University of New Orleans
ScholarWorks@UNO

University of New Orleans Theses and Dissertations | Dissertations and Theses

Fall 12-20-2019

A Transcriptomic Exploration of Hawaiian Drosophilid Development and Evolution

Madeline M. Chenevert
University of New Orleans, mmchenev@uno.edu

Follow this and additional works at: https://scholarworks.uno.edu/td

Recommended Citation

Chenevert, Madeline M., "A Transcriptomic Exploration of Hawaiian Drosophilid Development and Evolution" (2019). *University of New Orleans Theses and Dissertations*. 2687.
https://scholarworks.uno.edu/td/2687

This Thesis-Restricted is protected by copyright and/or related rights. It has been brought to you by ScholarWorks@UNO with permission from the rights-holder(s). You are free to use this Thesis-Restricted in any way that is permitted by the copyright and related rights legislation that applies to your use. For other uses you need to obtain permission from the rights-holder(s) directly, unless additional rights are indicated by a Creative Commons license in the record and/or on the work itself.

This Thesis-Restricted has been accepted for inclusion in University of New Orleans Theses and Dissertations by an authorized administrator of ScholarWorks@UNO. For more information, please contact scholarworks@uno.edu.

A Transcriptomic Exploration of Hawaiian Drosophilid Development and Evolution

A Thesis

Submitted to the Graduate Faculty of the
University of New Orleans
in partial fulfillment of the
requirements for the degree of

Master of Science
in
Biological Sciences

by

Madeline Chenevert

B.S. Agnes Scott College, 2016

December, 2019

Acknowledgements

It didn't take three and a half years researching Hawaiian biology to know that no man is an island. Biology is a field of deeply interconnected topics and people. I owe a great deal of thanks to my advisors: Dr. Anthony for her support at the beginning and end of my time here as a professor and an editor, Dr. Clancy for her understanding guidance and encouragement, and Dr. Atallah for welcoming me into his lab as we both started new chapters of our lives. Thanks to Bronwyn Miller, my first academic sibling, for the countless hours she spent helping me with troubleshooting and computational analysis. Our lab's computer Science graduate students, Jack and Manisha helped me with several coding errors I forgot from my first semester and helped me keep moving on my projects. I had the pleasure of working alongside several undergraduates, so thank you Kaci, Arielle, Yasmine, Rayce, Jacob, Delmy, Bianca, Kristen, Dana, Nichole, Mary, Khaquan, and Cydney. The lab has grown so much since I got here, but the chaos of a full lab feels like home. Thanks to Anna and Jessie, thanks for lending their hands in help on the good days and ears in commiseration on the bad ones. Ahmad Karkoutli was a huge help in the lab and in class and has been a great friend beyond both.

Thanks to Mariana from BioBam for all the technical support during Blast2Go analysis. Dr. Susan Lott worked with Dr. Atallah on the work I got to continue, and Dr. Don Price has coordinated with our lab to help us navigate the field and collect our specimens. Their collaboration has helped us make strides in our lab.

My family has been an incredible support system before and after graduate school. My sisters Camille, Gabrielle, and Komal have been my lifelong friends (because they had to). I couldn't have picked them better myself. Thanks to my parents for their continued support of my academic career and teaching me to follow my passions. Finally, I'd like to thank Contessa, who must have been the nicest surprise in a long journey of surprises. She's my partner in everything but crime because we are a boring and law-abiding couple. She helped me face the challenges in my way to become more confident in myself and in what I do.

I'm grateful to have been surrounded by a supportive community with the patience to help me grow through this experience. I couldn't have done it without y'all.

Table of Contents

List of Figures

List of Tables

Abstract

One in four known species of fruit flies inhabit the Hawaiian Islands. From a small number of colonizing flies, a wide range of species evolved, some of which managed to reverse-colonize other continental environments. In order to explore the developmental pathways, which separate the Hawaiian *Drosophila* proper and the *Scaptomyza* group that contains reverse-colonized species, the transcriptomes of two better-known species in each group, *Scaptomyza anomala* and *Drosophila grimshawi*, were analyzed to find changes in gene expression between the two groups. This study describes a novel transcriptome for *S. anomala* studies as well as unusual changes in gene expression in *D. grimshawi* relative to other species, revealing priorities of both species in early development.

Keywords:

Transcriptomics, biodiversity, evolution, drosophila, developmental biology

Background

Introduction to the Maternal Zygotic Transition

Worldwide, there are around 4,000 documented species of *Drosophila* fruit flies, and nearly 1,000 of these species are endemic to the Hawaiian Islands [1]. While only a small group of flies initially colonized Hawaii, the colonizing groups radiated, resulting in a rich variety of species and an excellent model system with which to study island endemics and adaptive radiation. Over the course of about 25 million years, these species have evolved into several different clades to suit different niches [2]. These groups of flies show diverse genetics, morphology and behavior across all stages of life, from egg to adult. Two clades, *Scaptomyza* and Hawaiian *Drosophila* proper, diverged in Hawaii. Eventually, the *Scaptomyza* genus colonized other Pacific islands and coasts [3]. The date of this divergence is debated, as is the manner of Hawaii's initial colonization. Island endemic species usually show a high degree of specialization and are particularly susceptible to extinction upon changes to their environments. As such, it is uncommon for an island endemic to leave the island, let alone successfully colonize other environments [4]. Molecular analysis may provide insight into what drove adaptations that made *Scaptomyza* a hardy colonizer and the Hawaiian species so diverse. This analysis, accompanied by published phylogenies and theories explaining how the islands were initially populated, can help elucidate how this species richness developed, and how the *Scaptomyza* flies were able to leave Hawaii and colonize new landscapes.

Over the course of the study of evolution, naturalists have increasingly used molecular techniques, and other advanced methods, to create phylogenies to map the changes in and development of species[5]. Species originally studied through visible morphological traits can now be analyzed at molecular levels. Techniques continue to improve, and new genetic information can narrow in on evolutionary time. Genomic information can highlight the development of mutations that could have led to species divergence, but not all changes would be immediately evident[6]. Upregulation and downregulation of genes can occur, causing physiologically evident changes without making large changes to the coding region of the genome[6]. Highly conserved genes will not mutate as quickly as less conserved genes, as damage to conserved genes can cause infertility or death. Non-coding genetic changes prove less harmful and may occur in sites that regulate gene transcription. While genomic regulation can be tracked using transcriptomic methods, focusing on one life stage or tissue is very important, since regulation differs across cells. Since many factors that could contribute to changes between species, such as fecundity, substrate, and developmental rate, all occur at early developmental stages, genomic regulation in early embryonic development may have contributed to species differentiation and success in a way that is not evident in protein-coding sequence changes.

In the earliest stages of development, the transcriptome is young, and tissues have yet to differentiate. Diversity in the transcriptome is low between cells and even individuals, since many genes, such as those involved in sex differentiation, have yet to be transcribed. While the early transcriptome is consistent within a species, it can vary between species due to changes in regulatory sequences. For instance, a transposable element of a few base pairs may insert itself into a region upstream of a gene and change a motif that acts as a transcription factor target, interfering with normal transcription of a gene. Since mutations driving changes in gene regulation occur more quickly than those in protein-coding genes, a transcription-based process early in development would be optimal to observe diversity in genetically similar species.

With few exceptions, every animal begins development as a single-celled zygote[7]. At this stage, the zygote is not yet transcribing its own genes or translating its own RNA. To give the egg resources to begin development, the mother must deposit factors into the zygote to begin transcription and translation[8]. While many animals add nutritional resources to provide energy for zygotic development, nutrition alone will not jump-start genetic replication. Maternally deposited RNA can be translated into proteins vital to development, activating zygotic genes (see below) or playing catalytic roles of their own[9].

At a certain stage, the embryo has developed enough of its own cellular machinery that it no longer needs the assistance of maternally deposited RNA. At this point, the embryo will break down the maternal deposits, using proteins and miRNAs is made on its own, as well as those deposited by the mother[9,10]. This process is referred to as the maternal-zygotic transition (MZT). In *Drosophilids*, the MZT can be tracked at easily identifiable life stages which are comparable across species[11,12]. Conserved reference points, before and after the maternal-zygotic transition, can be used to compare species and look for divergence in the maternally supplied and zygotically transcribed RNA transcripts[11].

Shortly after fertilization, *Drosophilid* eggs begin a series of nuclear divisions which can begin before the egg is laid. This study uses previously defined, highly conserved stages rather than mitotic divisions to estimate zygotic transcription levels, as they have easily identified physical characteristics which correspond to known levels of zygotic transcription [12]. Prior studies have shown that throughout species, the physical markers indicating fly embryo stages are consistent, even if the size and scale of developmental time vary [13].The egg reaches Stage 2 (St2) around the fifth cleavage cycle. The embryo forms a syncytium, a group of nuclei not separated by cell membranes and able to share organelles, transcription factors, and transcripts. At this point, the RNA transcripts are primarily maternal [14]. At 14 cleavage cycles, the egg reaches Stage 5 (St5), where the syncytium closes off. The nuclei have migrated to the sides of the egg and cell membranes form between them [12]. At this point, the transcripts are primarily zygotic. Zygotic transcription is well underway, but some maternal transcripts have yet to be degraded [14].

When comparing transcriptomic samples of St2 and St5, four possible results can indicate the origin and the fate of a transcript[11]. If a transcript is present in St2 and absent in St5, it was maternally deposited, but degraded by the zygote. If a transcript is present in St2 and less abundant in St5, the transcript was partially degraded. If a transcript is present in St2 and present in equal or greater portions in St5, it is both maternally deposited and zygotically produced. If the transcript was absent in St2 but present in St5, it is only produced in the zygote. Knowing when each transcript is produced can give insight on gene activation and maternal deposition. The time at which a gene is active can provide information on the role it plays in development.

During the MZT, development is regulated by a cascade of genes that incrementally build up the layout of the body [7]. Maternally deposited transcription factors lead to the activation of the gap genes, which are expressed in a localized pattern to create broad divisions of the embryo [7]. The gap genes activate pair-rule genes, which appear in the embryo as seven bands [7]. The pair-rule genes activate segment polarity genes, which polarize each segment and divide the embryo into 14 segments [7]. Gap, pair-rule, and segment polarity genes can activate homeotic, or Hox genes, which begin the differentiation of the embryonic segments into individual organ systems [7]. Hox genes cannot be transcribed until upstream genes activate them. The MZT is the beginning of a series of interlinked genes, proteins, and processes which are vital to the initial layout of the body.

Analysis of the MZT can show interesting patterns within a species, highlighting which genes play roles in the early development of a fly[11]. The enrichment of specific gene isoforms and ontologies

can reveal the developmental priorities and processes to improve the understanding of the embryology of that species. When compared to another species, comparison of gene representation during the MZT can yield insight into the species' development and evolution.

Diversity Across Hawaii

When organisms colonize an island environment, they can take advantage of new niches and resources to diverge into new and diverse species[15]. Islands regularly show higher abundance of species and more variations in traits within those species than do nearby continents[15]. In the absence of competition and predation, island-endemic species often lose many high-energy adaptations used for protection or escape. After these losses, they are usually not able to survive migrations back to older, more populated areas, especially mainland environments. Hawaii is home to nearly 1,000 of the world's 4,000 *Drosophila* species, each with unique and often elaborate adaptations[1]. These species have been studied as a prime example of island endemics[1]. Hawaii's Drosophilids are divided into two clades: Hawaiian *Drosophila* proper and *Scaptomyza*. Most islands tout rather delicate endemics at high risk of extinction[4,16]. Hawaii is no different in this sense, as many of its *Drosophila* species are endangered. Hawaii is different, however, in that *Scaptomyza*, while an endemic group, has managed not only to leave the islands and survive, but also diverge into a large genus on all continents except Antarctica[2]. An exploration into gene expression may reveal adaptations at the developmental level which allowed for *S. anomala's* hardy nature.

Members of the Hawaiian *Drosophila* tend to be larger, slower, and more adapted to specific diets and habitats than those of the *Scaptomyza* genus[3]. While the Hawaiian *Drosophila* proper are limited to the Hawaiian Islands, they show a wide array of adaptations[17]. Within the group are four clades, all with different adaptation types: AMC, modified mouthpart, Halekaele, and picture-wing flies[17].

The AMC clade is composed of approximately 95 species and three subgroups: *antepocerus*, modified tarsus, and ciliated tarsus [18]. Within the modified tarsus group, there are split-tarsus, bristle tarsus, and spoon tarsus species[19]. Internally, the species within the modified tarsus subgroup show little variation and are closely phylogenetically related. These subgroups differ primarily in the sexually dimorphic modifications males carry in the tarsi of their forelimbs[19]. Ciliated tarsus flies have particularly long setae on male forelimbs[17]. These modifications likely play a role in male courtship and are results of sexual selection. The modified mouthparts clade also shows modified mouthparts indicative of sexual selection[17]. Males in this group show ornamental changes in the bristles and projections around their labella[20].

Halekaele flies are identified by their shiny, dark, and slender bodies [21]. Males in this species typically lack anal sclerites [21]. Halekaele flies primarily breed on mushrooms, rather than on leaves or bark, though the exact substrates for each species are not all defined [21]. This 51-species subgroup is smaller and has less genomic information and fewer laboratory specimens than the others [21]. They were originally categorized as a more basal group, but mitochondrial DNA evidence suggests they are a sister group to the AMC clade [22].

The picture-wing clade contains 116 known species, characterized by dark brown markings on the wings [3,23]. Picture-wing flies tend to be larger than the average fruit fly, often measuring a centimeter or longer [3]. While in some species, such as *Drosophila grimshawi*, both sexes show similar wing patterns, markings can also be sexually dimorphic [23].

These clades have diversified through sexual behavior, dietary changes, and the substrates on which they feed and reproduce. This diversity is represented through body size, coloration, wing patterns, and developmental morphology such as ovariole number, egg size, and structure [24]. Many of these adaptations have reproductive and developmental roles, and other developmental patterns may show changes which parallel *Drosophila's* island adaptations. By comparing the developmental genes

represented in different species, an intersection of genetics and physiology can add to the body of information aiming to uncover evolutionary timelines. Picture-wing flies, which are relatively well-studied and easier to trap than the other Hawaiian flies, are particularly useful for this type of research program.

Picture-Wing *Drosophila*

Picture wing *Drosophila* were originally classified by host plant, mating behavior, and male genitalia, but recent genetic studies have allowed more specific classification[25]. Picture wing *Drosophila* fall into five major species groups: *adiastola*, *plantibia*, *picticornis*, *nudidrosophila*, and *grimshawi*, not to be confused with the individual *D. grimshawi* species[17]. The diversification of the groups was rapid, starting around 4.7 MYA, and driven first by host plant specification, and then by geographical separation across the islands[26]. *Adiastola* is the most basal group[26]. *Picticornis*, *nudidrosophila*, *plantibia*, and early *grimshawi* groups began diverging around 3.8 MYA[26]. A second divergence from 2.1-3.1 MYA resulted in the *grimshawi* group proper[26]. The first divergence occurred when Kuai was the highest island, and the soil had eroded enough to allow floral growth[26]. The first species adapted to the new growth of plants and have maintained specificity to their host plants as evolution continued[26]. As the islands aged, plant growth spread, and the *Drosophila* followed[26]. As the groups spread across the islands, pre-mating reproductive barriers grew, further diverging the species[26].

The species of each group usually breed on a specific type of plant[17]. While the adult fly's diet is not as specific, specialized host species are required to induce egg-laying and promote larval growth. Picture wings typically breed in the bark of Hawaiian trees, but a few species can breed on monocots or in the sap flux of the trees[17]. *Adiastola* flies lay eggs only on *Campanulaceae* plant species and *plantibia* breed primarily on *Araliaceae*[17]. *Picticornis, Nudidrosophila,* and *Grimshawi* show more diversity in host plants within their species groups, but only four *Grimshawi* species are considered generalists within Hawaii [26].

A previous study clustered the species according to gene expression levels within the transcriptome[11]. When compared to the phylogenies constructed using traditional genomic methods, some species were placed into new subgroups, forming connections differently from previously established clades[11]. The expression patterns in zygotic-only samples also revealed a set of highly conserved genes represented across several species [11]. If *S. anomala*'s highly represented genes show more similarities to those of mainland species or to those of *D. grimshawi*, timelines of divergence may be inferred on species divergence. Transcriptomic information may show previously unexplored connections between species that may provide new insight into the functions that drove Hawaiian *Drosophilid* diversification.

Comparisons in the Lab

This study starts with *D. grimshawi* as a picture-wing specimen because thanks to their role in a 12-genome generation project, they are currently the most studied picture-wing species[27]. They can be cultured in the lab with relative ease, as their generalist nature compared to other picture-wings allows them to survive on lab-made media. *D. grimshawi* is the only picture wing species with a complete genome available. Since transcriptomes need to be mapped back to genomes during analysis, this availability allows the most accurate map possible[27]. While *D. grimshawi* is an excellent model for Hawaiian flies, and particularly picture wings, few genetic studies have focused on them. Further genetic and transcriptomic analysis will make them better model organisms.

D. grimshawi is one of the more prolific species in Hawaii. It is one of the few picture-wing species native to three islands and has a generalist nature compared to other Hawaiian *Drosophila* species: it

has a wide range of host plants and can adapt to a slightly wider range of environments[26]. Despite its relative versatility, it still shows many of the energy-saving adaptations common to island endemics. They are large, slow, don't have many defenses against predation, and never migrated farther than the next island over. While these traits may have occurred as mutations gathered in coding regions of genes, modifying the amino acid sequences of proteins, non-coding sequences evolve more rapidly[28]. They do not impact protein function but can change which genes are activated at certain times and in certain tissues by mutating transcription factors. Relaxed selection factors may have aided in the accumulation of these changes. A study of a uniform set of tissues can show patterns which may reveal changes from mainland and *Scaptomyza* species, such as the patterns behind the loss of defensive functions. In early development, the maternal-zygotic transition can show which genes are represented before cells begin to differentiate. An analysis of zygotic-only gene representation offers a comparison to *Scaptomyza anomala,* a species present in both Hawaii and Japan, to determine what genes *S. anomala* upregulates or downregulates to survive in a wider range of environments.

Introduction to *Scaptomyza* Flies

Scaptomyza Drosophilids diverged from the Hawaiian *Drosophila* proper as early as 35 MYA, before the formation of the current high islands in the archipelago (Alternative theories suggest that *Scaptomyza* arose from by a second colonization event, separately from *Drosophila*)[2]. Island endemic species tend to grow highly specialized to one environment and can easily die off when the environment is disturbed[16]. *Scaptomyza* flies appear to have overcome this evolutionary "dead end"—around 20 MYA, some Scaptomyza species emigrated from Hawaii to colonize neighboring pacific islands and continents[2]. The Hawaiian *Drosophila* species only managed to migrate to neighboring Hawaiian Islands.

