, LOC106533362, LOC106513763, LOC106528769, LOC106522682	
Functional annotation:	ephrin receptor signaling pathway 1.19E-02 cell-cell adhesion 1.79E-02 movement of cell or subcellular component 3.33E-02	



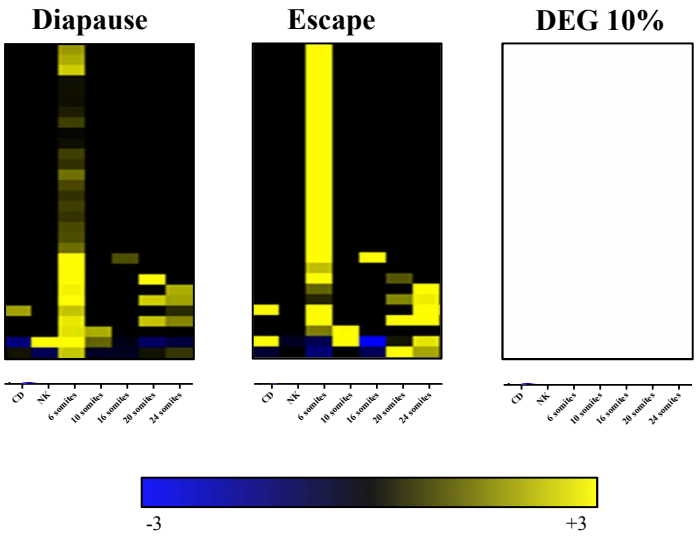
C.13: Composition statistics for Pink module. Heat map shows significant differential expression (DE) between phenotypes for each gene and stage of development *p* values < FDR 10% (red).

Pink (336 genes)		
Zsummary preservation statistic		13.6
Cluster count summary (FPKM):	**Across all stages and trajectories**	**Only 20 °C Neural keel**
mean	108.7	41
median	2.9	1.5
max	13,450.10	3,969.30
min	0.3	0
Hub genes	LOC106535526, *dnah11*, *ikzf4*, LOC106528897, LOC106521872, *luzp2*, LOC106526437, LOC106518574, LOC106528881, LOC106526435	
Most abundant:	LOC106524240, LOC106524241, LOC106530118, *hspb1*, LOC106536988, LOC106530346, LOC106528554, LOC106528458, LOC106528552, LOC106525612	
Functional annotation:	No GO terms with an enrichment BH adjusted *p*-value <= 0.05	



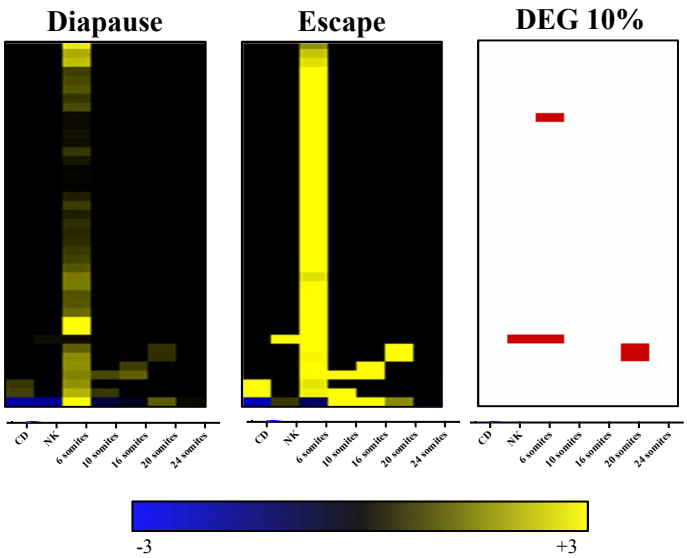
C.14: Composition statistics for Royal blue module. Heat map shows significant differential expression (DE) between phenotypes for each gene and stage of development *p* values < FDR 10% (red).

Royal blue (30 genes)		
Zsummary preservation statistic		9.1
Cluster count summary (FPKM):	**Across all stages and trajectories**	**Only 20 °C 6 somites**
mean	1	2.1
median	0.5	1.1
max	9.1	16.1
min	0.3	0.1
Hub genes	LOC106527563, LOC106531698, LOC106515784, LOC106515865, LOC106532600, LOC106530157, LOC106520243, LOC106525502, LOC106536785, LOC106533896	
Most abundant:	LOC106535129, LOC106536116, LOC106511698, LOC106512897, LOC106511907, LOC106522776, LOC106511369, LOC106536785, LOC106511351, LOC106512350	
Functional annotation:	No GO terms with an enrichment BH adjusted *p*-value <= 0.05	



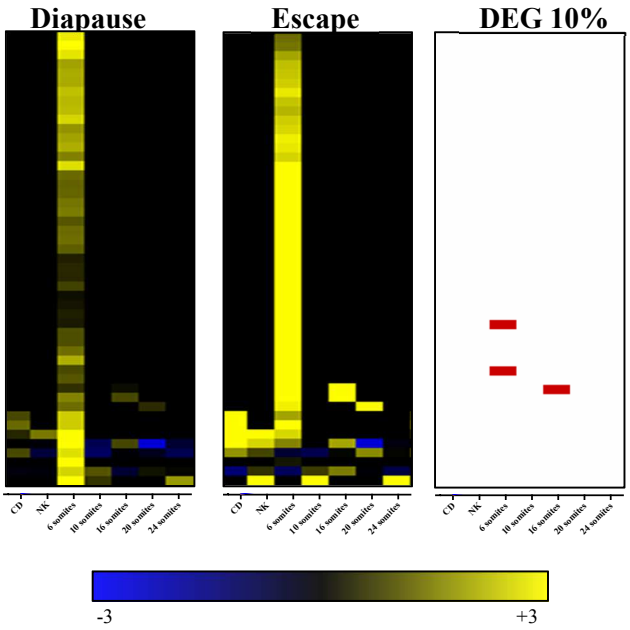
C.15: Composition statistics for Grey60 module. Heat map shows significant differential expression (DE) between phenotypes for each gene and stage of development *p* values < FDR 10% (red).