Scaptomyza might have met success in populating new areas simply because their morphology allows them a better opportunity to compete with local species than *Drosophila.* Hawaiian *Drosophila* tend to be slower and larger[3]. *Scaptomyza* average 2.5 mm in length, and *Drosophila* averaged 3 mm at the times they began migrating across the Hawaiian Islands[3]. *Scaptomyza* require less food and can therefore be transported on a smaller volume of substrate. Rather than laying eggs on bark and consuming yeast, many *Scaptomyza* adapted to herbivory [29]. *Scaptomyza* no longer needed yeast or other microorganisms to eat and could directly eat their host plant. The new variety of substrates, such as the *Pisonia* fruit, are smaller, stickier, and more portable than bark[29]. These substrates could stick to migratory birds to spread larvae to new environments [29]. Other plants such as sandalwoods have spread across pacific islands to Hawaii, so it is possible *Pisonia* followed a similar pattern[30].

New substrates and small body sizes made *Scaptomyza* more likely to leave Hawaii than Hawaiian *Drosophila,* but certain traits allowed them to thrive in new environments on top of simply leaving their initial ones[25]. *Scaptomyza* have shorter lifespans than Hawaiian *Drosophila*—in the laboratory, they develop from egg to adult in about 15 days, while Hawaiian *Drosophila* may take 23-29. Generalized oviposition and fast development allowed for more offspring to be produced at a rapid rate. The short generation time gave *Scaptomyza* more opportunities to evolve traits suited to their new ecosystems.

Addressing the roots of diversification

Two groups of thinking address the initial population of Hawaii: the single origin theory and the multiple origin theory [31]. The single origin theory posits that both genera originated from a single common ancestor. The multiple origin theory suggests one or more additional initial population events, with the two genera diverging before their arrival on the island. Mitochondrial DNA sequences support the single origin theory, indicating that *Scaptomyza* and *Drosophila* are sister clades that diverged from a single common ancestor around 23 MYA. The genera are separated by 99 changes within 134 sequences from a single common ancestor [3,31]. The species that colonized other islands, as well as every continent but Antarctica, diverged between 12-3 MYA.

While populations started from such small numbers usually succumb to the founder effect, Hawaiian *Drosophila* showed an unexpected diversification. A few hypotheses point to Hawaii's volcanic nature as a source of diversity [32]. Since Hawaii was made up of newly-formed land with low species diversity, early Hawaiian *Drosophila* likely experienced no predation pressure or competition [4]. The heat and ash from the volcanoes activated heat shock proteins which enabled transposable elements to move around the genome [32]. This may have given rise to new mutations that caused rapid change, allowing the *Drosophila* to quickly adapt to the abundant new space [32].

These species vary from other Hawaiian Drosophilids not only on genetic, but also on physiological and ecological fronts. *Scaptomyza* flies can feed on diets of yeast and bacteria, but some species have

adapted to leaf-mining live mustard species, and others exhibit carnivory of insects and spider egg sacs [2,33]. When a new species evolves on a volcanic island, they do so with fewer predators and competitors. Theoretically , this means that they are not suited to compete with populations on previously-populated land, but in reality, the barrier to colonization by diversity may only apply at a local level—larger scales may provide habitat suitability not available at smaller scales [4]. While many Hawaiian *Drosophila* species would succumb to desiccation or predation, *Scaptomyza* differ in that, while they evolved in Hawaii, they diversified into 21 different subgenera, some of which emigrated to the American pacific coast and other Pacific islands such as New Zealand and succeeded in populating their new environments. Clades that place island species within mainland clades serve as evidence for initial colonization, while mainland species within island clades are the result of reverse colonization [4]. In a similar manner, species found on an older island embedded in a newer island clade could be evidence of reverse colonization. *Scaptomyza* flies have not been frequently used as a model organism in genetic and developmental studies, but research into this genus will offer unique insights into Hawaiian fly evolution because some of its species left the islands after evolving there. It is rare for an island species to leave its island, and even rarer for the species to successfully colonize their new environments.

This study analyzes *S. anomala*, which is part of the subgenus *Bunostoma,* which has individual species in Hawaii, Australia, and Brazil. The *Scaptomyza Bunostoma* subgenus contains flies that breed on both flowers and fungi. *Bunostoma* endemic to Hawaii show a counterintuitive evolutionary pattern: they appear to have originated on younger islands and migrated to older ones [2]. *Scaptomyza*'s ancestors are estimated to have left Kauai 8 times and migrated back twice [2].

This study focuses on *Scaptomyza anomala* because it is easily reared in laboratory conditions, providing ready access to embryos. *S. anomala* diverged before leaving the Hawaiian Islands (citation) and may reveal what changes took place during the divergence between *Drosophila* and *Scaptomyza* before emigration took place. *S. anomala* is a member of the *Bunostoma* subgenus, which contains members native to Brazil and Oceania [34,35]. *Scaptomyza elmoi* was a good study candidate for similar reasons, but low sample quality led to data that could not prove useful in this study. Since *S. anomala* is so closely related to species that left Hawaii, comparisons of both Hawaiian *Drosophila* proper and non-Hawaiian *Scaptomyza* may determine whether *S. anomala* and its Hawaiian relatives share more with members of a geographically close island group or their farther-off relatives. The results of the comparison could inform previous theories of divergence. If *S. anomala* has more in common with the Hawaiian species, this could support the single-origin hypothesis, while increasing similarity with non-Hawaiian *Scaptomyza* or a more basal *drosophilid* species could support the multiple-origin hypothesis. Future *S. anomala* comparisons could reveal some of *Bunostoma's* migration between individual islands in the Hawaiian archipelago.

Exploring *S. Anomala* through the MZT

Scaptomyza flies have proven to be hardy species, and many factors played a role in their success in new environments. Developmental aspects such as egg substrate, size, and shape determine specialization or generalization within their environments[2,3]. Filaments help the embryo access oxygen, substrates feed newly hatched larvae, and egg size determines how much energy is stored for the embryo, as well as its initial hatching size. These traits are not determined by the embryo, but instead its mother, to give the embryo its best chance at developing. The mother also determines which transcripts go in the embryo to contribute to translation and transcription processes while the embryo's own cellular machinery develops. Since *Scaptomyza* flies have shown a great ability to adapt to new

environments, the maternal RNA deposits may also show adaptations to survive a wider range of conditions.

A transcriptomic analysis of the changes between the maternal and zygotic RNA can allow for fold-change comparisons that could yield insight into which genes play important roles at each stage[11]. Gene ontology analyses of genes upregulated at either maternal or zygotic stages can reveal which processes (at the molecular, cellular, and biological levels) predominate at each stage. Analyzing the isoforms of transcribed genes can show what each gene is used for at each stage. If the maternal and zygotic stages show little divergence, it may be possible the maternal deposits were meant to act as backups if a zygote's genes are defective. Hypothetically, divergence between the two stages could indicate a prioritization of rapid development.

The biological priorities of the early embryo might reveal traits that contributed to *S. anomala's* adaptivity, as well as those that helped *Scaptomyza* overcome the limits of island adaptation and spread worldwide.

Materials and Methods

Drosophila grimshawi Materials and Methods

Drosophila grimshawi husbandry

Upon receiving the *D. grimshawi* (Stock Number 15287-2541.00) stocks from the San Diego *Drosophila* Species Stock center, the flies were maintained in accordance with a modified version of stock center protocols[36]. This line was a replacement donated by Dr. Ken Kaneshiro from the same stock that was used in sequencing the *D. grimshawi* genome for the 12 *Drosophila* Genomes Project [37]. The flies were kept in vials of Wheeler-Clayton food [38], prepared as for *S. anomala* to adjust for New Orleans' climate conditions. The vials were papered with Kimwipes and changed to new vials twice per week. Flies were kept at 12-16 individuals to a vial to increase mating without causing frustration due to crowding. Once the larvae reached the third instar, their vials were moved to a jar containing two layers of heavy-duty paper towels cut to fit the jar, and then covered with 2 cm deep of a 50/50 mixture by volume of oolite and aragonite sand to pupate. The sand was kept moist by spraying in water from a squeeze bottle once or twice per week until the sand and towels were damp but showed no puddles. Spare cotton flugs (not cotton balls) were used to absorb any excess water, and then disposed of. Clean, damp flugs were used to maintain a humid environment for flies after eclosion. After the first eclosions, the vials with larvae were moved to a new sand jar, and a small petri dish of Wheeler-clayton top layer was added to the jar to feed new flies as they emerged. The flies were initially quickly moved to vials to continue population expansion, but often became stuck in the paper or condensation due to the humidity of New Orleans. To avoid losing flies before they reached sexual maturity, the flies were kept in the jar for two weeks and moved to large egg collection cages with a base of apple-agar media. After pouring the apple media, Wheeler-Clayton food was melted and deposited in the center of each plate in a 2-mL drop to provide food for the adult flies, as well as a nutrient-rich media to lay eggs. Population expansion could not take place in bottles, as the flies tended to get stuck and the food fell too easily. The flies laid their eggs in the Wheeler-Clayton pellets, which were transferred from the apple media with a spatula. The pellets were washed with water or a 50% solution of apple cider vinegar to inhibit the viability of mold spores travelling through the cage cover. The pellet was cut up with a spatula and mixed into vials of Wheeler-Clayton food. The larvae were more tolerant of the hypoxic environment under the food than the mold, and most vials prepared in this manner produced dozens of larvae and little to no mold. Larvae progressed as if they were laid directly in the vials. After the first egg collection, all flies were moved to cages and the initial method in which flies laid directly into vials was phased out. All larvae afterwards were raised from eggs in Wheeler-Clayton food pellets collected from cages or jars.

Egg collection

In order to produce statistically significant results, four libraries per stage, per species were produced with the aim of having three high-quality transcriptomes per stage. In order to reduce within-sample variation, transcriptomes were produced from single-embryo samples. The small samples can produce such low RNA content that they may not have enough to create an adequate cDNA library. Approximately 20 embryos were collected per stage and species, and only the highest-quality RNA and DNA samples were kept.

To collect the embryos, flies were held at room temperature in an egg collection cage lined with a petri dish filled with apple-agar media. About 2 ml of a paste of live yeast and water was spread in the center of the plate. Since St2 embryos can be easily confused with later stages, the cages were timed to ensure all eggs were laid within one hour. The flies were held on live yeast for one hour, after which the first plate was discarded and replaced with a new plate. St5 embryos were collected from these plates, and St2 embryos were only collected after plates were replaced once more.

In order to stage embryos for selection, chorions were removed. A 50% bleach solution was poured onto the collection plate, and an eyeliner brush was used to break up the yeast pellet and dissolve it into the solution. While smaller eggs could be dechorionated in 120 seconds after the addition of the bleach, *D. grimshawi* eggs tended to have tougher chorions to protect the larger eggs. Dechorionation could take upwards of three minutes and was usually stopped when the filaments were completely dissolved and no longer visible. To stop the dechorionation process, the bleach solution and the eggs were poured through a square of 70-micron nylon mesh stretched over a 1" PVC adapter (Cantex 514015) and secured with a 1" PVC threaded bushing (Cantex 5134202). The eggs were rinsed with deionized water for at least 30 seconds, until bleach was no longer detectable by smell. The dechorionated eggs were moved onto a drop of halocarbon oil (Sigma) on a microscope slide. The slide was placed on a dissection scope over refracted light. The embryos were selected using a set of morphological characteristics which are conserved across all species [12]. St2 embryos were identified by the empty poles on each end of the egg and their lack of visible cell membranes and evagination. St5 embryos were collected at the later end of St5. Traits required for St5 embryos were 2 layers of well-defined cell membranes and formation of the pole cells. Embryos showing any damage, further cell development, or any evagination were rejected.

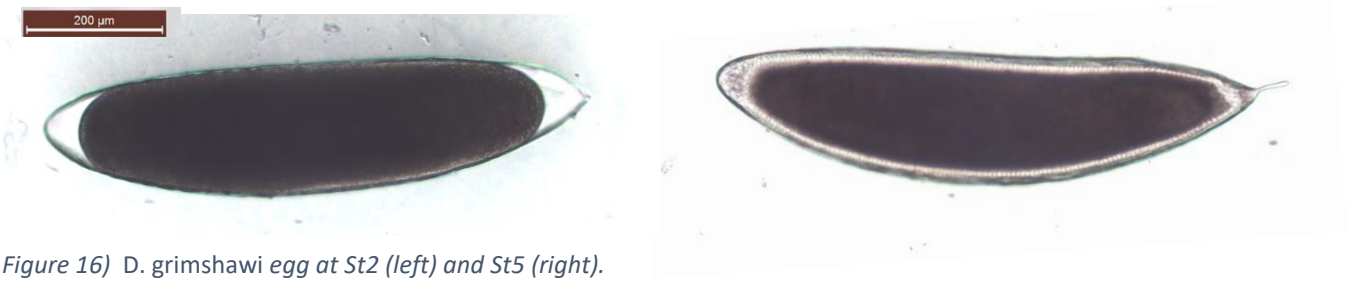


Figure 16) D. grimshawi *egg at St2 (left) and St5 (right).*

A 1.5 mL microcentrifuge tube was filled with 800 ul of Trizol reagent (Ambion) and labelled with the sample number. The labelling scheme contained three components: a letter marking the species, in this case G for *grimshawi,* a number marking the stage, either 2 or 5, and a final letter to indicate the individual sample (G5H would indicate the sample came from a grimshawi embryo at St5 and was labelled as sample G). Additional third letters were added when over 26 individual samples were collected (G2AC would indicate the alphabet was used once, so the second set would contain the letter A, then a new individual letter).

Each embryo was placed on the corner of a clean microscope cover slip. A 0- or 000-point paintbrush was used to lift and lower the embryo in a new place two to three times to remove excess halocarbon oil. The embryo was then moved to the center of the slide for lysing. 3 ul of Trizol were pipetted from the labelled 800-ul aliquot and placed in a droplet on top of the embryo. The embryo was then lysed with a sterile, 30-gauge medical lancet (Reli-On Ultra-Thin). The embryos were left to dissolve for 5 minutes. After the embryo was completely dissolved, an additional 3 ul of Trizol from the

tube were added to the slide, and all 6 ul were pipetted back up and into the remaining 794 microliters and pipetted up and down. 6 microliters were again removed from the same tube and pipetted onto the place of the embryo, re-collected and placed back into the aliquot. This rinsing step was repeated two times to ensure thorough collection of embryonic tissues. Labelled samples were held at -80°C until RNA purification.

RNA extraction

RNA was extracted per the manufacturer's instructions using a Trizol phenol-chloroform extraction (Invitrogen). The method was modified to accommodate an initial volume of 800 ul of Trizol and the low RNA content of the samples. During RNA precipitation, 10 uL of 20 ug/ul glycogen (Invitrogen™ UltraPure™ Glycogen) was used, and the samples were spun in an Eppendorf 5424 R centrifuge at 21130 RCF (maximum speed) for a full 60 minutes. If a sample showed no visible pellet at this point, it was spun for another 30 minutes. After resuspension in UltraPure water, 15.5 ul were kept for processing, and 3 aliquots of 1.5 ul each were taken for quality analysis.

RNA Quality analysis

Between 10 ng and 1 ug of RNA is needed to create a cDNA library using NEBNext® Ultra™ RNA Library Prep Kit for Illumina® (New England BioLabs). To determine which samples could go on to library processing, RNA aliquots were tested for concentrations using a Qubit 4 fluorometer with a Qubit high-sensitivity RNA assay. All samples collected showed concentrations of at least 4 ng/ul (62 ng RNA per sample), so the concentration alone was high enough to process, but only high-quality, low-degradation samples could go into libraries.

To determine the quality of the RNA, samples were tested on an Experion bioanalyzer using a Bio-Rad Experion RNA High-Sensitivity kit. During the research, the Experion stopped working, and analysis was continued on an Agilent 2100 bioanalyzer using a Bioanalyzer RNA 6000 Pico assay (Agilent). Both samples were inspected for quality by finding the two large peaks for the 28s and 18s ribosomal subunits and smaller peaks representing the 5S and 5.8S subunits (fig. 2). Other peaks and peaks that were low, absent, or too close together would indicate degradation. It is important to consider that insect RNA electropheragrams will show different patterns than human or mammalian eloctropherograms. First, for insect RNA there is no equivalent to the RNA integrity Number (RIN) which is used to indicate RNA quality for mammalian samples[39]. Peaks for mammalian 18 S and 28 S subunits show greater separation on the electopherogram than the equivalent *Drosophila* 18 S and 28 S subunits [39]. Since *Drosophila* peaks occur so close together, it is important to account for this before dismissing close peaks as degraded.

This study is concerned with only messenger RNA (mRNA), which occurs at a far lower concentration tin ribosomal DNA. While mRNA levels would not be visible in an Agilent assay, the ribosomal RNA is expected to degrade at the same rate as mRNA, so samples with high-quality rRNA were assumed to have high-quality mRNA suitable for library construction. Samples with high RNA concentration and no signs of contamination or degradation were processed into cDNA libraries.

Sample	Bio-analyzer Concentration (ng\µl)	Qubit concentration (ng\µl)
G2D	8.308	6.7
G2O	24.41	13
G2AB	9.629	8.08
G2AF	13.142	11.2
G5F		6.98
G5H		8.4
G5L	7.659	8.1
G5AA	10.16	12.0

Table 4) RNA concentration of D. grimshawi RNA samples from which libraries were generated, in nanograms per microliter. A rough estimate was taken during testing on the bioanalyzer, but Qubit concentration readings are regarded as significantly more accurate than Agilent Bioanalyzer concentration estimates. Samples with errors on the Qubit were discarded, but samples with bioanalyzer errors were kept if they contained no degradation.

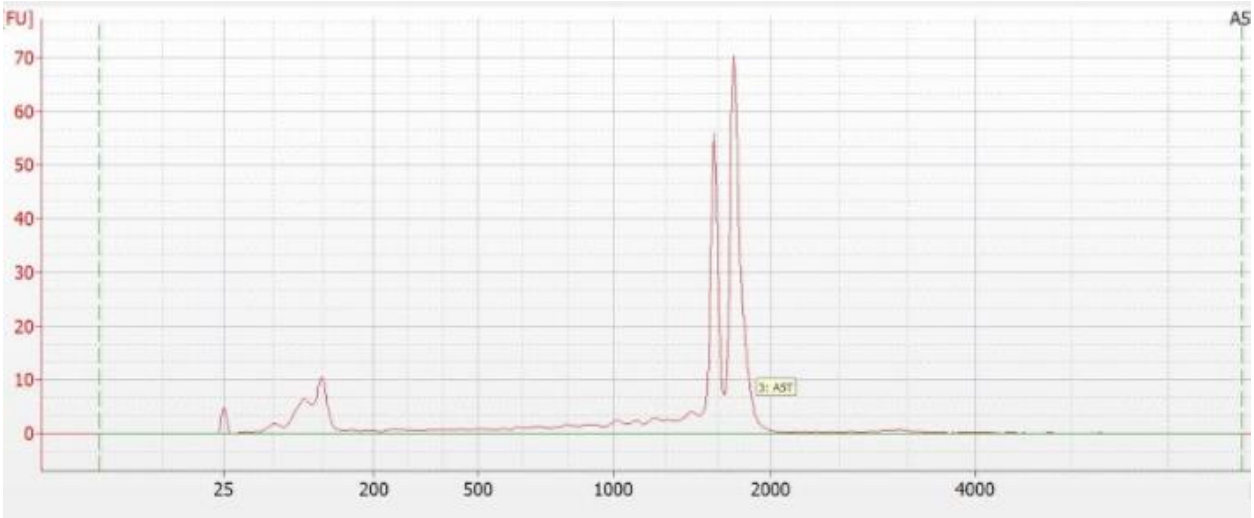


Figure 17) An optimal RNA bioanalyzer electropherogram (sample A5T). RNA size in base pairs is measured on the X axis and fluorescence is labelled on y.

cDNA libraries

Sequencing was performed at Novogene using Illumina cDNA sequencing methods. As such, RNA and library quality had to meet Novogene's standards. To remove any DNA contamination before

library construction, a Turbo DNA-Free kit was used according to directions before immediately continuing to library construction using the NEBNext Ultra RNA Library Prep kit for Illumina. Adaptors and barcodes were added using NEBNext Multiplex Oligos for Illumina, sets 1, 2, and 3. During the PCR enrichment of adaptor ligated RNA, the procedure used the individual sets rather than the set of 96 index primers. The starting concentrations were low, so the denaturation and annealing/extension steps were repeated the maximum of 15 times in a Bio-Rad T100 Thermocycler. Three 1.5 ul aliquots of library product were taken for quality analysis, and 15.5 ul were reserved for sequencing.

The samples show main peaks around 300 BP—about the fragment length needed for high-throughput Agilent sequencing. The quality of the libraries was checked using the Agilent 2100 High-Sensitivity Pico DNA Kit.

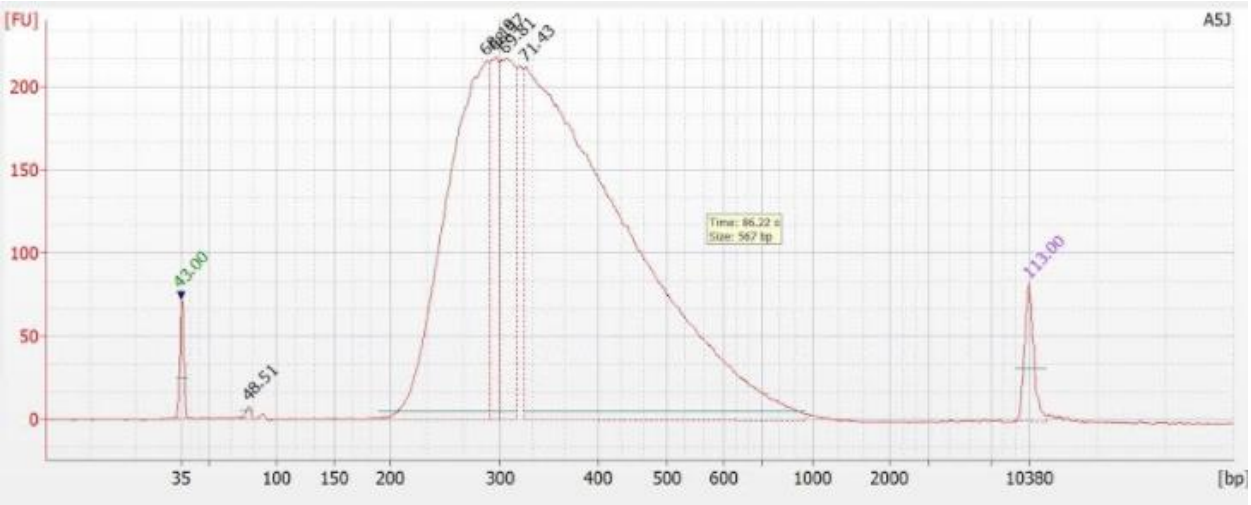


Figure 18) Bioanalyzer reading of a high-quality DNA library (sample A5J).

Sequencing

Libraries were sent to Novogene in two batches for sequencing on two lanes. The four highest-quality libraries for each stage were sequenced in the hope that three replicates per sample could be used to avoid sampling bias. Only libraries from samples showing minimal RNA degradation were selected. The samples varied in concentration, so those with a similar concentration were selected for sequencing. Concentrations from both Qubit and the Bioanalyzer electropherograms were considered.

Bioinformatic Analysis

Bioinformatic analysis was carried out using methods outlined in a previous transcriptomic paper on the MZT [11]. Genome and annotation files were previously produced in the 12 *Drosophila* Genomes Project, and were used as the scaffolding in this analysis [27] The analysis required quality control and trimming of the raw data, creating a transcriptome, aligning the transcripts to the transcriptome, counting the numbers of each transcript, and performing a differential expression analysis. Cutadapt was used on the raw reads to remove adapter contamination [40].

After adapter removal, transcriptome analysis began with the Tuxedo Suite [41]. Tophat2 [42] was used to leverage BowTie2 [43], aligning the reads to the reference genome [27] and find the splice junctions present in the transcripts. Cufflinks was used with the -N upper-quartile normalization option to create

a transcriptome assembly for each sample [44]. The samples were combined by species and stage using CuffMerge [41]. To analyze representation, Cuffnorm calculated the FPKM measures [41]. The expression levels were compared to one another in differential analysis using Cuffdiff [41]. Gene expression was determined by combining the expression of all gene isoforms. Orthologs were assigned using Flybase's orthology table. The gene was only studied if 10 of the 15 species (*Drosophila* species *melanogaster, simulans, sechellia, yakuba, santomea, erecta, ananassae, pseudoobscura, persimilis, willistoni, mojavensis, virilis, miranda, mauritiana,* and *grimshawi*) in the comparison had a 1-1 ortholog. Orthology results were compiled into a table for analysis.

R was used to create Spearman rank sum correlation coefficients, which were plotted out with the R heatmap2 package [45]. Heatmaps were used to visualize the changes in representation of genes between stages and species.

Scaptomyza anomala Materials and Methods

Fly husbandry

Upon arrival from the San Diego Fly Species Stock Center, adult *Scaptomyza anomala* (stock number 33000-2661.00) flies were transferred to vials containing standard potato media and about five grains of baker's yeast (Fleischmann's). The flies stuck to the potato media, so they were moved to Wheeler-Clayton media [38]. The Wheeler-Clayton food dislodged from the bottles and crushed the flies during transfer, so they were raised in several vials rather than fewer bottles. *Scaptomyza anomala* are native to Hawaii and are adapted to a cool mountain climate. As this study took place in New Orleans, LA, the flies were raised in conditions designed to protect them from the higher temperature and humidity of the local climate. To adjust for the temperature, specimens were kept in a 19° C Percival Intellus incubator to protect them from the risk of heat shock. Flies were kept in a 12-hour light, 12-hour dark cycle. To counter the humidity, which could have caused the flies to stick to the food, Wheeler-Clayton food was prepared using the maximum recommended amount of agar (14.5 g per liter of water). The cooked potato media was prepared with 456 g of filtered water for every 100 g of powdered mix to ensure a consistency that prevented food from falling during transfers. High doses of ethanol and propionic acid (6.5 ml per liter of water) were also used in the top layer recipe to prevent mold growth.

The flies showed increased oviposition with the addition of about 5 granules of live yeast (Fleischmann's) to the vials. Three parallel, shallow slashes with a scalpel into the top layer provided a place for small larvae to burrow. Flies were kept at 30-40 individuals per vial and transferred to new food every two to three days, usually on Mondays and Wednesdays. When the flies were removed, a small spatula spoon was dipped in water and used to scrape down eggs and larvae on the side of the vial to food level. The spatula was then used to cut a deep crosshatch into the food to give more surface area to larvae. The spatula was cleaned with 70% ethanol and dried between vials. Papering the vials containing adults led to early mold growth, so flies were kept upright in vials with no paper. Damp 2 cm by 5 cm strips of plain white printer paper were slid between the food and the vial wall when the larvae reached their third instar and wandering larval stages. This offered a secure location for pupation and kept the pupae away from the food and potential mold.

Rescue of Mold-infected Cultures

Despite precautions taken against mold, some vials still became infected. Treatments varied depending upon the life stage of the insects in the vials, the severity of the mold, and the health of the remainder of the culture. If the mold occurred before the larvae reached the third instar, the larvae were too small to safely clean or transfer. A spatula was used to remove any small spots of mold forming and monitored for continued mold growth. If the mold growth continued or became fuzzy, moldy vials were euthanized at -20°C overnight. If the healthy adult population was too small to minimize the impact of losing one vial (under three vials), larger growths were brushed down with water daily until the larvae reached the third instar or began forming pupae. At this stage, about 4 ml of 50% apple cider vinegar were poured into the vial, loosening the larvae and pupae with a brush. The solution and all the pupae and larvae would be poured into a petri dish, after which the larvae and pupae were moved to another dish of clean water. After rinsing, the larvae and pupae were transferred to a fresh, papered vial of Wheeler-Clayton food: pupae on the paper and larvae in the food itself. This process was also applied to vials in which mold appeared after pupation occurred. To prevent the accumulation of mold in the digestive systems of newly-eclosed adults, all new eclosions were housed in cleansing

vials, which were changed daily for 7 days before transfer to the normal Wheeler-Clayton food and changing schedule. The cleansing vials contained approximately 1.5 cm of apple juice-agar media topped with an even, generous sprinkling of fine-grained live yeast (Red Star). This was chosen so the flies would eat heavily to purge their digestive tracts, and because the apple media is easier to prepare and less likely to fall during transfer. Since the flies did not lay eggs for about a week after eclosion, there was no need to keep the vials to raise offspring, so larval viability was not a necessary consideration.

Egg collection

When they reached a week old, approximately 200 *S. anomala* flies were placed into a large egg collection cage with a smear of live yeast paste. The timing, extraction, and dechorionation of the eggs proceeded just as in the *grimshawi* protocol, but dechorionation was stopped at 120 seconds.

RNA extraction

RNA was extracted using a Trizol phenol-chloroform extraction (Invitrogen), as described in the *grimshawi* protocol.

RNA Quality analysis

RNA quality was checked just as the *D. grimshawi* RNA was. The concentrations at each stage were lower, because each sample still used only one embryo, but the embryos of *S. anomala* are much smaller and therefore contain less RNA than those of *D. grimshawi*.

Sample	Bio-analyzer Concentration (ng\µl)	Qubit concentration (ng\µl)
A2J	2.718	4.6
A2L	2.929	5.0
A2P	3.475	
A2AC	2.514	9.56
A5E	5.570	8.1
A5H	4.330	5.58
A5M	3.638	5.0

Table 5) Concentrations of S. anomala *samples measured from the Agilent Bioanalyzer and Qubit*

cDNA libraries and sequencing

Libraries were generated using the same protocol used for *D. grimshawi* and sent to Novogene for sequencing. Half of the samples were sent with the *D. grimshawi* library, and the other half were sequenced alongside *Scaptomyza elmoi* samples used in a different study on a separate lane.

FASTQc

To check the quality of the reads and the success of the Illumina sequencing, a quality-control check was carried out with FASTQC [46]. FastQC is run by uploading the FASTQ files to be checked into the Java program. It can measure quality scores across bases, per sequence, and on GC content, N content, length distribution, and duplicate sequences. Across-base quality score should have a lower quartile above a Phred score of 10 for any base, and the median should be above a Phred score of 25. The most frequently observed per-sequence quality should have a mean above a Phred score of 27, with an error rate less than 0.2%. While not all the samples met these qualifications, trimming and cleanup applications were used to improve the sequence quality.

Data analysis

Alignment of Reads

Because *S. anomala* is not a model species, there is no high-quality genome to use as a scaffold for alignment. In the absence of a genome, an RNA alignment was assembled *de novo* using Trinity assembly software[47]. The FASTQ files sent in from Novogene were loaded in two sets: the forward and reverse reads of the St2 samples and the forward and reverse reads of the St5 samples. The St2 and St5 samples were each processed in a separate run. The forward files were input first, then the corresponding reverse files were input. A paired-end Trinity run was carried out for each set, using contigs of 200 base pairs. 24 CPUS and 230 gb of the available 256 gb of RAM were dedicated to the run. The –full_cleanup option was used to retain only the Trinity FASTA file while discarding temporary files to conserve memory. The two runs resulted in two FASTA files containing the transcriptomes for the St2 and St5 groups. A map coordinating genes and their isoforms was also created.

Trinity was normalized in silico with the default options for the paired-end mode, which runs automatically during trinity. This reduced universally low reads and duplicate sequences to preserve computational memory. Trimmomatic was also run using the –trimmomatic option to remove adaptor sequences and reduce sequence errors and low-quality nucleotides. A FASTQ file containing all the trimmed sequences was produced.

While it is possible to run Trinity in Blast2GO, these runs were conducted using command line shell scripts to ensure all necessary parameters were met, and to optimize memory use and speed by eliminating the need for a graphical user interface.

Quality Control

FASTQC

A second FASTQC run was used to determine whether the sequences were improved with Trimmomatic and Trinity's normalization methods[46].

BUSCO

S. anomala does not have a well-annotated genome, and there is therefore no template to directly determine whether all the transcripts present in the sample were measured. A BUSCO analysis compares sets of highly-conserved sequences in similar species to a dataset to check for orthologs of genes that would typically be represented [48]. A higher percentage of BUSCO matches would indicate the sample contains transcripts expected in its order, and that the library fairly represents the sample [48]. The analysis was run in transcriptome assessment mode, rather than genomic, to account for the fact that transcriptomes will show a lower number of individual genes than a genome. The samples were compared to a *Diptera* dataset with an e value set at the default of .01. While there was an option to compare the transcriptome to that of a specific species, limiting the analysis to one species could have limited the scope of the assessment, so no species was specified within the order. A *D. melanogaster* genomic library is expected to make a 98% score [48]. The *S. anomala* transcriptome may show a high percentage of incomplete transcripts could be the result of potential new transcripts. The embryonic samples do not represent the whole transcriptome, but instead the transcriptome at a very specific life stage. As such, the embryonic transcriptomes may not be as complete as the data provided to BUSCO [49].

Annotation

All annotations were conducted using Blast2Go Pro [50]. A GFF file was created to sort out genes from other features. Some genes contained multiple isoforms. To conserve computing power, a custom script was written to create a new FASTA file with the sequences of the longest isoform of each gene, assuming the isoforms of the genes have similar functions. Sequences were matched to those in the CloudBlast database. An InterProScan was used to detect motifs to find additional GO terms associated with the sequences. GO mapping was used to retrieve the GO terms associated with the sequences found in the BLAST search.

CloudBlast

The Blast-X option was used to match nucleotide sequences in the uploaded file to amino acid sequences in the Cloudblast protein database. Searching for nucleotide sequences can find non-protein gene products such as small RNAs and ribozymes that my not show up in an amino acid search. While this could be useful in future studies, this study aimed only to find sequences of protein-coding genes. The Blast-X search reduced the computational power needed for the search. The database was searched for non-redundant sequences with an expectation value of 1.0 e^{-3} and 10 Blast hits. The top 10 blast hits for each feature were exported for later analysis.

InterProScan

A FASTA file containing sequence information was loaded to Blast2Go. The scan was carried out using the default settings and the annotation was saved as an XML file. The file was then merged with the preexisting information on the sequences in Blast2Go.

Gene Ontology

GO terms were mapped back to the feature names, linking GO descriptions to the sequences. The mapping had no parameters, so it was simply run on the BLASTed file.

Counting the Reads

To start differential expression analysis, a count table had to be assembled to determine the number of times a given read appeared in each sample. To create the count table, the gene isoform map was loaded into Blast2GO with the untrimmed trimmed FASTQ files for each sample, forward and reverse.

Differential Expression Analysis

The count table and the annotations were loaded into Blast2Go. A pairwise differential analysis was performed using an experimental design file dividing the samples into groups St2 and St5, and then into individual samples (table 3). Blast2Go ran the EdgeR program to perform differential expression analysis [51]. A Count per Million (CPM) filter was used to remove samples with low counts across all samples. The filter was set to 1 to remove any samples which had all CPM values below one, both to cut out low counts and to eliminate bias in the calculation of fold-change that could occur in calculations with counts below one. The samples reaching CPM filter was set to three, the lowest number of samples per group, as recommended by B2G. A Simple Design Type was used since only two groups were defined. The simple design type sets two experimental conditions, in this case St2 and St5, and performs a pairwise comparison between the sets. A trimmed mean of M values was selected as the normalization factor, an Exact Test was set as the Statistical Test.

Differential expression analysis was run at the transcriptomic level, using St5 as the contrast condition (in which higher representation will be considered upregulated) and St2 as the reference condition (in which higher representation will be considered downregulated). The analysis was not strand-specific. The differential expression toolbar was used to create an MDS plot to represent sample separation. At this point, it was noticed that sample 5-3 clustered more tightly with St2 samples rather than St5 samples and was likely mislabeled. After the differential expression analysis was run, the results were opened with the annotations by selecting and opening them together in the file manager. The merged file was used in the enrichment analyses.

Experimental Design

Sample	Lib. size (pre-filter)	Lib. size (post-filter)	Norm. factor	Condition	Individual
A2AC_L4	21,982,883	21,740,113	1.128	Stage_2	4
A2J_L1	14,481,005	14,380,528	1.109	Stage_2	1
A2L_L4	15,292,562	15,200,645	1.219	Stage_2	2
A2P_L1	16,867,011	16,779,060	1.013	Stage_2	3
A5E_L4	18,368,385	18,138,658	0.861	Stage_5	1
A5H_L1	16,810,370	16,513,538	0.882	Stage_5	2
A5M_L1	16,954,519	16,694,764	0.852	Stage_5	3

Table 6) Experimental design for differential analysis.

Enrichment Analysis

A two-tailed fisher's exact test was run on the differentially expressed data using the datasets in which St5 was the primary target, then in the dataset using St2. In order to determine which features were significantly under-or over-represented, the results were filtered using a false discovery rate (FDR) less than or equal to 0.05. Bar Charts were created using the least and second-least specific GO terms to determine GO enrichment.