Grey60 (41 genes)		
Zsummary preservation statistic		11
Cluster count summary (FPKM):	**Across all stages and trajectories**	**Only 20 °C 6 somites**
mean	0.6	1.3
median	0.4	0.9
max	5.5	9.1
min	0.3	0
Hub genes	LOC106511015, LOC106530114, LOC106533663, LOC106511727, LOC106516671, LOC106523276, LOC106514416, LOC106527880, LOC106517025, LOC106528444	
Most abundant:	LOC106511015, LOC106530114, LOC106533663, LOC106511727, LOC106516671, LOC106523276, LOC106514416, LOC106527880, LOC106517025, LOC106528444	
Functional annotation:	No GO terms with an enrichment BH adjusted *p*-value <= 0.05	



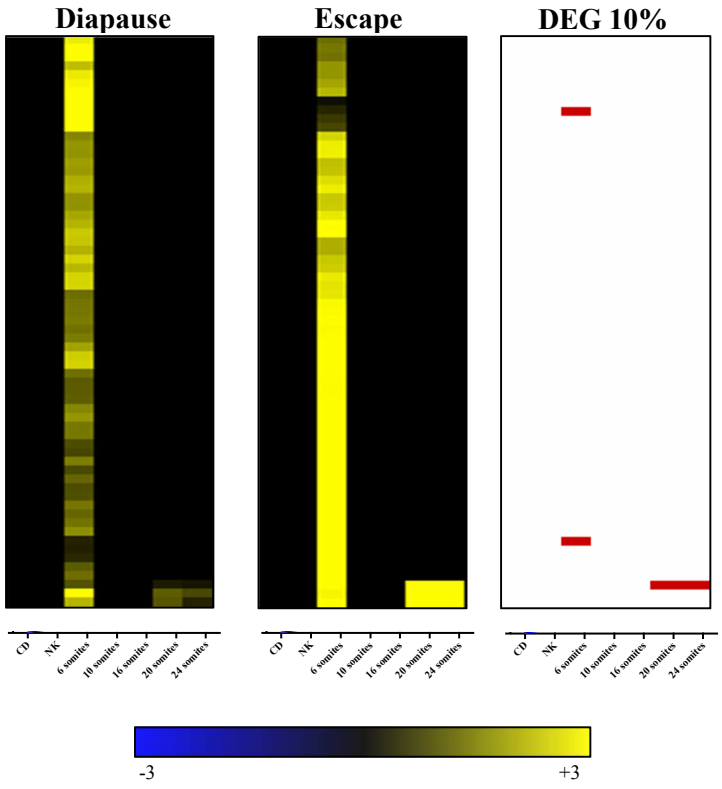
C.16: Composition statistics for Light cyan module. Heat map shows significant differential expression (DE) between phenotypes for each gene and stage of development *p* values < FDR 10% (red).

Light cyan (49 genes)		
Zsummary preservation statistic		13.1
Cluster count summary (FPKM):	**Across all stages and trajectories**	**Only 20 °C 6 somites**
mean	0.7	2.1
median	0.4	1.5
max	6.3	8.8
min	0.3	0.2
Hub genes	*ankrd11*, *spa17*, *vwa1*, *pkp2*, LOC106527183, LOC106518533, LOC106517551, LOC106513926, *trnam-cau*, *lacc1*	
Most abundant:	*bre*, *trmt12*, LOC106521124, LOC106511589, LOC106511408, LOC106529259, LOC106520601, LOC106518955, LOC106518812, LOC106533526	
Functional annotation:	No GO terms with an enrichment BH adjusted *p*-value <= 0.05	



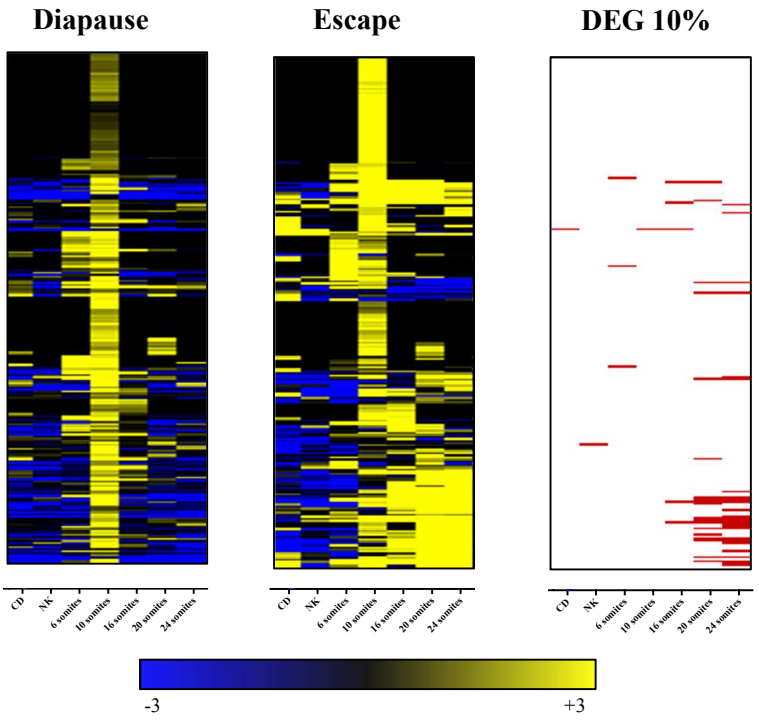
C.17: Composition statistics for Midnight blue module. Heat map shows significant differential expression (DE) between phenotypes for each gene and stage of development *p* values < FDR 10% (red).

Midnight blue (65 genes)		
Zsummary preservation statistic		9.5
Cluster count summary (FPKM):	**Across all stages and trajectories**	**Only 20 °C 6 somites**
mean	0.4	1.9
median	0.3	1.7
max	1.8	5.2
min	0.3	0.4
Hub genes	LOC106528217, LOC106529661, LOC106520451, *fbxl4*, LOC106514382, LOC106526542, LOC106530305, *slc4a1*, *tex2*, LOC106513879	
Most abundant:	*crybb1*, LOC106531830, *e2f7*, LOC106513614, LOC106528055, LOC106533921, LOC106536084, LOC106515728, *egr2*, *nox4*	
Functional annotation:	No GO terms with an enrichment BH adjusted *p*-value <= 0.05	



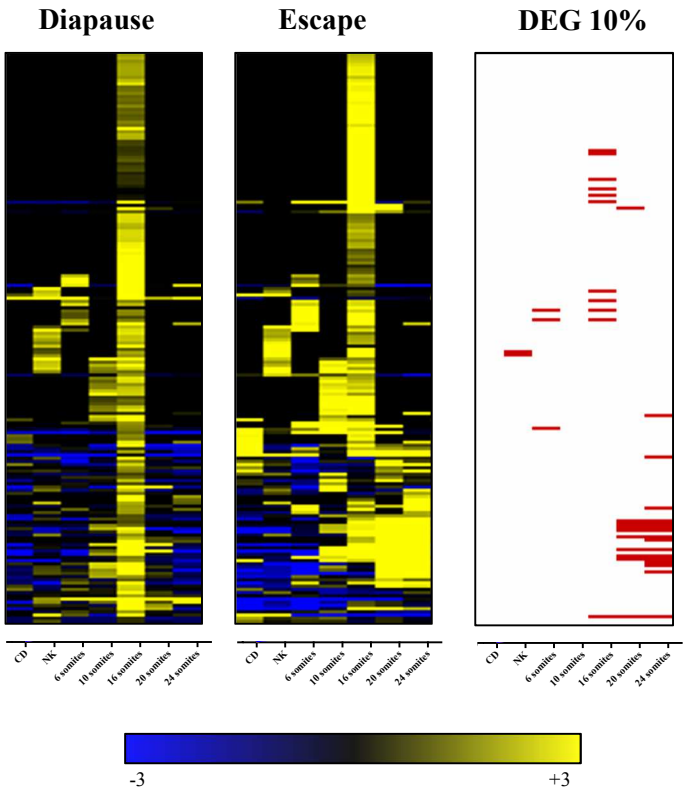
C.18: Composition statistics for Green yellow module. Heat map shows significant differential expression (DE) between phenotypes for each gene and stage of development *p* values < FDR 10% (red).