Inter-Species Analysis

To find relevant genes to compare expression levels with other species, transcripts with representation under 1 TPM were filtered out in Blast2Go. Augustus was used to find the sequences most likely to be coding regions of genes[52,53]. These sequences were saved as FASTA files, and a new count table was created from the original set of transcripts. Blast results were appended to these sequences, and a custom script was used to pair the *S. anomala* genes to their orthologs in other species.

Results

D. grimshawi Results

Representation Between Stages

Once the transcripts were mapped back to the transcriptome and counted, an initial test was run to ensure the samples were consistent with their stages. St5 sample were expected to show higher degrees of variability within their respective stages due to the extended length of St5 in *D. grimshawi*. Samples were selected as late into St5 as possible (around cellularization, just before gastrulation), but some samples resembled late St5 for as long as an hour, during which time many transcriptomic changes could have occurred. St5_3, which clustered with the St2 samples, was removed from the analysis. In the results presented here, we also removed St5_2 and St2_1, which showed lower levels of read alignment to the reference.

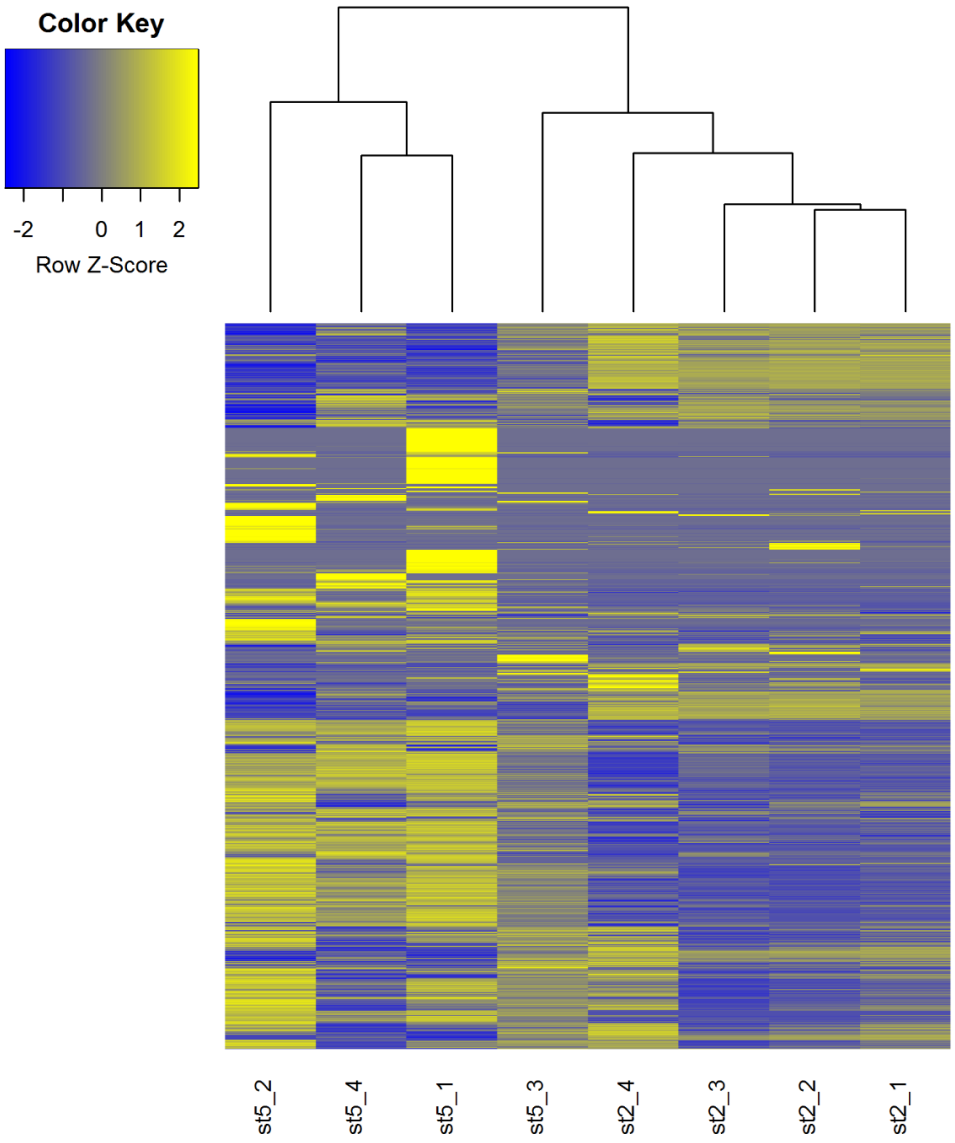


Figure 19) A heatmap of gene representation across the D. grimshawi samples.

Differential expression at all stages

When comparing FPKM levels (fragments per kilobase per million reads mapped) – a measure of transcript level) by hierarchical clustering [54] of all genes represented at St2 (St2) and St5 (St5) of each species, *D. grimshawi* forms an outgroup at each stage (fig. 5). While the dendrogram shows that species transcriptomes may display different patterns of clustering at St2 and St5, at both stages the other species cluster together more tightly than with *D. grimshawi*. Importantly, *D. grimshawi* does not cluster with its two closest relatives, *D. virilis* and *D. mojavensis*, suggesting strong transcriptomic divergence in this species.

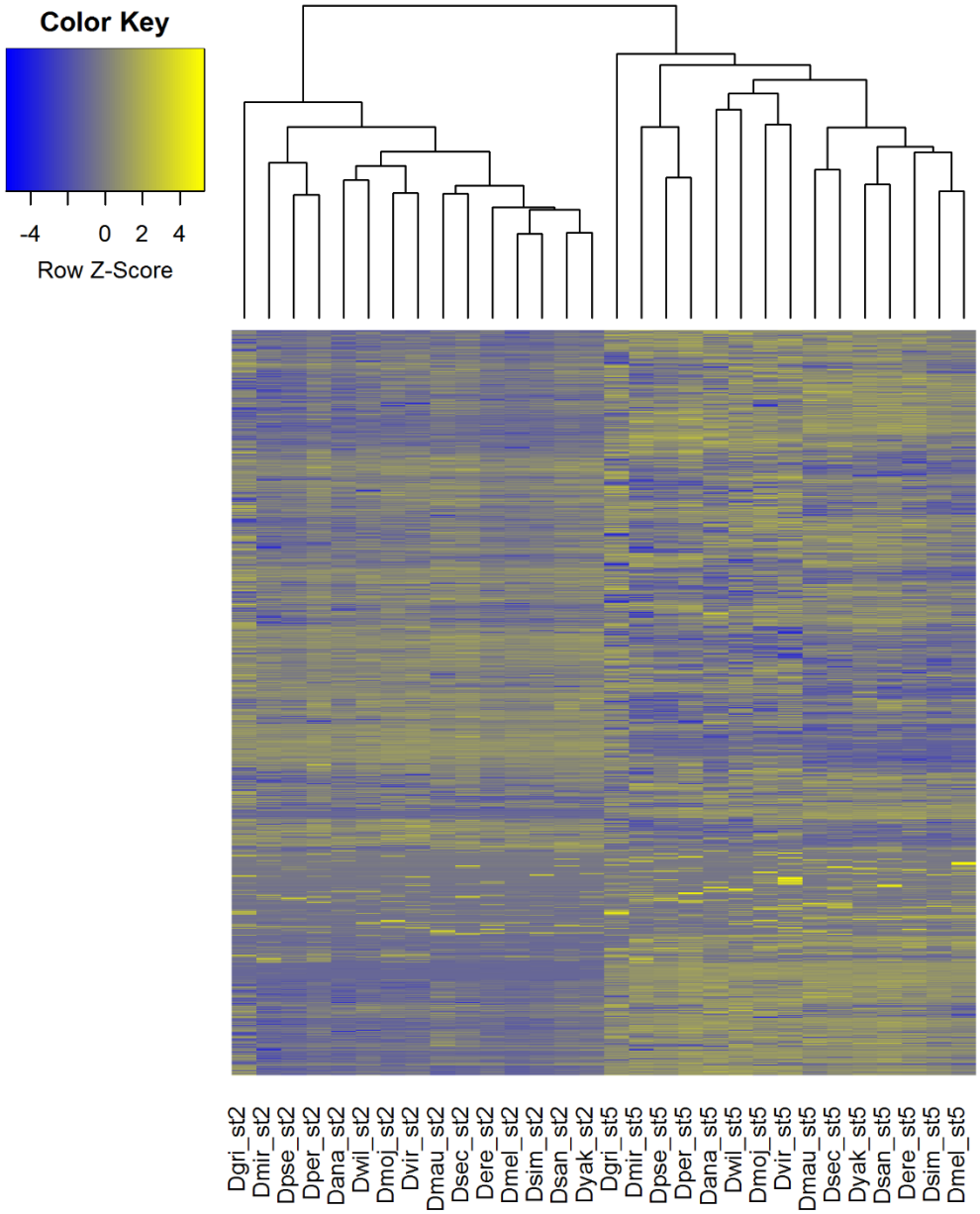


Figure 20) Hierarchical clustering of St2 and St5 transcriptomes from 15 Drosophila species.

A)

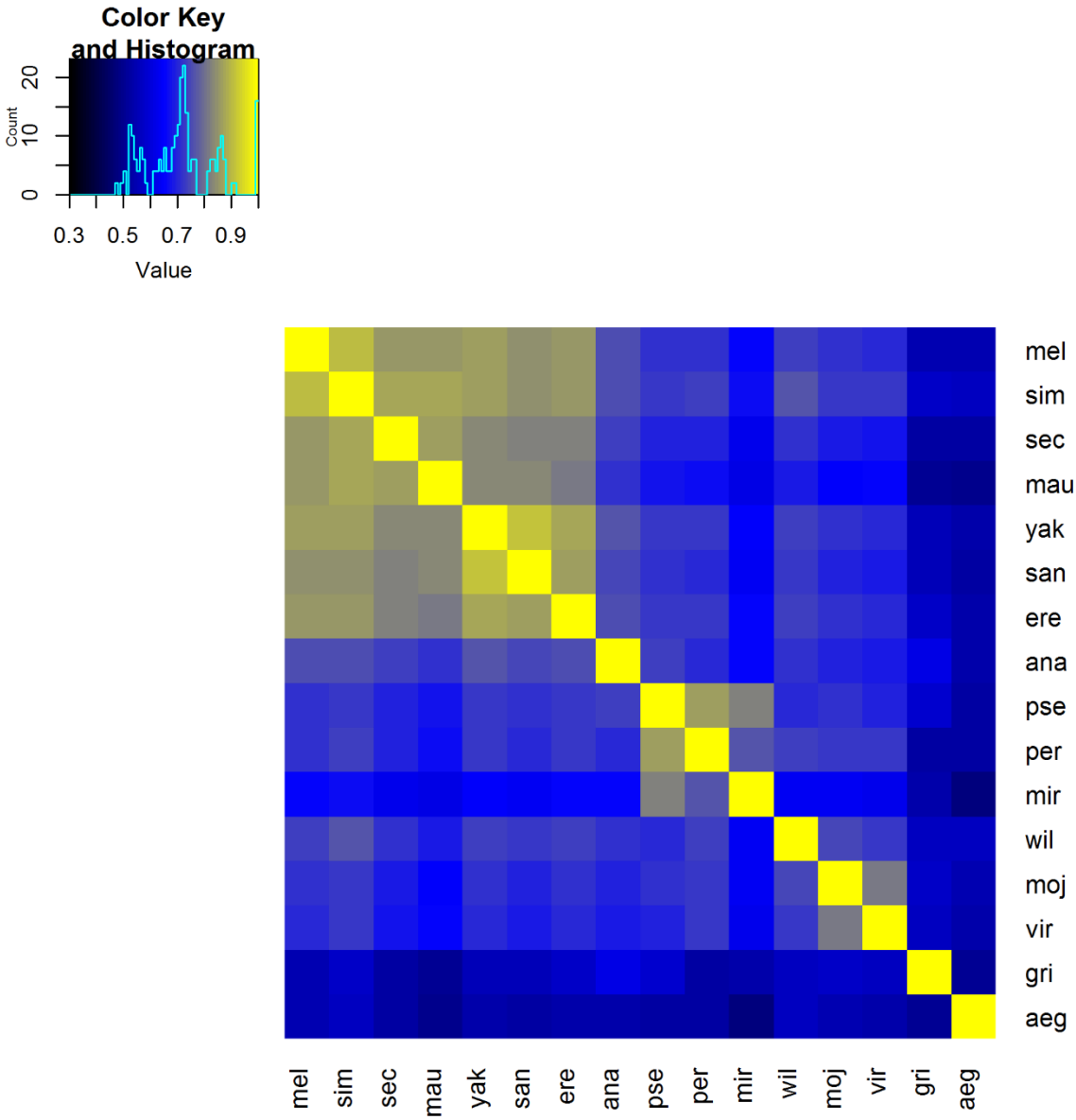


Figure 21) A) A heat map comparing the Spearman rank sum correlation coefficients of overall gene representation in St2 across 15 fly species, along with Aedes aegypti (Akbari et al., 2013) as an outgroup for comparison.

B)

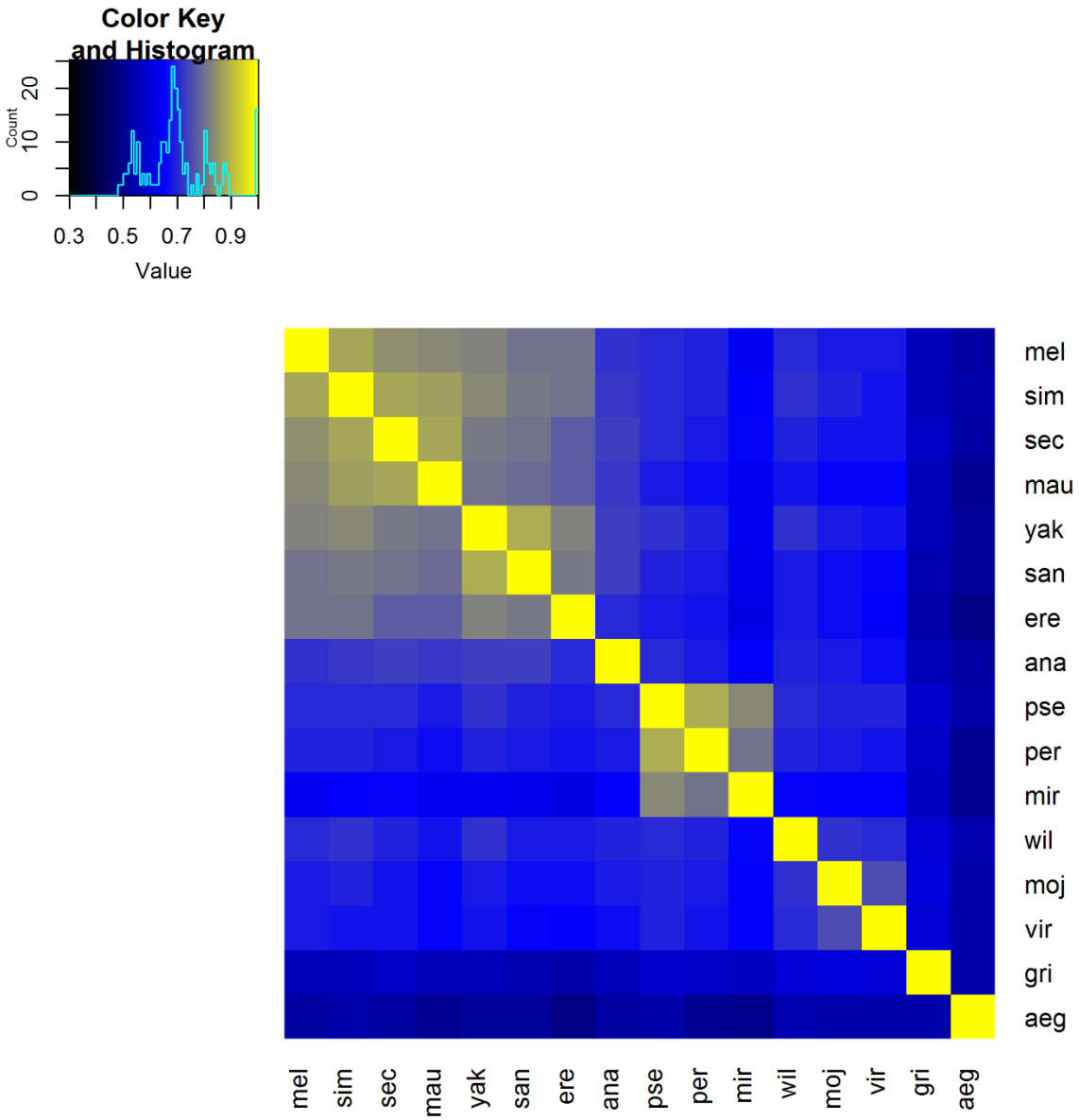


Figure 6 Cont.) B) A heat map comparing the correlation of overall gene representation in St5 between all 15 fly species, along with Aedes aegypti as an outgroup for comparison.

When comparing the overall representation of genes at each stage between species, *D. grimshawi*'s outgroup status remained constant. When compared to transcriptomic data from the more basal *Aedes aegypti*, *D. grimshawi* showed a wide margin of difference – *D. grimshawi* showed lower correlation with *A. aegypti* than it did with other D*rosophila* species (fig. 6a). However, *D. grimshawi* was still less correlated with the other *Drosophila* species than they were with one another. Similar patterns arose when comparing the species at St5 (fig. 6b). Despite *D. grimshawi's* differences at these stages, the correlations were relatively consistent compared to *A. aegypti*.

Zygotic only

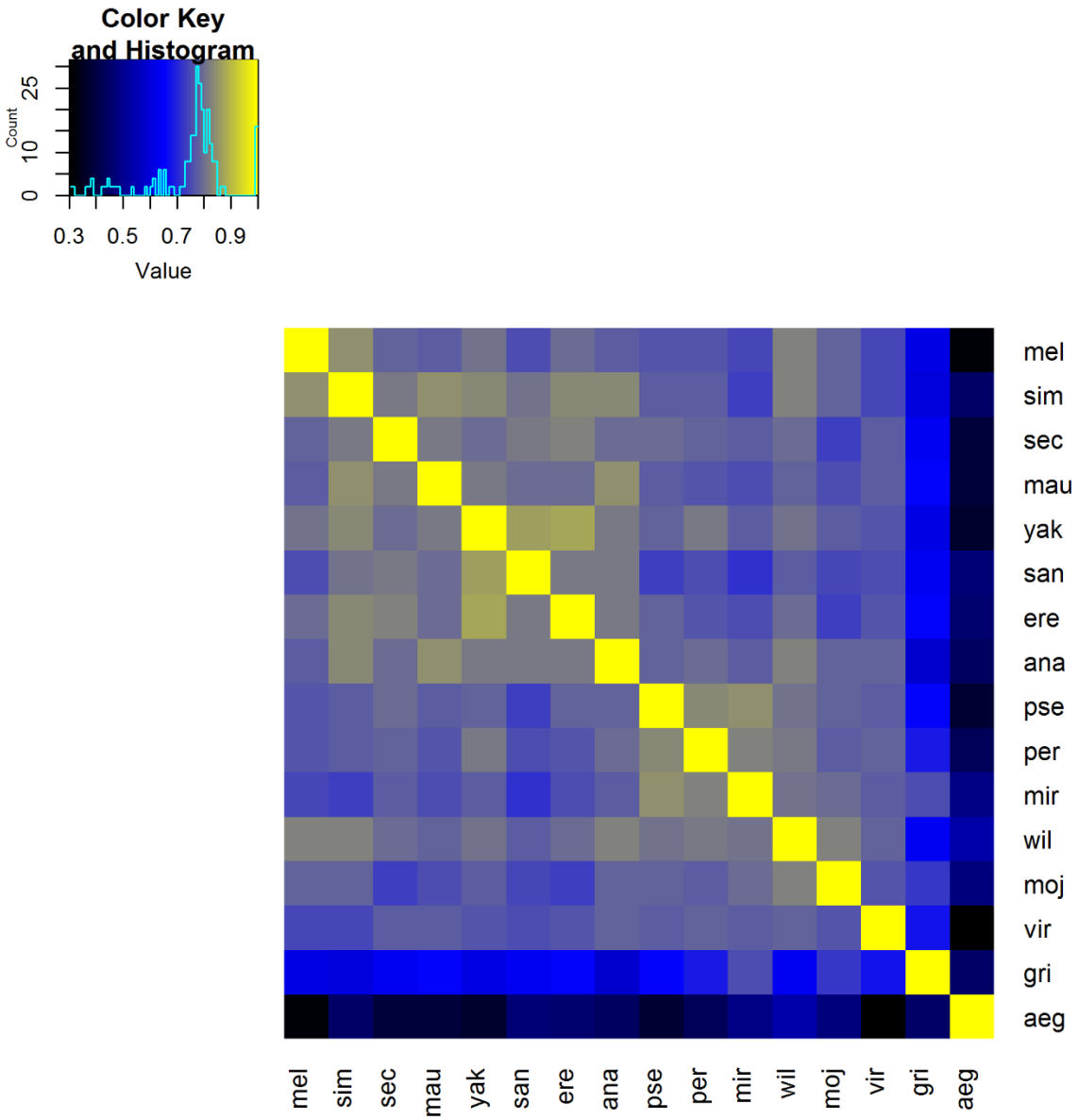


Figure 22) A heat map showing Spearman correlation coefficients of transcripts represented only at the zygotic stage (St5), with A. aegypti as an outgroup for comparison.

Zygotic only genes are those that are not represented in the maternal deposits and exclusively produced by the zygotes [55]. Here, genes not represented at St2 but present at St5 were considered zygotic only. These genes are highly conserved among *Drosophila* species [55].

The zygotic-only genes showed the lowest degree of correlation with *Aedes aegypti* (fig. 7). These genes are highly conserved within *Drosophila*, but the patterns of expression are markedly different in other insect groups. This split between *D. grimshawi* was less defined than the other gene

groupings, (fig. 7) but *D. grimshawi* was still an outgroup, even in a conserved group in which consistent correlations should have been expected.

	st2 gain	st2 strong gain	st2 loss	st2 strong loss	st5 gain	st5 strong gain	st5 loss	st5 strong loss
mel	28	12	12	6	74	16	20	3
sim	34	22	40	32	17	8	69	33
sec	52	32	19	9	48	19	23	9
mau	47	31	30	22	45	21	33	14
yak	32	19	11	4	28	9	16	3
san	24	13	11	5	35	9	15	3
ere	99	51	19	10	55	24	48	14
ana	54	32	37	20	60	28	60	24
pse	34	18	24	15	36	15	19	7
per	73	38	30	23	63	28	38	22
mir	79	46	74	60	72	34	83	47
wil	52	24	46	31	46	17	56	31
moj	83	44	17	10	84	34	77	37
vir	107	52	35	23	179	70	52	25
gri	87	49	28	16	77	35	91	53

Figure 23) Number of genes showing unique gains and losses in each species. Only genes with one-to-one orthologs in at least 10 of the 15 species are included.

When examining the differences between species, one metric of difference was the number of unique gains and losses of gene representation across species. A gain would mean that the FPKM level was greater than 1 in the noted species and less than 1 in all other species. A loss would mean that the FPKM level is below one in the noted species and above one in all others. Gains are considered strong when the FPKM change is greater or less than three instead of one. Only genes with one-to-one orthologs in at least 10 of the 15 species were included. While *D. grimshawi* does not stand out in terms of losses at St2, it shows a greater number of gains than average at St2 and St5 and more losses of St5 representation than any other species. (fig. 8).

Hox Genes

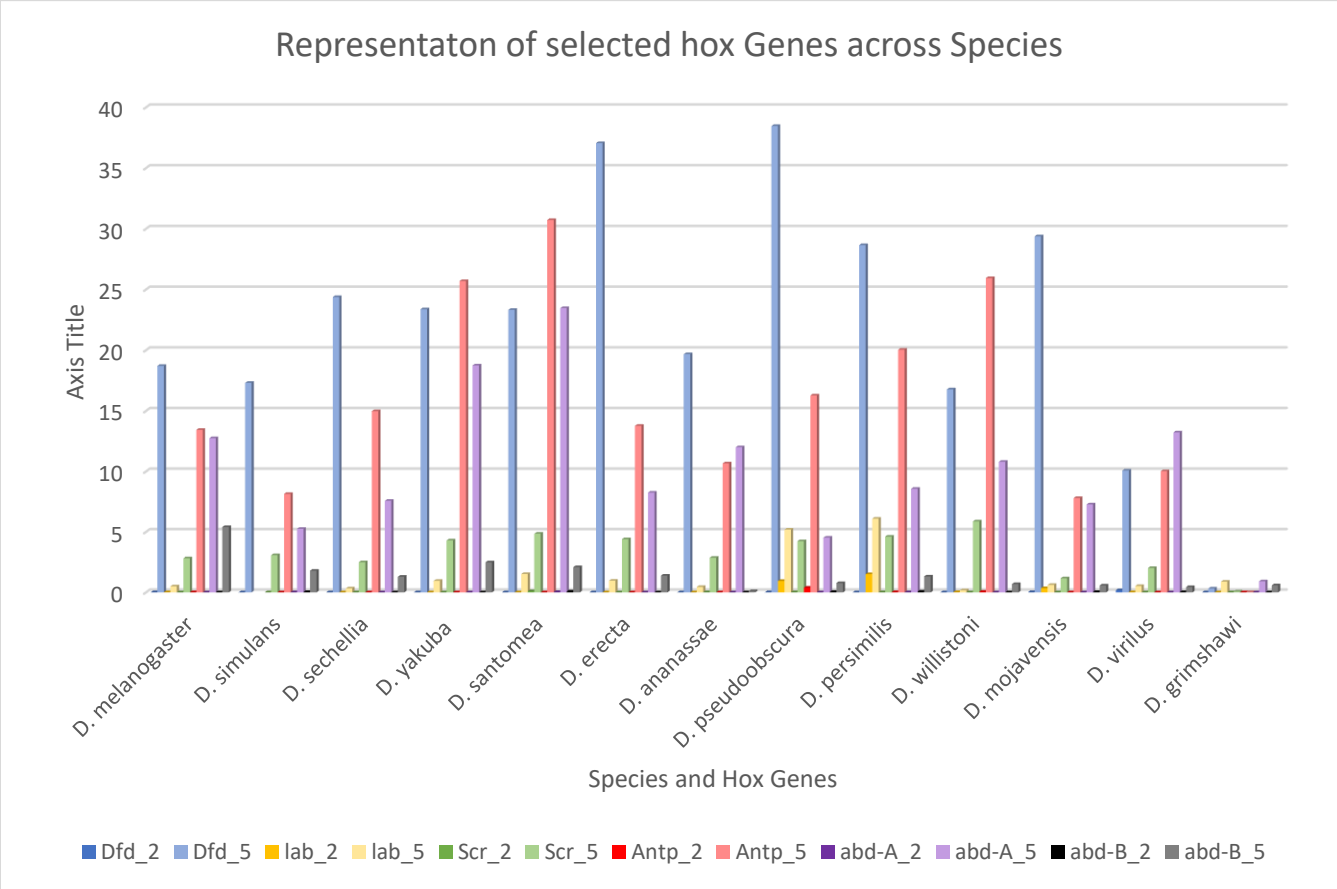


Figure 24) Representation of Hox genes across 13 species. St2 representation levels are represented in saturated colors and their St5 counterparts are represented in a corresponding pastel.

Included among the St5 losses in *D. grimshawi* are the iconic Hox genes. Hox genes specify anteroposterior segment identity, being the most downstream members of the segmentation cascade [7]. *D. grimshawi* has a longer lifespan than mainland flies and lower predation pressure. As such, *D. grimshawi* embryos may not need to develop as quickly. Energy can be allocated to the development of traits such as imaginal disc wing patterning before body planning is a necessity. However, the delay in expression of the Hox genes suggests a heterochronic shift in development, since it is not merely a delay in absolute time, but a delay in gene activation relative to defined orthologous stages.

The absence of Hox gene representation at St5 in *D. grimshawi* contrasts sharply will all other species. This raises the question of whether upstream members of the canonical anteroposterior segmentation cascade are also reduced in *D. grimshawi*. We find that St5 expression of segment polarity genes, including *wingless* (*wg*) and *hedgehog* (*hh*), is relatively low in *D. grimshawi* (fig. 9). This suggests that activation of the entire segmentation cascade may be heterochronically shifted in *D. grimshawi*.

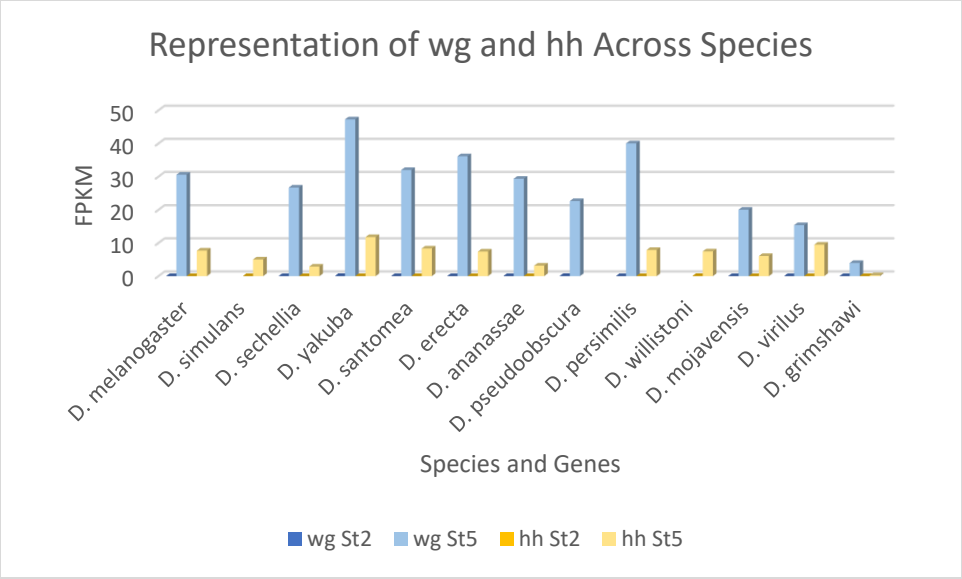


Figure 25) Representation across 13 species for segment polarity genes wg *and* hh. *St2 samples are shown in saturated colors, and St5 samples are shown in corresponding pastels.* wg *samples for* D. willistoni *and* hh *samples for* D. pseudoobscura *were not available.*

The heterochronic shift apparent in some of the Hox genes seems to be consistent upstream in the *Drosophila* development cascade. Segment polarity genes, the last member of the signal cascade to activate Hox genes, too show lower expression than expected. None of the species display any representation of the segment polarity genes in St2, but *D. grimshawi* shows lower representation of both *wingless* (*wg)* and *hedgehog* (*hh*) than any of the other species (fig. 10)

Potential Knockout Targets

This study aimed to find genes which can be used in future Crispr experiments. Although the Hox genes showed interesting differences in expression, in order to study novel gene function through knockout experiments the gene would need to be upregulated in the species of interest. A knockout of a gene that was already absent would not yield useful results. Other genes showed heightened expression, either at St2 or St5 alone or at both stages. Knockouts for these genes may prove viable into adulthood, which would make embryo collection more feasible. Knockout embryos may show how upregulation in these early phases contributed to island diversification. These genes play roles in sensory neurons, cytoskeletal structure, cell polarity, and wing development. While these genes were not formally grouped together, a future study can investigate their roles in well-studied protein pathways.

Neural Formation

Rutabaga (rut)

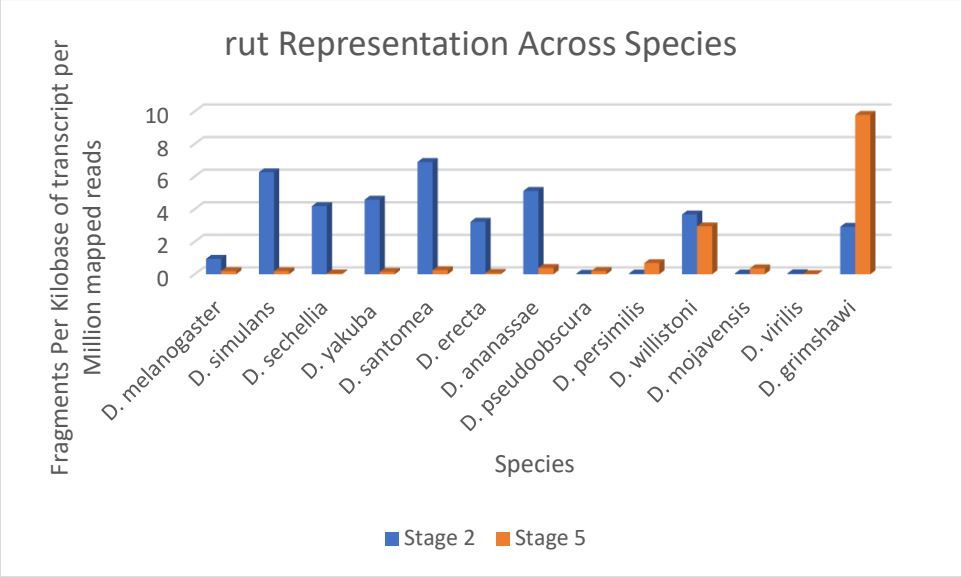


Figure 26) Representation of the rut gene across Drosophila species.

The *rut* gene encodes for a Ca2+/calmodulin-responsive adenylyl cyclase [56]. This protein is concentrated in the fly's mushroom bodies, which are believed to play a role in memory formation, especially in connection with the olfactory system [56]. Rut is either absent at both stages (as in *D. pseudoobscura* and *D. persmilis* in the *obscura* group) or deposited maternally with near-complete degradation by St5. Only in *D. grimshawi* does *rut* show evidence of zygotic transcription (fig. 11.) In *D.* melanogaster, *rut* is active in embryonic muscles and neuromuscular junctions [57]. While *rut* was shown earlier in *D. grimshawi* than in *D. melanogaster,* it is possible the early appearance of *rut* could serve the same functions in the embryo at an earlier stage. Continued expression may play a role in the early development of *D. grimshawi* olfactory systems or muscular junctions.

Cousin of *atonal* (*cato*)

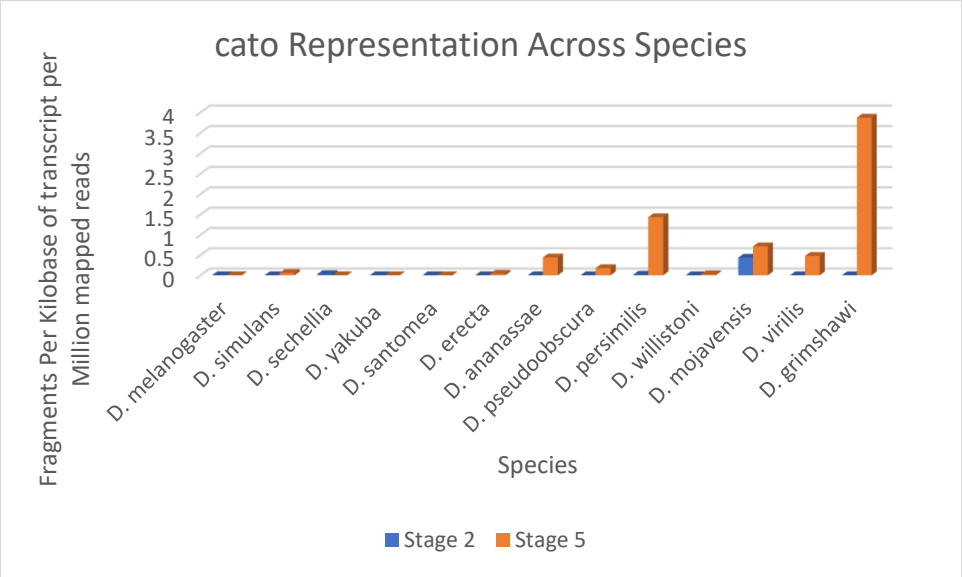


Figure 27) Representation of the cato gene across Drosophila species.

The gene *cato* plays a role in the development of precursors to sensory neural cells in the peripheral nervous system (PNS) [58]. It is active after proneural cells are selected, but before dendrite and axon terminals are differentiated [58]. Except in *D. sechellia*, *cato* is upregulated only in St5, if at all (fig. 12). While *D. persimilis* shows the next-highest St5 expression levels at 1.4 FPKM, *D. grimshawi* shows 3.9 FPKM at St5 (fig. 12). As in *rut*, the high representation of *cato* may reveal a priority for neural growth in the developing embryo over other developing systems. This could be the result of a safer environment—if there are fewer predators, there may not have been pressure to develop other genes right away, and early *cato* representation posed a benefit in *D. grimshawi.*

Cytoskeletal structure

milkah (mil)

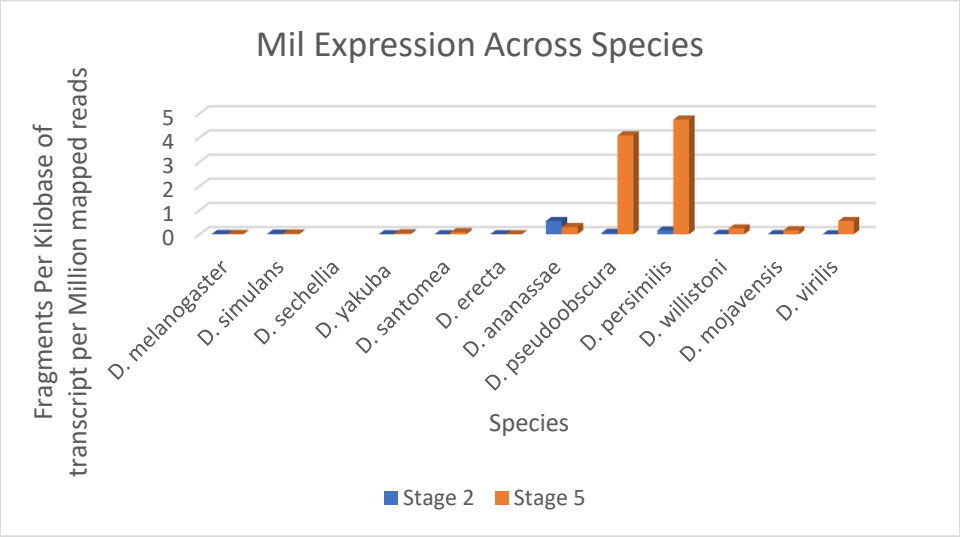


Figure 28) Representation of the mil gene across Drosophila species.