Green yellow (251 genes)		
Zsummary preservation statistic		8.5
Cluster count summary (FPKM):	**Across all stages and trajectories**	**Only 20 °C 10 somites**
mean	9.1	11.3
median	1.2	3.3
max	332.8	300.9
min	0.3	0
Hub genes	LOC106537237, *clpb*, LOC106515109, LOC106530247, LOC106531068, *bspry*, LOC106519623, LOC106516544, *zbtb49*, LOC106531718	
Most abundant:	LOC106527471, LOC106535846, LOC106515210, *rsl1d1*, LOC106511040, *pim1*, LOC106512826, LOC106532232, LOC106533706, *rrp15*	
Functional annotation:	maturation of LSU-rRNA 3.77E-02 poly(A) RNA binding 6.29E-03	



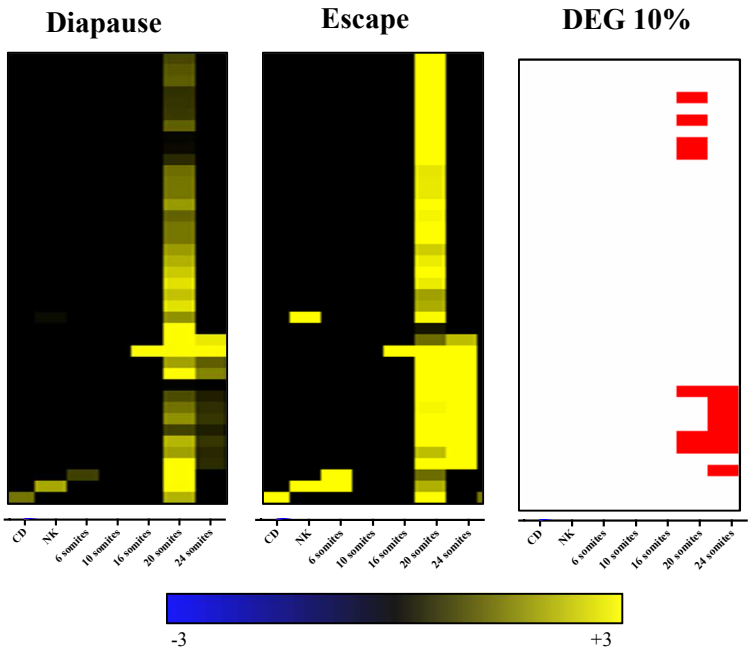
C.19: Composition statistics for Salmon module. Heat map shows significant differential expression (DE) between phenotypes for each gene and stage of development *p* values < FDR 10% (red).

Salmon (179 genes)		
Zsummary preservation statistic		8.9
Cluster count summary (FPKM):	**Across all stages and trajectories**	**Only 20 °C 16 somites**
mean	3.3	5.2
median	0.7	2.6
max	100.2	123.5
min	0.3	0
Hub genes	LOC106536978, LOC106534480, *arid5b*, LOC106525484, LOC106523099, LOC106517413, *ascc3*, *proca1*, LOC106530082, *sema3b*	
Most abundant:	LOC106536606, *ik*, LOC106527267, LOC106515020, *hibch*, LOC106518985, *haus2*, *ttf1*, LOC106533743, *qsox2*	
Functional annotation:	No GO terms with an enrichment BH adjusted *p*-value <= 0.05	



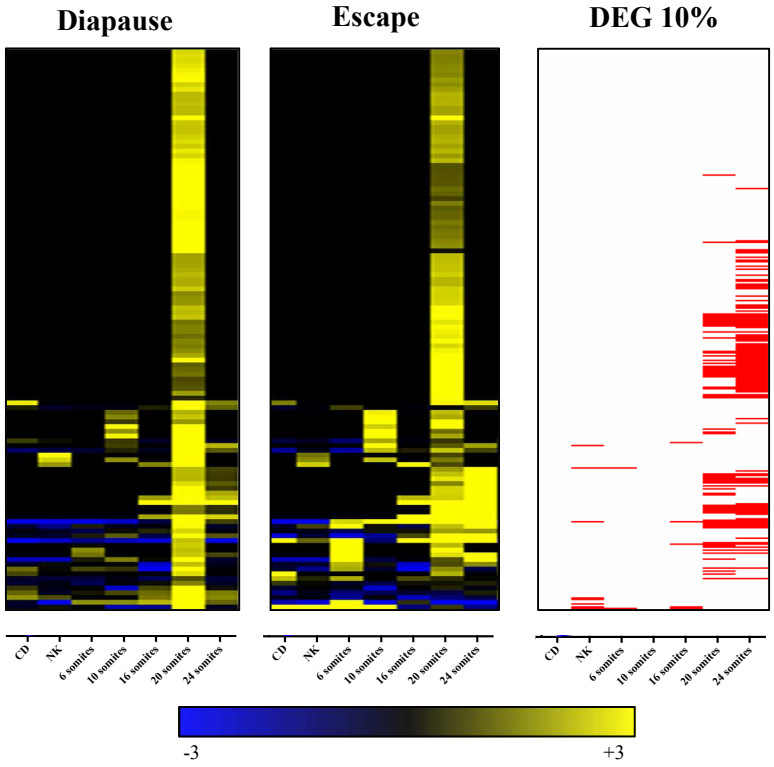
C.20: Composition statistics for Light green module. Heat map shows significant differential expression (DE) between phenotypes for each gene and stage of development *p* values < FDR 10% (red).