In adult flies, *mil* plays a role in the testes in spermatogenesis. It elongates the cytoskeleton and reorganizes histones to condense the nucleus in sperm cells [59]. *D. grimshawi* shows high levels of *mil* compared to the other species both at St2 and St5 (fig. 13). It is unlikely *mil's* role here was in spermatogenesis, as the pole cells that will go on to form the gonads have only just formed. *mil's* role in the nucleosome assembly protein (NAP) family contributes not only to spermatogenesis, but also to memory, mating behavior, and, in mammals, detoxification [59,60]. Other memory systems such as those encoded in rut already show some prioritization, so the upregulation of *mil* could be memory related. Another explanation lies in its putative detoxification role: since *D. grimshawi* take longer to develop, they spend more time in decaying substrate. Early *mil* representation may help *D. grimshawi* survive its longer developmental period by preparing for upcoming toxin exposure.

Zona Pellucida

The Zona Pellucida (ZPD) complex of proteins operates in the apical areas in developing epithelial cells [61]. These proteins control the interactions of the cell membrane and the extracellular matrix, managing actin formation and structure, and the formations of cellular extensions [61]. Genes in the ZPD such as *miniature* and *cypher* were found to show unusual representation patterns in *D. grimshawi*.

miniature (*m*)

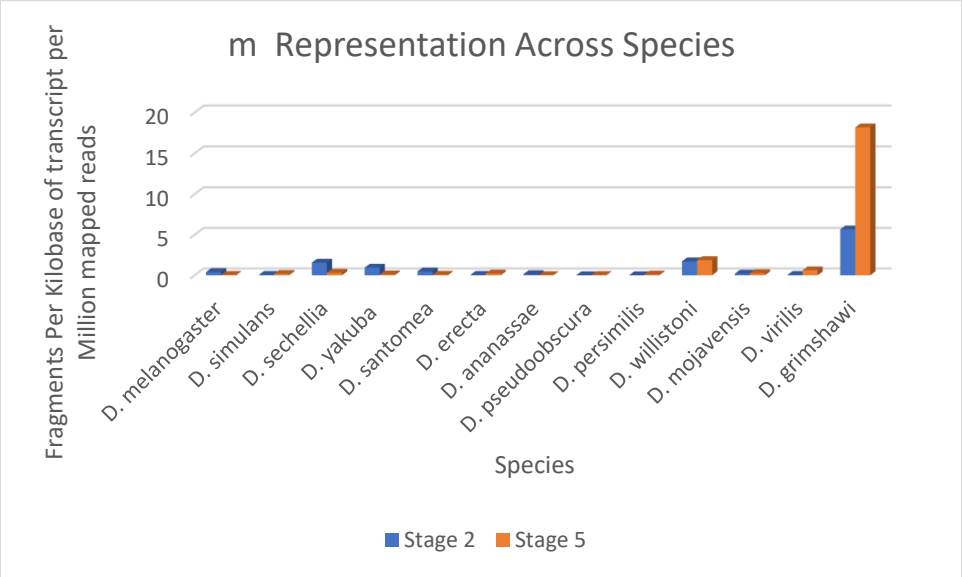


Figure 29) Representation of the m gene across Drosophila species.

The gene *miniature* interacts with other proteins in the zona pellucida complex to secrete the cuticle essential to the formation of the wing [62]. A decrease in *m* expression leads to a decrease in wing size [62]. *m* is expressed at 5.6 FPKM at St2 and 18.2 FPKM at St5 (fig. 14). No other species shows m expression over 2 FPKM at any stage (fig. 14). In *D. melanogaster, m* is not expressed until the later embryonic stages, and does not usually produce phenotypes at the early embryonic stage [57]. Given that *D. grimshawi* are so large, they may need stronger wings to fly. The high representation of *m* so early on could represent a head-start in wing formation to hold up to the large fly, or simply because it can because it is not transcribing other, more energetically expensive genes. *m* could play a role in establishing the wing patterns present in picture-wing flies.

Cypher (*cyr*)

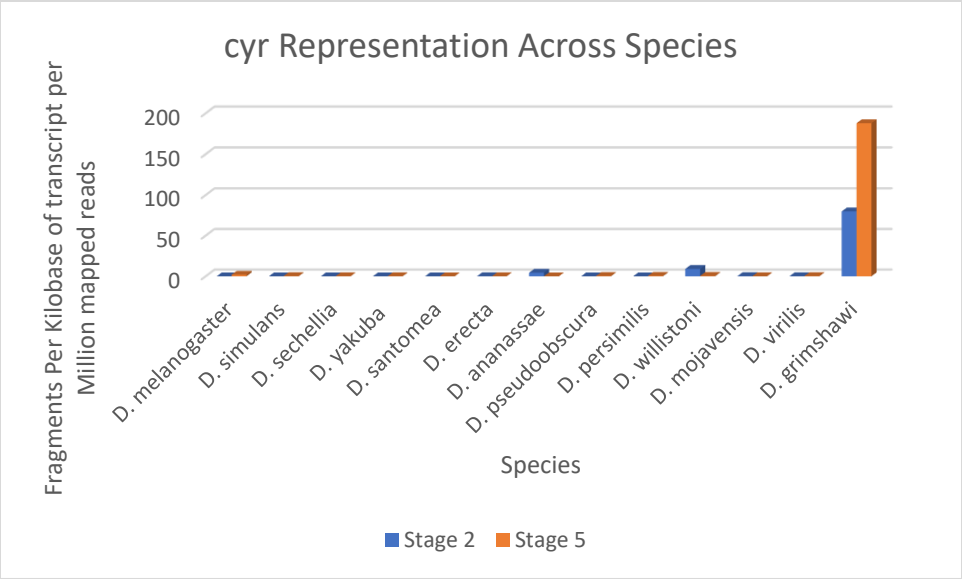


Figure 30) representation of the cyr gene across Drosophila species.

In *D. melanogaster*, *cyr* works with *m* in the ZPD, in dorsal cells at Stage 15 and in ventral stripes at Stages 16 and 17 [61]. It activates *m* transcription and plays a role in wing pigmentation [61]. *cyr* shows a pattern like *m*'s, showing high representation at both stages whereas other species show little to none (fig. 14, fig. 15). Like *m,* in *D. melanogaster, cyr* is lowly expressed in the early embryo, and no cyr-related phenotypes have been described at these early stages [57]. Since the ZPD proteins work in tandem to interact with the extracellular matrix, it may show upregulation due to *D. grimshawi*'s unique surface area-to-volume ratio. A stronger connection with the intracellular matrix may have been needed to contain the higher volume of the larger cell.

Multiple Wing Hairs (*mwh*)

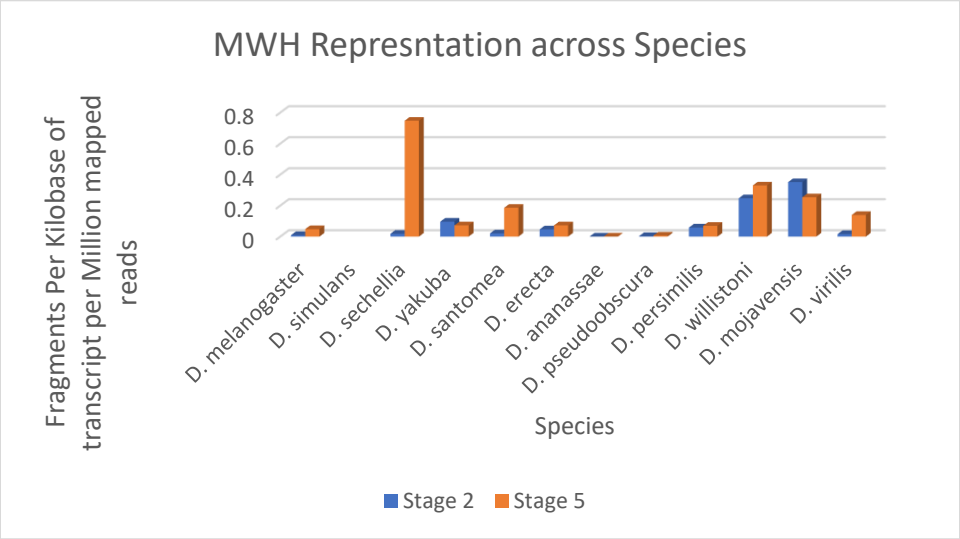


Figure 31) Representation of the mwh gene across Drosophila species.

The *mwh* gene regulates planar cell polarity (PCP) by activating cytoskeletal activity in an asymmetrical pattern [63]. In pupal flies, it forms two separate complexes which regulate the direction of growth and number of wing hairs on the distal and proximal sides of the wings, respectively [63]. Reduction in *mwh* expression leads to multiple wing hairs [63]. Other species show little to no representation of *mwh* at either stage, but *mwh* is present at 15.6 FPKM at St2 and 44.3 FPKM at St5 (fig. 16). Oddly, at the adult level, *D. grimshawi* wings appear to show multiple wing hairs and a great deal of variation in hair length and width corresponding with wing coloration. Despite the evidence of low expression later, MWH is strongly activated in the early embryo. While it may be related to wing formation, mwh's role in planar cell polarity may play another part in embryonic development.

Other Genes

ptr

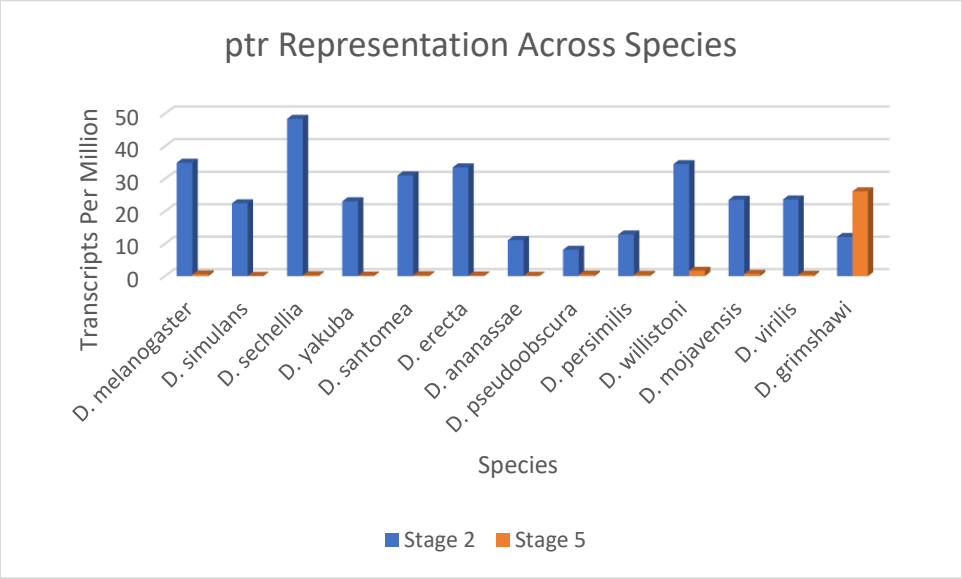


Figure 32) Representation of the ptr gene across Drosophila species.

The gene *ptr* (not to be confused with *Ptr*, or *Patched-Related*), is not yet well defined. In every other *Drosophila* species, *ptr* is upregulated in St2 and downregulated in St5 (fig. 17). Upstream genes may offer insight into its function.

S. anomala Results

Data quality analysis

BUSCO Scores

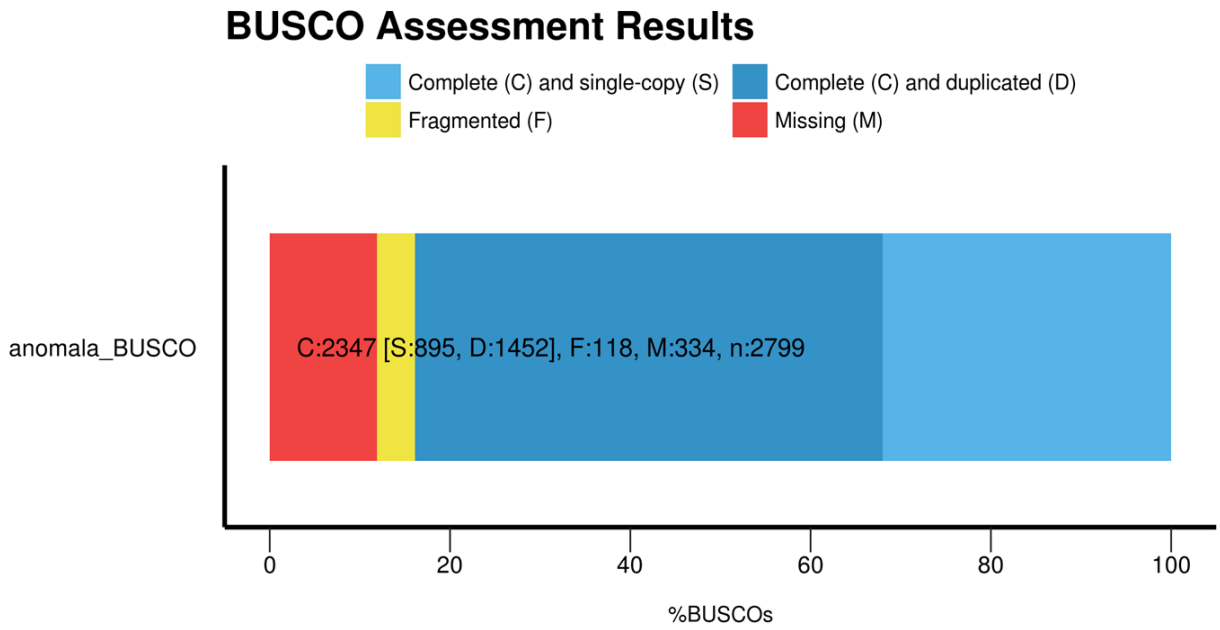


Figure 33) BUSCO Scores for S. anomala, using Diptera predictions as a reference.

A Benchmarking Universal Single-Copy Orthologs (BUSCO) assessment was used to ensure that the libraries contained a high number of highly conserved orthologs expected to be represented in similar species (fig. 18). This is typically used to ensure that levels of degradation are low and that expected genes are evenly represented in the library. Here, the orthologs compared to the library were taken from the order *Diptera* and would be the most specific comparison for any fly species. Fewer than 20% of the predicted genes were missing or fragmented (red and yellow, fig. 18). About one quarter of the dataset matched once to an ortholog predicted in dipterans (light blue, fig. 18), and about one half matched more than once (dark blue, fig. 18).

Most of the dataset reaches the standards predicted for a dipteran species. A *Drosophila melanogaster* gene set would match about 98% of the orthologs [48]. The match percentage would be expected to be lower in less-studied and more derived species [48]. BUSCO analysis is meant to be robust, so some orthologs may be marked as absent or degraded to avoid false positives. Given that levels of degradation are low and few of the expected orthologs are missing, the BUSCO assessment would indicate that the library is high-quality for a non-model, derived species and analysis can move forward.

Differential expression quality

After normalization using a trimmed mean of M values method, the maternal group contained 68,100,356 reads and St5 contained 52,133,247 reads. The lower read count in St5 is due in part to one

eliminated sample. These reads were attached to structured blocks of data to be composed into a total of 283,528 features. These features could be genes, gene isoforms, or an RNA product [64].

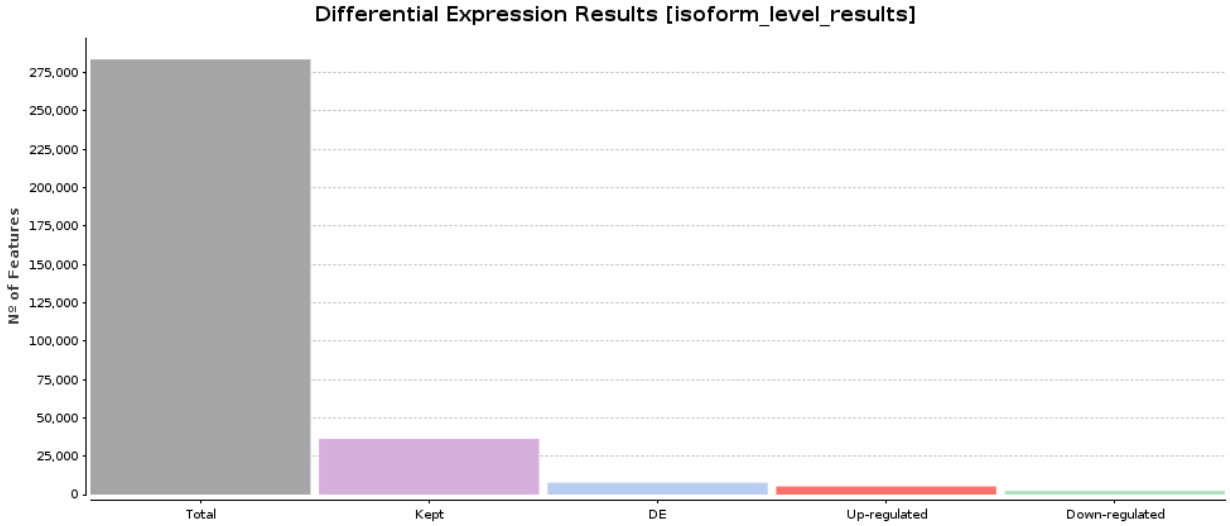


Figure 34) Results of differential expression at the isoform level. All features passing the filtering method are in lavender. Of those kept, the first blue bar represents all differentially expressed genes, and the red and blue bars represent upregulated and downregulated genes, respectively.