Light green (40 genes)		
Zsummary preservation statistic		9.2
Cluster count summary (FPKM):	**Across all stages and trajectories**	**Only 20 °C 20 somites**
mean	1	1.1
median	0.4	0
max	14.3	36.3
min	0.3	0
Hub genes	LOC106520446, afap1, LOC106527336, LOC106519571, LOC106526953, spred2, LOC106536282, LOC106536893, LOC106537054, LOC106512845	
Most abundant:	LOC106525212, LOC106534051, LOC106522867, LOC106536251, LOC106532683, LOC106522139, LOC106535787, LOC106512451, LOC106516396, LOC106534522	
Functional annotation:	No GO terms with an enrichment BH adjusted *p*-value <= 0.05	



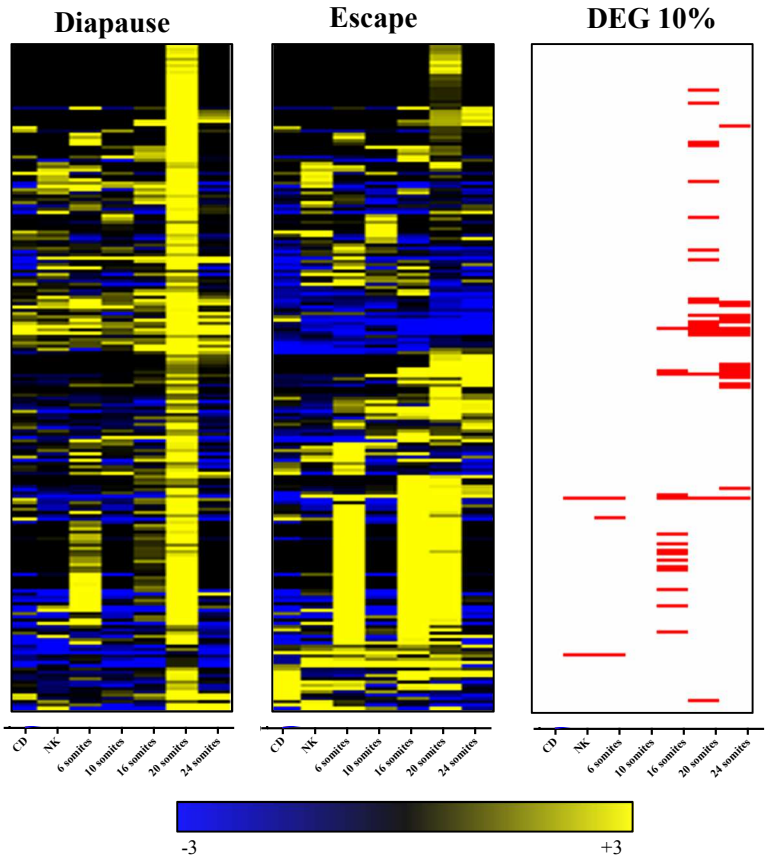
C.21: Composition statistics for Cyan module. Heat map shows significant differential expression (DE) between phenotypes for each gene and stage of development *p* values < FDR 10% (red).

Cyan (118 genes)		
Zsummary preservation statistic		16.8
Cluster count summary (FPKM):	**Across all stages and trajectories**	**Only 20 °C 20 somites**
mean	2.9	1
median	0.4	0
max	192.2	18.5
min	0.3	0
Hub genes	*prkar1b*, *slc35f2*, *avpr2*, *pkd2l1*, LOC106528218, *nek10*, LOC106511505, *adamts6*, *asb13*, LOC106530503	
Most abundant:	LOC106528894, LOC106525180, *polr2i*, *xab2*, LOC106528184, *lats1*, LOC106515689, LOC106520979, LOC106528990, *hip1*	
Functional annotation:	No GO terms with an enrichment BH adjusted *p*-value <= 0.05	



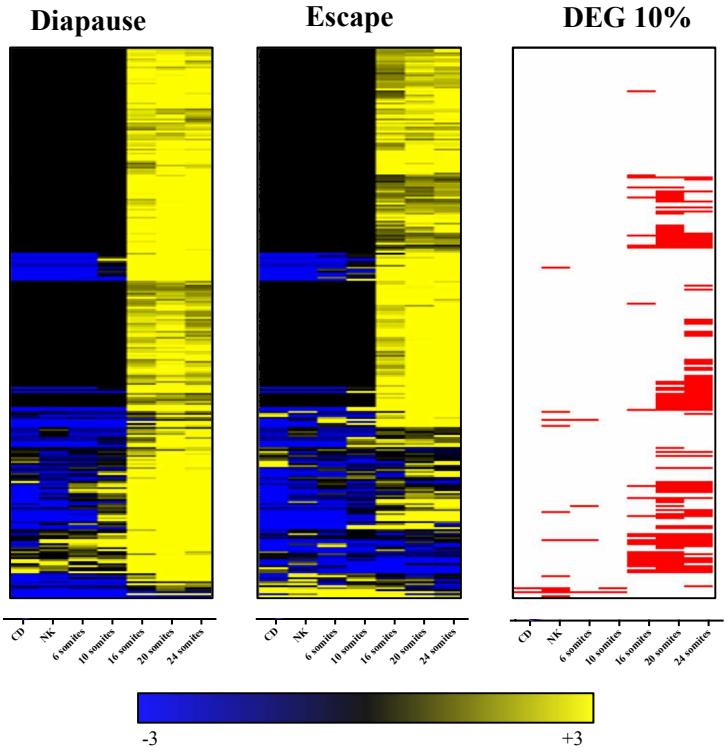
C.22: Composition statistics for Tan module. Heat map shows significant differential expression (DE) between phenotypes for each gene and stage of development *p* values < FDR 10% (red).

Tan (205 genes)		
Zsummary preservation statistic		5.5
Cluster count summary (FPKM):	**Across all stages and trajectories**	**Only 20 °C 20 somites**
mean	36.8	5.7
median	3.7	2.3
max	5,772.90	125.7
min	0.3	0
Hub genes	LOC106511466, LOC106531814, *neurod6*, LOC106518019, *acvr1c*, *pcsk1*, *fbxl22*, LOC106529194, *tmem55a*, LOC106534647	
Most abundant:	LOC106524776, LOC106515119, LOC106534647, *fbxo32*, *kiaa0513*, *tmem55a*, LOC106523305, LOC106512205, *tfdp1*, LOC106528734	
Functional annotation:	No GO terms with an enrichment BH adjusted *p*-value <= 0.05	