Library size and sequencing depth may cause inflation of certain transcripts without real differential expression. Before any sorting, the reads were structured into 283,528 features, shown in the grey bar (fig. 19). Before features can be shown to be differentially expressed, a filter is applied, sorting out any results with counts that round to zero in both the maternal stage and zygotic stage, since low expression across both stages cannot be differential. Next, differences in expression resulting from a sampling error, taking into account the size of each library, differences in representation resulting from uneven sequencing depth, and other scaling factors using a trimmed mean of M method [65,66]. After low expressions and bias-related changes in expression are removed, 36,232 features remained, shown by the lavender bar (fig. 19). EdgeR determined that of these, 7,708 were differentially expressed (FDR≤.05, blue) (fig. 19). 5,281 were upregulated in St5 (FDR≤.05, red) and 2,427 were downregulated in St5 (FDR≤.05, green) (fig. 19).

Of the initial features, under 13% were expressed at both stages and showed no misrepresentation from scaling factors in the libraries. While the initial libraries were high-quality, they were still susceptible to sampling error. Only 7,708 of the 36,232 trimmed features (about 20%) showed differential expression. This is to be expected, since many essential genes would function at similar levels at both stages. Over twice as many genes were upregulated at St5 than at St2. This would indicate higher expression of more genes later in development. Since there are more cells, each with increasingly differentiated functions as development moves forward, it makes sense that St5 would show a higher degree of upregulation.

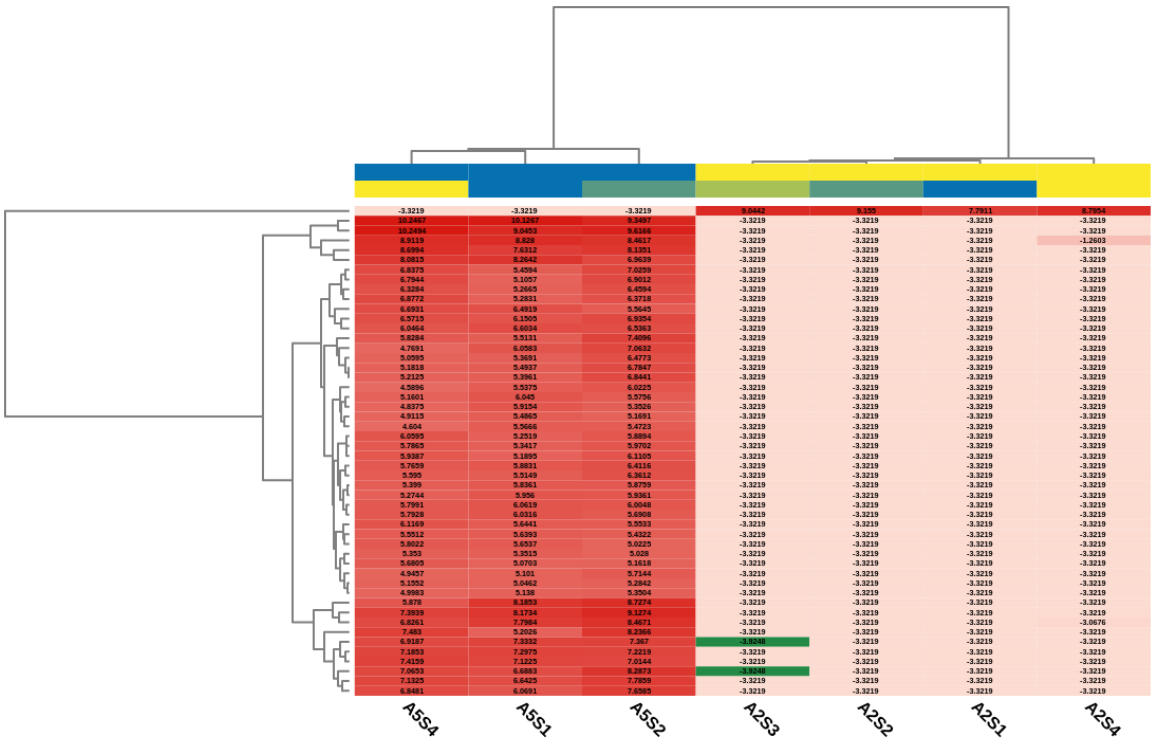


Figure 35) Heatmap of the 50 most differentially expressed gene isoforms, using St5 as the reference set. Green indicates lower levels of expression, and red indicates higher levels. The numbers in the cells are the log fold change for each isoform's representation.

The top 50 most differentially expressed isoforms were sorted using a pairwise differential expression analysis in EdgeR and ranked by log10 of the fold change (fig. 20). St5 was used as the reference set, so upregulation here means that a transcript was upregulated in St5. Upregulation is indicated in red, and downregulation is indicated in green. Darker shades of red indicate higher upregulation than lighter shades. The first three columns represent the St5 samples, and the last four columns represent the samples from St5 (fig. 20). The names on the right are the names of the differentially expressed transcripts (fig. 20). Clustering of the gene sequences are represented in the dendrogram on the left, and clustering between the samples are represented by the dendrogram on the top (fig. 20). The samples cluster into two main groups—maternal stage (2) and zygotic stage (5) (fig. 20). This supports the initial premise that there is, in fact, definitive differential expression between these two stages and that the samples have more in common within their groups than outside of their groups.

Of the top 50 transcripts, only one is upregulated in the maternal stage (fig. 20). The St2 samples of this gene show consistent, red shading and the St5 samples show a consistent pink—the coloring and log change for all three samples are identical (-3.3219) (fig. 20). When the zygotic genes are upregulated, the St2 groups show a narrower variation of shading than their sSt5 counterparts, indicating that the St2 genes are expressed more consistently between samples than St5 genes (fig. 20). Since the maternal deposits are dependent on only one process—maternal deposition—and the St5

deposits are dependent on maternal deposition as well as zygotic transcription and degradation, it makes sense that St5 samples would separate farther from one another. The dendrogram shows that the distance between the maternally upregulated gene is farther between the maternally upregulated transcript and the zygotically upregulated transcripts than the distance between any two of the remaining zygotically upregulated genes (fig. 20). This implies that, first, more genes are upregulated in St5 than in St2, and that the independent zygote expresses more genes than those deposited by the mother. Second, the sequences of the genes in St5 are more like other St5 genes than the sequences of the St2 genes, pointing towards a need for different genes and functions at St5, and a greater variety of genes represented later in development. Third, downregulation is stronger in St2 than St5—when St2 shows upregulated genes, they are eliminated in St5. St2 genes still show some representation when their St5 counterparts are upregulated. This could be a result of the maternal and zygotic deposits breaking down the initially maternally deposited genes.

a)

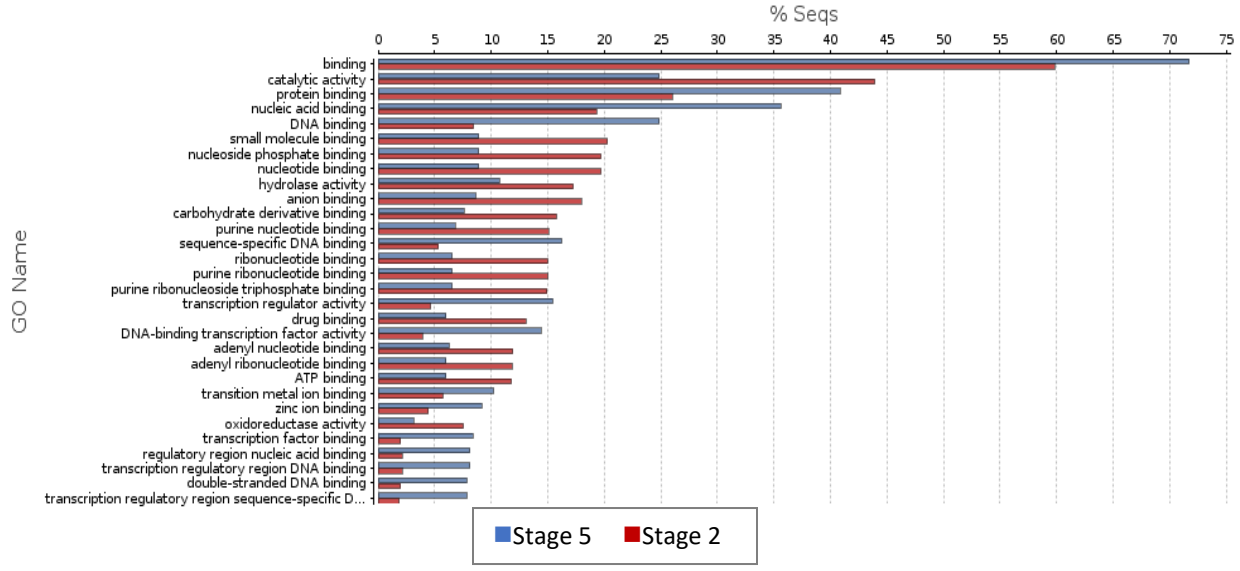
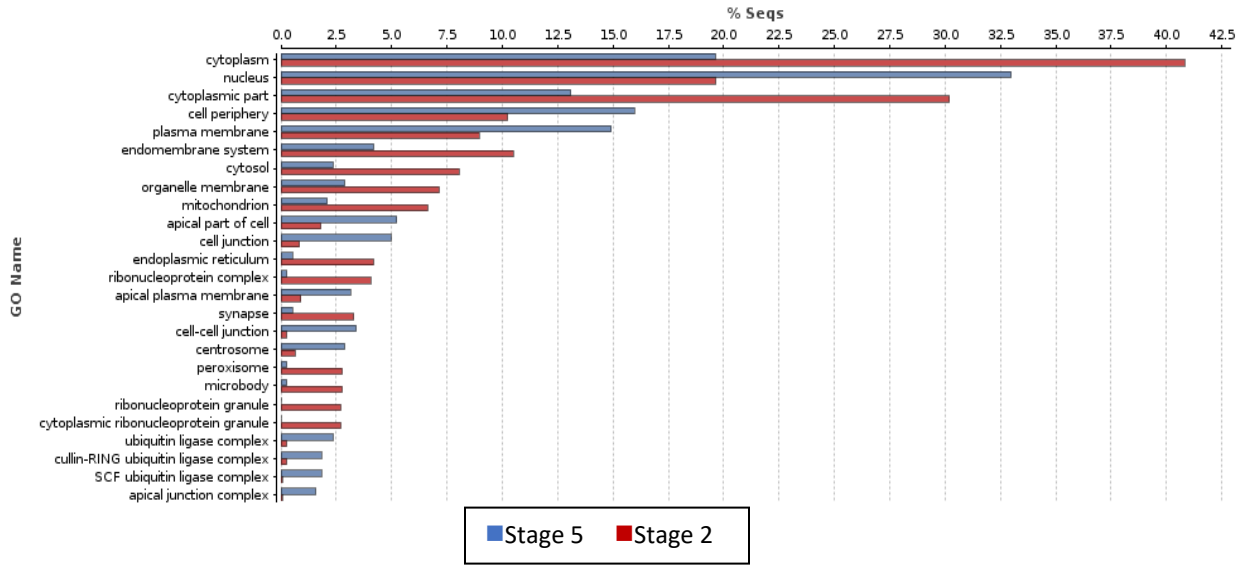
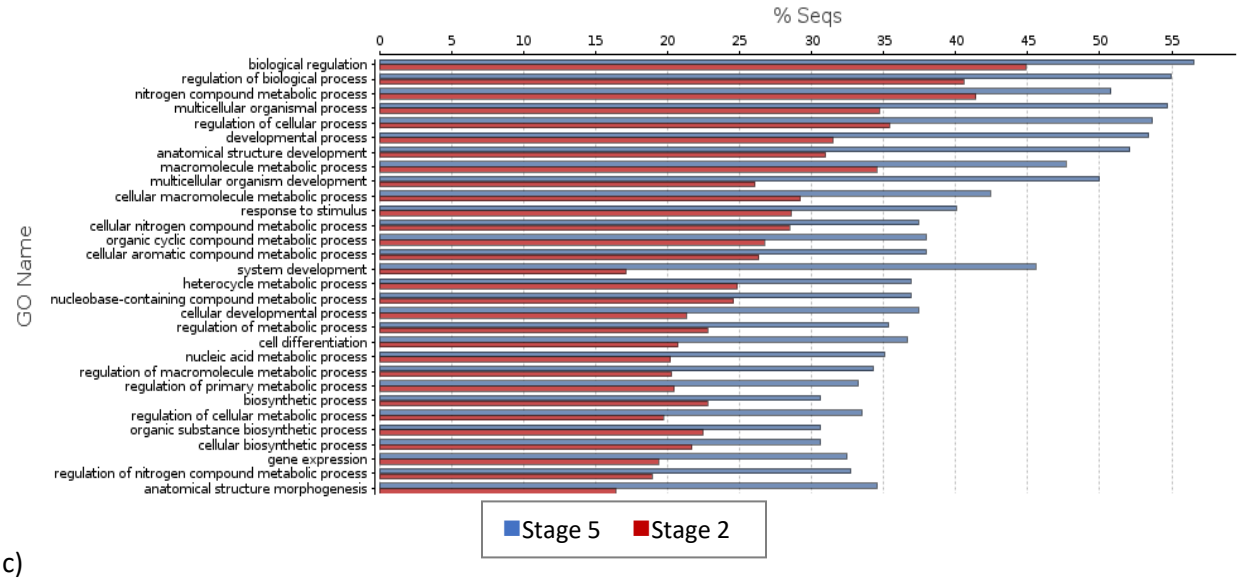


Figure 36) GO enrichment of zygotically upregulated gene transcripts. The X axis shows percentage of annotated sequences associated with a GO term, with zygotic enrichment as the test set in blue and maternal enrichment as the reference set in red. The Y axis represents the significantly enriched GO terms (FDR ≤ .05). A) molecular function-level GO enrichment of upregulated gene products when St5 is the primary target at the lowest level of specificity.

b)





c)

Figure 21 Cont.) B) Cellular component zygotic GO enrichment. C)Biological process GO enrichment.

A Fisher's Exact Test was used to determine statistical significance of GO enrichment. Tests were performed at the lowest level of specificity to explore the broadest GO priorities at the molecular function, cellular component, and biological process levels.

At the first molecular function level, molecular binding monopolized both zygotic and maternal function, with 60% and 70% of transcripts playing a binding role, respectively (fig. 21 A). Nucleic acid binding increases from 19 to 36 percent between stages 2 and 5, pointing to an increase in processes using DNA such as replication or transcription (fig. 21 A). Catalytic activity slowed in St5 (fig. 21 A). If the catalytic activity involved the breakdown of maternal deposits, this could indicate that the maternal RNA is mostly broken down. ATP binding was also emphasized in earlier stages, suggesting higher

energy use at earlier stages (fig. 21 A). Many metabolic processes are enriched at St5 (fig. 21 A), which could indicate a priority on growth.

At the cellular component level, there is a major shift from activity in the cytoplasm to activity in the nucleus. Activity reduced in the cytoplasm from the maternal stage at 41% to 19% at the zygotic stage (fig. 21 B), which could be a result of the overall decrease in the ratio of cytoplasm-related genes after the construction of cell membranes and the increase in nucleus and organelle-based activity. Plasma membrane-linked genes rose from 9% to 15%, following the pattern which would occur as the syncytium is divided into individual cells (fig. 21 B). Ribonucleoprotein activity spiked, which would assist in creating the proteins necessary to translate the newly produced RNA and encourage cellular growth (fig. 21 B).

The biological functions of the transcript confirm many of the patterns the molecular and cellular transcripts suggest. While the other GO categories showed upregulation in both the maternal and zygotic stages, the biological processes showed increases in the zygotic stage across the board (fig. 21 C). Biological regulation was the highest priority at both stages, and the zygotic stage showed increases in several types of metabolism, cell differentiation, nucleic acid production, and cellular and system growth (fig. 21C). As the zygote grows, the biological processes grow more specific, laying the groundwork for organ systems and stimulus response (fig. 21 C). The increase in cellular growth and macromolecular production corroborates the signs of growth present in the previous GO levels.

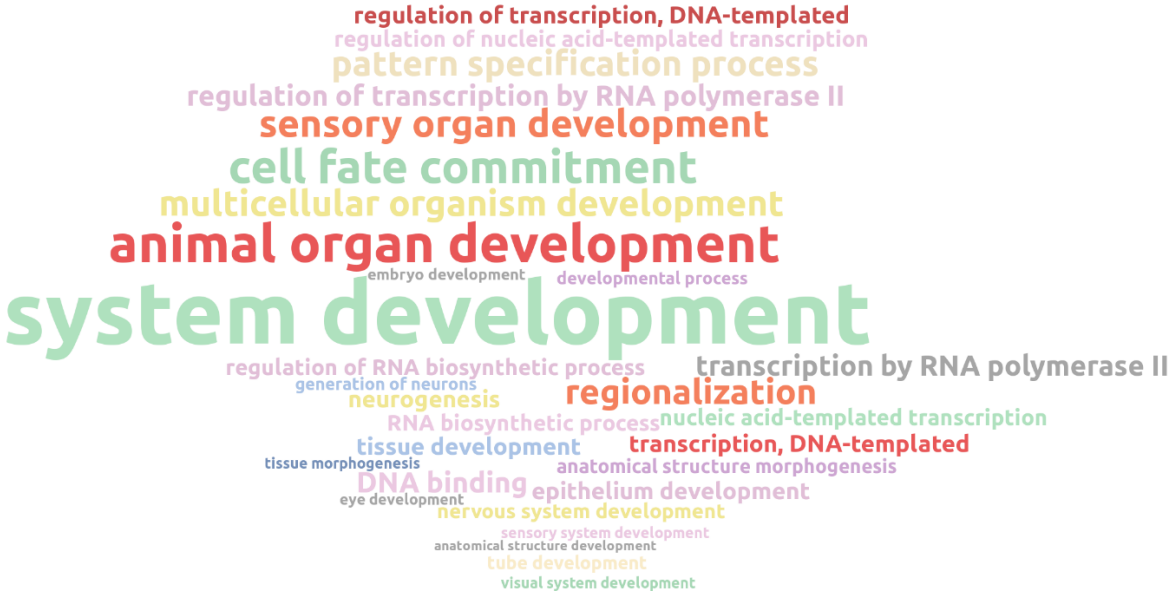


Figure 37) A word map of GO processes highlighted in S. anomala's transcriptome

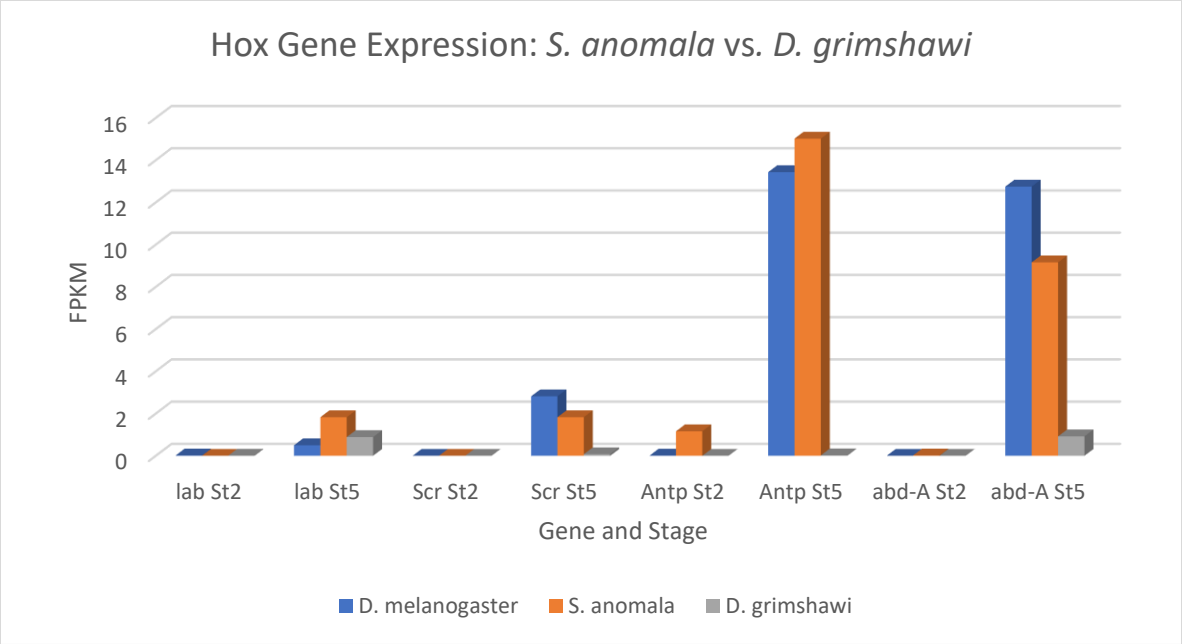


Figure 38) Representation of lab, Scr, Antp, and abd-A in S. anomala, D. grimshawi, *and* D. melanogaster *for comparison.* D. grimshawi *and* D. melanogaster *gene levels are represented in FPKM and the* S. anomala *gene levels in transcripts per million (TPM). Not pictured is Dfd, which occurs at 56 FPKM at St2 in* S. anomala *and below one in the other species, and 18, 251, and below 1 FPKM in St5 in D. melanogaster, S. anomala, and D. grimshawi respectively.*