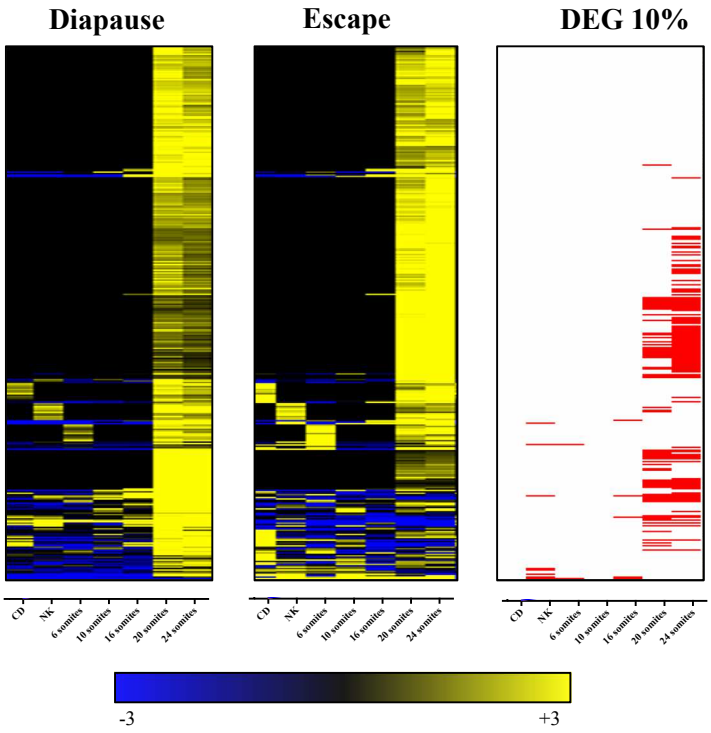
C.23: Composition statistics for Magenta module. Heat map shows significant differential expression (DE) between phenotypes for each gene and stage of development *p* values < FDR 10% (red).

Magenta (205 genes)		
Zsummary preservation statistic		29
Cluster count summary (FPKM):	**Across all stages and trajectories**	**Only 20 °C 20 somites**
mean	9.1	17.1
median	1.8	4.8
max	303.8	833
min	0.9	0.8
Hub genes	LOC106514548, LOC106536151, *loxl3*, LOC106519509, LOC106519464, *cd248*, LOC106527639, LOC106516029, LOC106529314, *mylpf*	
Most abundant:	LOC106528963, LOC106532129, LOC106525143, LOC106514906, LOC106515726, *ivns1abp*, LOC106531054, LOC106534396, LOC106516639, *mylpf*	
Functional annotation:	collagen catabolic process 1.40E-03 collagen fibril organization 1.12E-03 extracellular matrix organization 4.42E-03 cellular response to amino acid stimulus 3.06E-02	



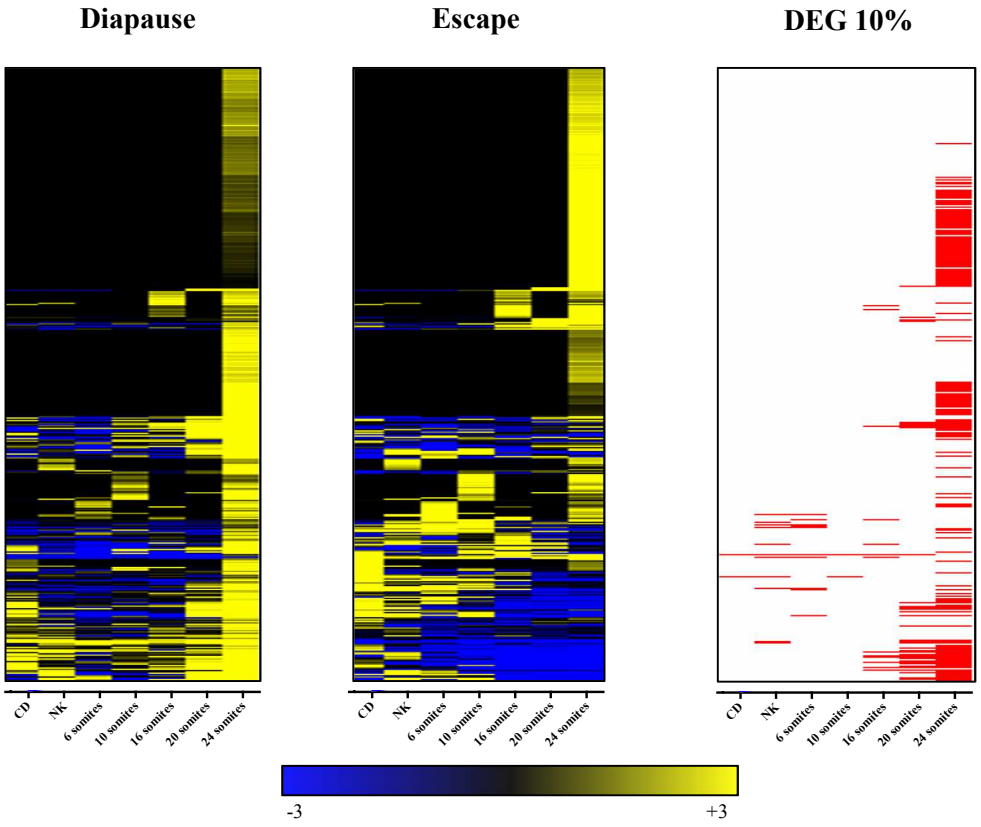
C.24: Composition statistics for Red module. Heat map shows significant differential expression (DE) between phenotypes for each gene and stage of development *p* values < FDR 10% (red).

Red (375 genes)		
Zsummary preservation statistic		31
Cluster count summary (FPKM):	**Across all stages and trajectories**	**Only 20 °C 20 somites**
mean	10.7	15.5
median	0.9	2.3
max	1,918.40	2,361.20
min	0.6	0
Hub genes	*extl3*, *tnk2*, LOC106513916, *lama4*, LOC106531418, LOC106534004, LOC106512958, LOC106521921, *tenm4*, *ebf3*	
Most abundant:	LOC106535353, *ccng1*, LOC106511248, *sepp1*, LOC106517379, *ddx21*, *mrps5*, *pnisr*, *bdh2*, *snd1*	
Functional annotation:	poly(A) RNA binding 3.48E-02	



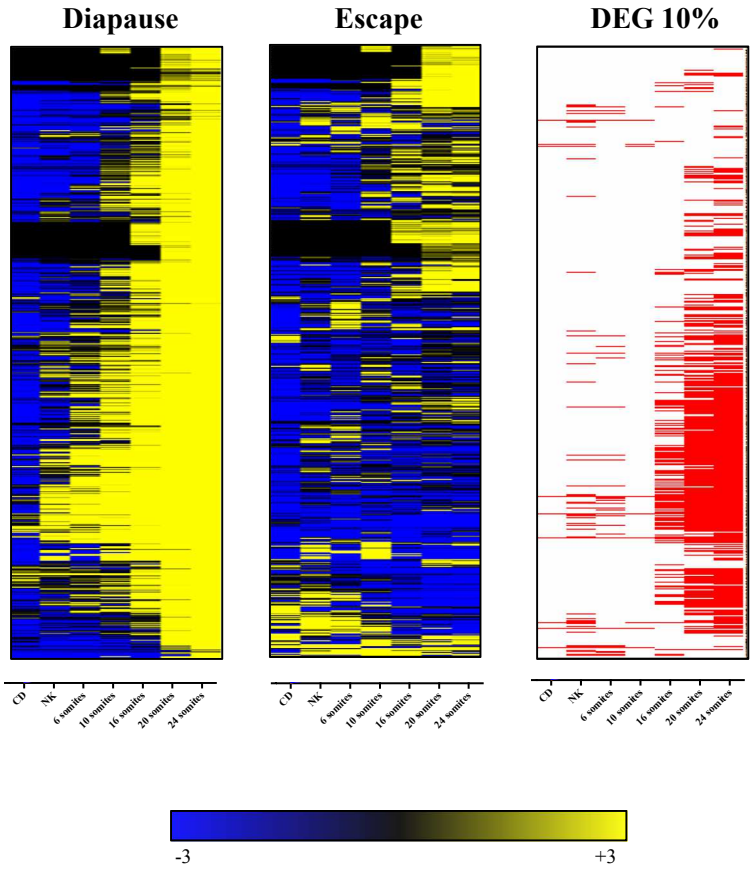
C.25: Composition statistics for Green module. Heat map shows significant differential expression (DE) between phenotypes for each gene and stage of development *p* values < FDR 10% (red).

Green (475 genes)		
Zsummary preservation statistic		21.2
Cluster count summary (FPKM):	**Across all stages and trajectories**	**Only 20 °C 24 somites**
mean	8.7	15.2
median	0.6	2.8
max	390.5	567
min	0.3	0
Hub genes	*sipa1*, *lmod3*, LOC106514489, *col6a1*, LOC106515305, *calcb*, LOC106513661 *clmp*, *sp3*, *gab2*	
Most abundant:	LOC106519604, LOC106516986, LOC106535465, LOC106522318, *ndufs6*, *pithd1*, *higd2a*, *qdpr*, LOC106518809, *pfdn6*	
Functional annotation:	No GO terms with an enrichment BH adjusted *p*-value <= 0.05	



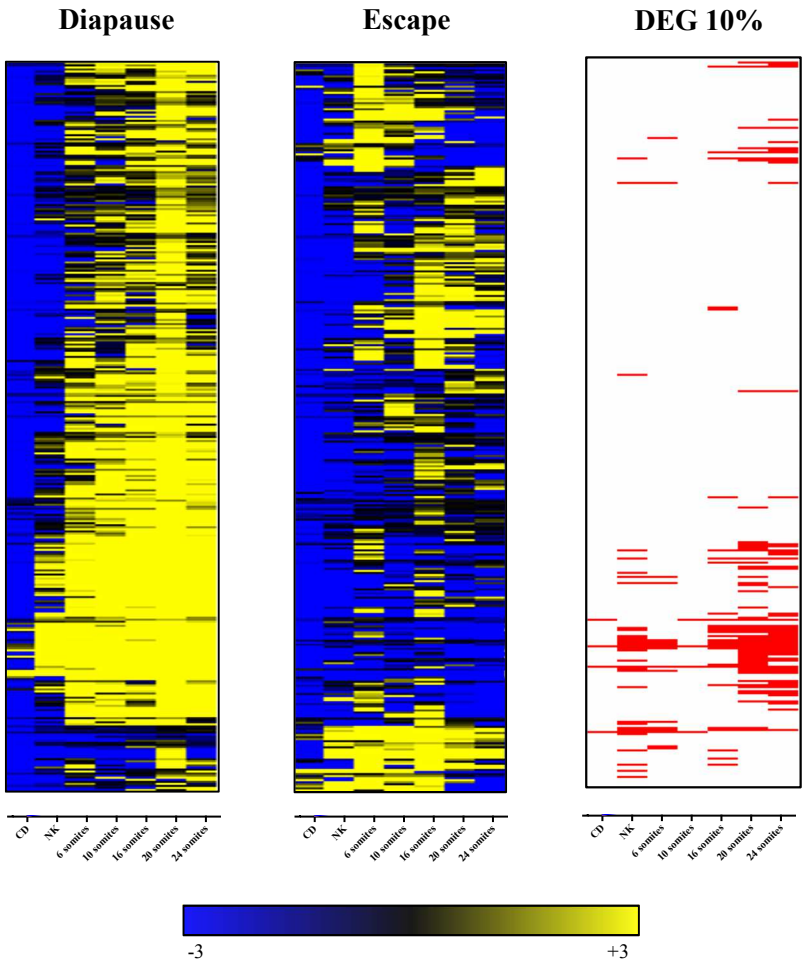
C.26: Composition statistics for Brown module. Heat map shows significant differential expression (DE) between phenotypes for each gene and stage of development *p* values < FDR 10% (red).

Brown (634 genes)		
Zsummary preservation statistic		4.5
Cluster count summary (FPKM):	**Across all stages and trajectories**	**Only 20 °C 24 somites**
mean	79.7	154.7
median	10.3	24
max	7,384.60	14,533.60
min	0.6	0.3
Hub genes	*nes*, LOC106530705, *klf11*, *tmem38a*, *gapdh*, *fgfr2*, LOC106528100, *cpamd8*, LOC106531160, *hoxd3*	
Most abundant:	LOC106535866, *rpl14*, LOC106529138, *rpl4*, *rps4x*, LOC106514001, *rpl27*, *eef1g*, *rpl23a*, LOC106535589	
Functional annotation:	protein binding 7.57E-09 poly(A) RNA binding 4.84E-06	



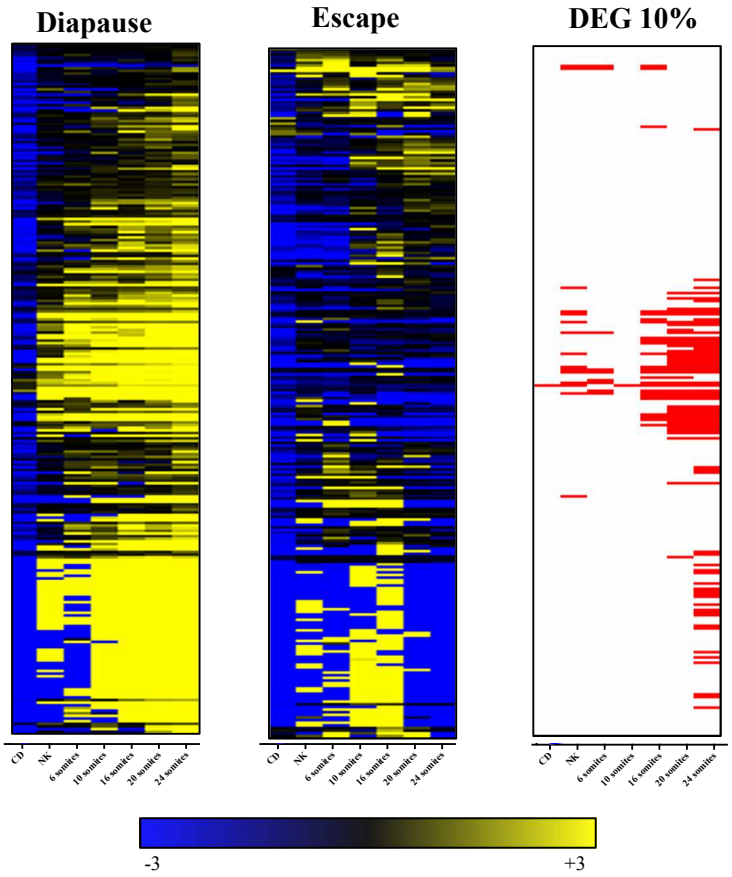
C.27: Composition statistics for Black module. Heat map shows significant differential expression (DE) between phenotypes for each gene and stage of development *p* values < FDR 10% (red).

Black (356 genes)		
Zsummary preservation statistic		11.2
Cluster count summary (FPKM):	**Across all stages and trajectories**	**Only 20 °C 20 somites**
mean	108.9	-not stage specific
median	17.1	-
max	8,112.50	-
min	1.7	-
Hub genes	*szt2*, *taf1a*, *med24*, LOC106534337, *kiaa0391*, *kif7*, LOC106520777, *trmt1l*, *snrnp70*, LOC106512744	
Most abundant:	*rpl35a*, *rplp1*, *rps20*, *rps28*, *rpl22l1*, LOC106512630, *rpl22*, LOC106512744. *snrnp70*, LOC106525147	
Functional annotation:	No GO terms with an enrichment BH adjusted *p*-value <= 0.05	



C.28: Composition statistics for Purple module. Heat map shows significant differential expression (DE) between phenotypes for each gene and stage of development *p* values < FDR 10% (red).

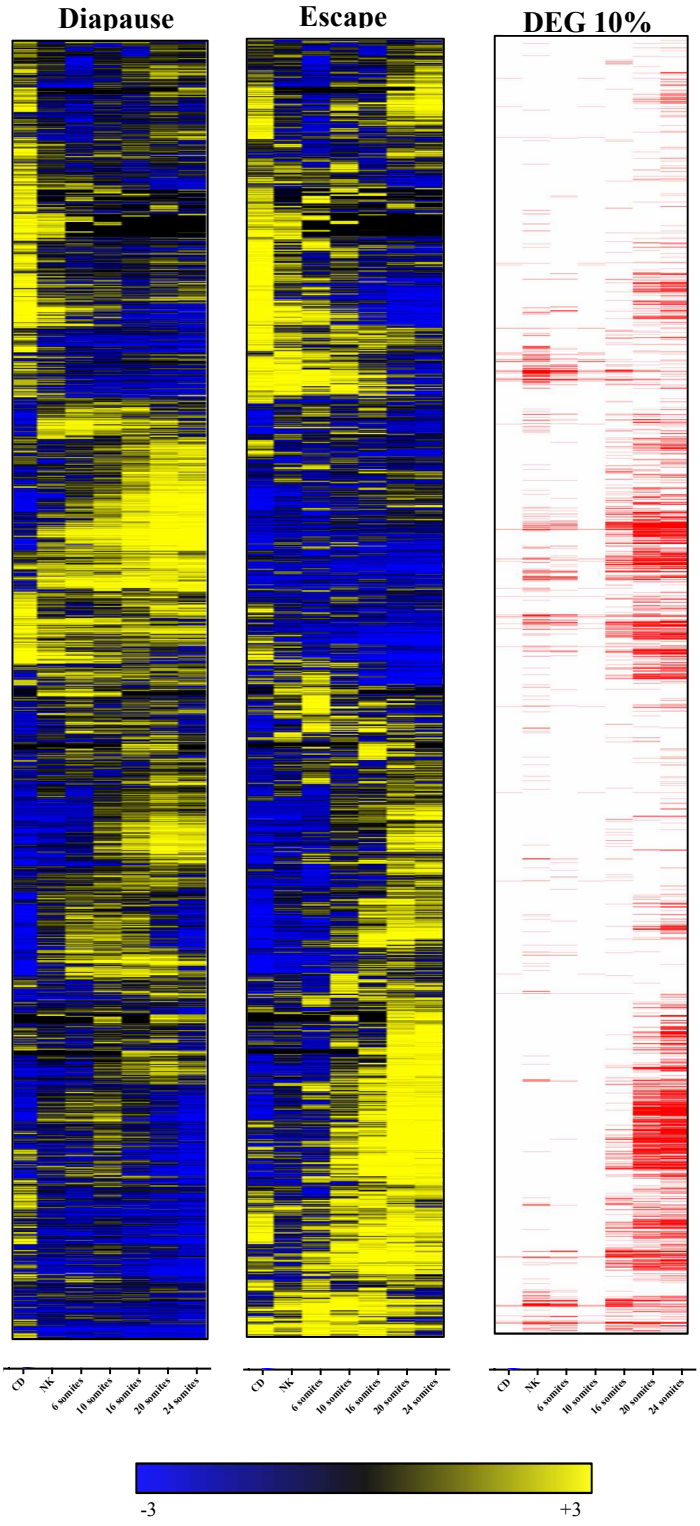
Purple (261 genes)		
Zsummary preservation statistic		12.4
Cluster count summary (FPKM):	**Across all stages and trajectories**	**Only 20 °C 20 somites**
mean	1,179	-not stage specific
median	23	-
max	9,352	-
min	2	-
Hub genes	*rpl30*, LOC106511333, *rps18*, *rpl15*, *rps3a*, *rpl24*, *rpl7a*, *rpl13*, *rpl12*, *rps8*	
Most abundant:	*rps12*, *rpl11*, *rps7*, LOC106527967, *rps8*, *rps14*, *rpl12*, *rpsa*, *rps27a*, *rpl35*	
Functional annotation:	No GO terms with an enrichment BH adjusted *p*-value <= 0.05	



C.29: Composition statistics for Grey module. Heat map shows significant differential expression (DE) between phenotypes for each gene and stage of development *p* values < FDR 10% (red).

Grey (9,474 genes)		
Zsummary preservation statistic		9.6
Cluster count summary (FPKM):	**Across all stages and trajectories**	**Only 20 °C 20 somites**
mean	31.2	-not stage specific
median	7.4	-
max	23,922.4	-
min	0.3	-
Hub genes	*pycr1*, *eftud2*, *chd4*, *itga7*, LOC106535666, LOC106519000, LOC106513161, *axl*, *ptpro*, LOC106532504	
Most abundant:	LOC106513764, *ahsg*, LOC106517545, LOC106531825, *apoa4*, LOC106529351, LOC106519283, LOC106526789, LOC106512994, LOC106512408	
Functional annotation:	No GO terms with an enrichment BH adjusted *p*-value <= 0.05	

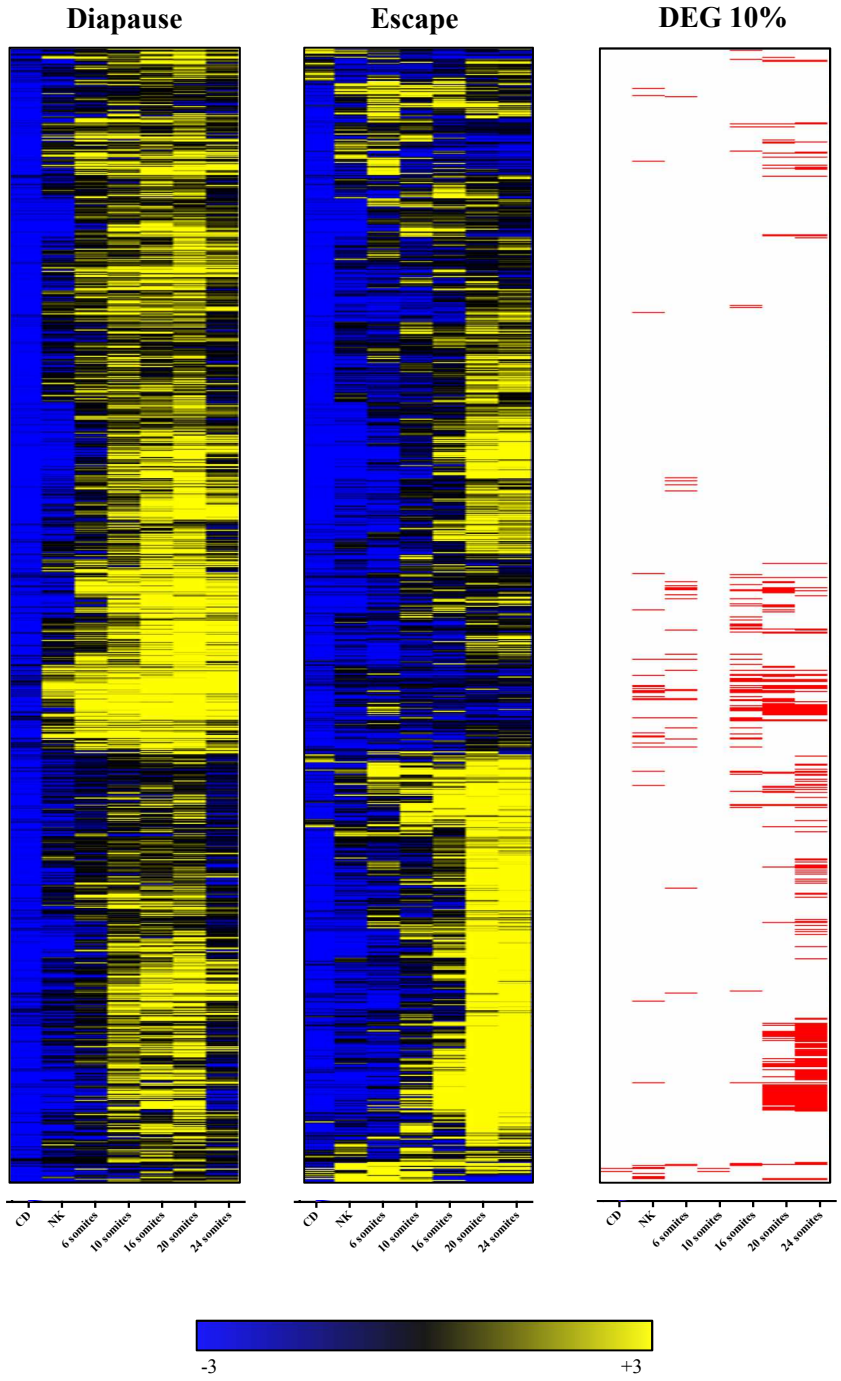
C.29: Continued



C.30: Composition statistics for Blue module. Heat map shows significant differential expression (DE) between phenotypes for each gene and stage of development *p* values < FDR 10% (red).

Blue (1,126 genes)		
Zsummary preservation statistic		18
Cluster count summary (FPKM):	**Across all stages and trajectories**	**Only 20 °C 20 somites**
mean	13.4	-not stage specific
median	6.3	-
max	530.4	-
min	1.0	-
Hub genes	*robo1*, *srgap1*, LOC106518367, LOC106516884, *jag1*, LOC106522850, LOC106527329, LOC106528155, *smarcd3*, *boc*	
Most abundant:	LOC106512448, LOC106531614, LOC106511047, *fabp3*, *pck1*, LOC106533361, *hmgn1*, LOC106526031, LOC106536610, LOC106526939	
Functional annotation:	No GO terms with an enrichment BH adjusted *p*-value <= 0.05	

C.30: Continued



APPENDIX D:

Supplemental Files

Supplemental File 2-1: The poly-A transcriptome of *A. limnaeus* 1-2 cell stage embryos. FPKM = fragments per kilobase of transcript per million mapped reads. Gene descriptions come from the *A. limnaeus* annotated genome. Microsoft Excel ©2017 spreadsheet; 1.4 MB.

Supplemental File 2-2: The sncRNA transcriptome in 1-2 cell stage embryos of *A. limnaeus*. CPM = copy per million. Alignment information is in regards to the nearest gene or its position within the gene or genomic region. Microsoft Excel ©2017 spreadsheet; 930 KB.

Supplemental File 2-3: Number of genes designated in the minimal stress proteome identified in the *A. limnaeus* 1-2 cell stage transcriptome. Microsoft Excel ©2017 spreadsheet; 734 KB.

Supplemental File 3-1: Potential mRNA gene targets for microRNA miR-430 binding in *A. limnaeus* genome. Microsoft Excel ©2017 spreadsheet; 113 KB.

Supplemental File 4-1: Poly-A transcriptomes along temperature-induced development of *A. limnaeus*. Transcriptomes are columns A1 through N3 and represent each sample library (n=42) in FPKM values. Microsoft Excel ©2017 spreadsheet; 10.5 MB.