When comparing the Hox genes found to show low representation in *D. grimshawi* to the same genes in *S. anomala, S. anomala* tends to show patterns of representation more like those in mainland fly species (fig. 23). *S. anomala* showed increases in St5 representation of all the Hox genes previously described in *D. grimshawi.* This could indicate a higher degree of divergence between Hawaiian *drosophila* and mainland flies than *Scaptomyza* and mainland species.

Discussion

Observations on the early *S. anomala* transcriptome

When observing which gene ontologies were upregulated at the zygotic stage, there is an emphasis on transcription and organ development. The upregulation would be expected—since so little of the RNA in the zygote remains from the mother, it would follow that the zygote is now transcribing its own RNA. Even though the cells are still mostly undifferentiated in St5, the beginnings of cell differentiation are becoming evident. Organ development, cell fate determination, and system development are all taking place. Genes involved in brain development, sensory organs, and neuron development are strongly represented (fig. 22). While many of the highly conserved zygotic-only genes remain in *S. anomala*, those that show changes to the expected patterns may add additional layers to the insights into the zygote's priorities. Future studies can compare zygotic-stage priorities of other species with modified zygotic-only RNA complements to explore whether this upregulation expedites neural development.

Other changed, commonly conserved genes could point to other gene ontology changes to explore in transcriptomic studies. Future experiments could focus on structural differences in *Scaptomyza* brains and sensory organs, as well as other systems brought up by unusually maternally enriched genes. Before making any decisions on genes in *S. anomala* which could be experimentally modified, further comparisons need to be carried out to determine which genes differ in representation between species, rather than the genes that differed within species in this preliminary study.

The unique transcriptomics of *D. grimshawi*

Since levels of transcript representation are more likely to evolve than protein-coding sequence, it was expected that some of the genes with novel gains in representation of zygotic transcription would have been transcription factors. This turned out not to be the case, although several the losses of gene representation at stage 5 (including the Hox genes), are transcription factors. While the downward changes in representation would be unremarkable if observed in absolute time, the fact that *D. grimshawi* has re-prioritized its RNA production at a conserved staged is a heterochronic shift.

While mainland species allocate their energy towards the production of Hox genes and their upstream components, *D. grimshawi* uses the same developmental stages to focus on genes related to neuromuscular communication, planar cell polarity, and interactions between cell membranes and the intracellular matrix. This shift in priority could be the result of different facets of island evolution. The increase in *D. grimshawi's* body size shifts the proportion of cells in the syncytium to the volume of the extracellular matrix. In order to accommodate the size change, the proteins in the ZPD may need to be activated sooner in order to maintain the embryo's structural integrity and continue shaping the interior of the embryo to accommodate the next stages in development. Lower competition and predation in island environments reduce the need to rapidly reproduce. *D. grimshawi* therefore have longer lifespans, taking longer to hatch, develop to adulthood, and reach sexual maturity. Some of the changes in gene representation could simply be the result of the embryo taking its time: with no pressure to develop and reproduce before being eaten, there may have been no evolutionary push to filter out fly species with rapidly developing body plans. Other changes could be the result of this slow development: since developing embryos and larvae spend more time in the larval substrate, they may have prioritized the production of proteins that provide protection against toxins or oxidation. In this

case, there may be a selective pressure, but to prevent out-competition by bacteria rather than other flies. Whatever the root causes of these changes are, continued investigation of the transcriptome and other fly species will help discover the priorities of Hawaiian *Drosophilid* development.

Future studies on Hawaiian *Drosphila* transcriptomes

Future studies will include knockout models in *D. grimshawi* flies, continued transcriptomic investigation of Hawaiian *Drosophila* and *Scaptomyza* species, and possibly studies looking beyond the representation of protein-coding genes. Not all the open reading frames found matched an ortholog, so revisiting these unpaired genes could expand the base of information.

Animal genomes can splice exons in different combinations to create different isoforms of a protein. An analysis including the variations in splicing can show which isoforms dominate during certain stages and can illuminate the roles of dominant gene fragments during their respective developmental periods. The availability of alternative isoforms can improve genome construction for D. grimshawi and inform future genome construction for the *Scaptomyza* species.

While many of the genes studied could be candidates for a knockout study, the Hox genes will need to be studied by upregulation. Aside from the fact that a knockout model of an already-underrepresented gene would not prove very exciting, these genes are vital to the proper development of the fly and may likely prove lethal in a knockout model. While these genes were not excellent candidates for CRISPR experiments, they could point towards a pattern of energy allocation resulting from *D. grimshawi'*s island evolution. With fewer selective pressures, the *D. grimshawi* embryos may not have to begin the processes of body patterning as quickly, or they may have to complete other steps first to compensate for their larger sizes. Additional transcriptomic analysis can be performed at later stages to find the point at which the Hox genes begin to appear in *D. grimshawi. Cato* starts St2 with no representation but shows a steep representation increase in St5, so it may be a good gene to start the knockout studies, since only St5 representation will change compared to the control.

The variety of results in this initial exploration opens numerous possibilities into future studies of the evolution of Hawaiian *Drosophilids*. With multiple genes to explore, more species to investigate, and a mysterious evolutionary path to trace, the transcriptomic investigation of Hawaiian flies can reveal information about both this Dipteran lineage and the overall patterns of island migration and evolution.

Works Cited

1. Kaneshiro, K. Y. R. C. L. Perkins' legacy to evolutionary research on Hawaiian Drosophilidae (Diptera). *Pac. Sci.* **51**, 450–461 (1997).
2. Lapoint, R. T., O'Grady, P. M. & Whiteman, N. K. Diversification and dispersal of the Hawaiian Drosophilidae: The evolution of Scaptomyza. *Mol. Phylogenet. Evol.* **69**, 95–108 (2013).
3. O'Grady, P. & DeSalle, R. Out of Hawaii: the origin and biogeography of the genus Scaptomyza (Diptera: Drosophilidae). *Biol. Lett.* **4**, 195–199 (2008).
4. Bellemain, E. & Ricklefs, R. E. Are islands the end of the colonization road? *Cell* **23**, 461–467 (2008).
5. Obbard, D. J. *et al.* Estimating Divergence Dates and Substitution Rates in the Drosophila Phylogeny. *Mol. Biol. Evol.* **29**, 3459–3473 (2012).
6. Carroll, S. B. Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell* **134**, 25–36 (2008).
7. Gilbert, S. F. *Developmental Biology, Tenth Edition*. (Sinauer Associates, Inc., 2013).
8. Tadros, W. & Lipshitz, H. D. The maternal-to-zygotic transition: a play in two acts. *Development* **136**, 3033–3042 (2009).
9. Bushati, N., Stark, A., Brennecke, J. & Cohen, S. M. Temporal reciprocity of miRNAs and their targets during the maternal-to-zygotic transition in Drosophila. *Curr. Biol.* **18**, 501–506 (2008).
10. Giraldez, A. J. microRNAs, the cell's Nepenthe: clearing the past during the maternal-to-zygotic transition and cellular reprogramming. *Curr. Opin. Genet. Dev.* **20**, 369–375 (2010).
11. Atallah, J. & Lott, S. E. Evolution of maternal and zygotic mRNA complements in the early Drosophila embryo. *PLOS Genet.* **14**, e1007838 (2018).
12. Bownes, M. A photographic study of development in the living embryo of Drosophila melanogaster. *J. Embryol. Exp. Morphol.* **33**, 789–801 (1975).
13. Kuntz, S. G. & Eisen, M. B. Drosophila Embryogenesis Scales Uniformly across Temperature in Developmentally Diverse Species. *PLOS Genet.* **10**, e1004293 (2014).
14. Graveley, B. R. *et al.* The developmental transcriptome of Drosophila melanogaster. *Nature* **471**, 473–479 (2011).
15. Losos, J. B. & Ricklefs, R. E. Adaptation and diversification on islands. *Nature* **457**, 830–836 (2009).
16. Whittaker, R. J., Fernandez-Palacios, J. M., Matthews, T. J., Borregaard, M. K. & Triantis, K. A. Island biogeography: Taking the long view of nature's laboratories. *Science* (2017).
17. Magnacca, K. N., Foote, D. & O'Grady, P. M. A review of the endemic Hawaiian Drosophilidae and their host plants. *Zootaxa* **1728**, 1–58 (2008).
18. Stark, J. B. & O'Grady, P. M. Morphological variation in the forelegs of the Hawaiian Drosophilidae. I. The AMC clade. *J. Morphol.* **271**, 86–103 (2010).
19. Hardy, D. E. & Kaneshiro, K. Y. A review of the modified tarsus group of Hawaiian Drosophila (Drosophilidae: Diptera). *Proc. Hawaii. Entomol. Soc.* **XII**, 71–90 (1979).
20. Magnacca, K. N. & O'Grady, P. M. A Subgroup Structure for the Modified Mouthparts Species Group of Hawaiian Drosophila. (2006).
21. Hardy, E., Kaneshiro, K. Y., Val, F. C. & O'Grady, P. Review of the haleakalae species group of Hawaiian Drosophila (Diptera: Drosophilidae). *Bish. Mus. Bull. Entomol.* **9**, 1–136 (2001).
22. O'Grady, P. M. *et al.* Phylogenetic and ecological relationships of the Hawaiian Drosophila inferred by mitochondrial DNA analysis. *Mol. Phylogenet. Evol.* **58**, 244–256 (2011).
23. Edwards, K. A., Doescher, L. T., Kaneshiro, K. Y. & Yamamoto, D. A Database of Wing Diversity in the Hawaiian Drosophila. *PLOS ONE* **2**, e487 (2007).
24. Sarikaya, D. P. *et al.* Reproductive capacity evolves in response to ecology through common developmental mechanisms in Hawaiian Drosophila. *bioRxiv* 470898 (2018) doi:10.1101/470898.

25. O'Grady, P. & DeSalle, R. Hawaiian Drosophila as an Evolutionary Model Clade: Days of Future Past. *BioEssays* **2018**, (2018).
26. Magnacca, K. N. & Price, D. K. Rapid adaptive radiation and host plant conservation in the Hawaiian picture wing Drosophila (Diptera: Drosophilidae). *Mol. Phylogenet. Evol.* **92**, 226–242 (2015).
27. Stark, A. *et al.* Discovery of functional elements in 12 Drosophila genomes using evolutionary signatures. *Nature* **450**, 219–232 (2007).
28. González, J., Lenkov, K., Lipatov, M., Macpherson, J. M. & Petrov, D. A. High Rate of Recent Transposable Element–Induced Adaptation in Drosophila melanogaster. *PLoS Biol.* **6**, e251 (2008).
29. Goldman-Huertas, B. *et al.* Evolution of herbivory in Drosophilidae linked to loss of behaviors, antennal responses, odorant receptors, and ancestral diet. *Proc. Natl. Acad. Sci.* **112**, 3026–3031 (2015).
30. Harbaugh, D. T. & Baldwin, B. G. Phylogeny and biogeography of the sandalwoods (Santalum, Santalaceae): repeated dispersals throughout the Pacific. *Am. J. Bot.* **94**, 1028–1040 (2007).
31. Katoh, T., Izumitani, H. F., Yamashita, S. & Watada, M. Multiple origins of Hawaiian drosophilids: Phylogeography of Scaptomyza Hardy (Diptera: Drosophilidae). *Entomol. Sci.* **2017**, 33–44 (2016).
32. Craddock, E. M. Profuse evolutionary diversification and speciation on volcanic islands: transposon instability and amplification bursts explain the genetic paradox. *Biol. Direct* **11**, 44 (2016).
33. Hardy, D. E. *Insects of Hawaii*. vol. 12 (University of Hawaii Press, 1965).
34. Frota-Pessoa, O. Bunostoma brasiliensis n. sp. (Drosophilidaelong dashDiptera). *SUMMA Bras. Biol.* **1**, 175–178 (1946).
35. Grimaldi D. New Distributional Records of Scaptomyza-Australis from South Pacific Islands and Biogeographic Implications. *J. N. Y. Entomol. Soc.* **98**, 484–488 (1990).
36. Hawaiian Drosophila | The National Drosophila Species Stock Center. http://blogs.cornell.edu/drosophila/hawaiian-drosophila/.
37. Consortium, T. modENCODE *et al.* Identification of Functional Elements and Regulatory Circuits by Drosophila modENCODE. *Science* **330**, 1787–1797 (2010).
38. Wheeler, M. R. & Clayton, F. A new Drosophila culture technique. *Drosoph. Inf. Serv.* **40**, 98 (1965).
39. Fabrick, J. A. & Hull, J. J. Assessing Integrity of Insect RNA. (2017).
40. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10–12 (2011).
41. Trapnell, C. *et al.* Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* **7**, 562–578 (2012).
42. Kim, D. *et al.* TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36 (2013).
43. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
44. Ghosh, S. & Chan, C.-K. Analysis of RNA-Seq Data Using TopHat and Cufflinks. in *Plant Bioinformatics* (ed. Edwards, D.) 339–361 (Springer New York, 2016). doi:10.1007/978-1-4939-3167-5_18.
45. R Core Team. R: A language and environment for statistical computing. *R Found. Stat. Comput. Vienna Austria Httpwww R-Proj. Org* (2013).
46. Abraham, S. *FastQC: a quality control tool for high throughput sequence data*. (2010).
47. Grabherr, M. G. *et al.* Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nat. Biotechnol.* **29**, 644–652 (2011).
48. Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015).

49. Dikow, R. B., Frandsen, P. B., Turcatel, M. & Dikow, T. Genomic and transcriptomic resources for assassin flies including the complete genome sequence of Proctacanthus coquilletti (Insecta: Diptera: Asilidae) and 16 representative transcriptomes. *PeerJ* **5**, e2951 (2017).
50. Götz, S. *et al.* High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res.* **36**, 3420–3435 (2008).
51. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
52. Stanke, M. & Morgenstern, B. AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Res.* **33**, W465–W467 (2005).
53. Stanke, M. *et al.* AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Res.* **34**, W435–W439 (2006).
54. Chen, Z.-X. *et al.* Comparative validation of the D. melanogaster modENCODE transcriptome annotation. *Genome Res.* **24**, 1209–1223 (2014).
55. Atallah, J. & Lott, S. E. Evolution of maternal and zygotic mRNA complements in the early Drosophila embryo. 29.
56. Han, P.-L., Levin, L. R., Reed, R. R. & Davis, R. L. Preferential expression of the drosophila rutabaga gene in mushroom bodies, neural centers for learning in insects. *Neuron* **9**, 619–627 (1992).
57. Gelbart, W. M. & Emmert, D. B. FlyBase High Throughput Expression Pattern Data. *Fly Base Anal.* (2013).
58. Goulding, S. E., White, N. M. & Jarman, A. P. cato Encodes a Basic Helix-Loop-Helix Transcription Factor Implicated in the Correct Differentiation of Drosophila Sense Organs. *Dev. Biol.* **221**, 120–131 (2000).
59. Shuhei Kimura. The Nap family proteins, CG5017/Hanabi and Nap1, are essential for Drosophila spermiogenesis. *FEBS Lett.* 922–929 (2013).
60. Kuzin, B. A. *et al.* Combination of Hypomorphic Mutations of the Drosophila Homologues of Aryl Hydrocarbon Receptor and Nucleosome Assembly Protein Family Genes Disrupts Morphogenesis, Memory and Detoxification. *PLoS ONE* **9**, e94975 (2014).
61. Fernandes, I. *et al.* Zona Pellucida Domain Proteins Remodel the Apical Compartment for Localized Cell Shape Changes. *Dev. Cell* **18**, 64–76 (2010).
62. Roch, F. Drosophila miniature and dusky encode ZP proteins required for cytoskeletal reorganisation during wing morphogenesis. *J. Cell Sci.* **116**, 1199–1207 (2003).
63. Yan, J. *et al.* The multiple-wing-hairs Gene Encodes a Novel GBD–FH3 Domain-Containing Protein That Functions Both Prior to and After Wing Hair Initiation. *Genetics* **180**, 219–228 (2008).
64. Sequence Features. https://www.ncbi.nlm.nih.gov/IEB/ToolBox/SDKDOCS/SEQFEAT.HTML.
65. Robinson, M. D. & Oshlack, A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* **11**, R25 (2010).
66. Blast2GO - Expression quantification and differential expression analysis. https://www.blast2go.com/support/blog/22-blast2goblog/206-expression-quantification-and-differential-expression-analysis.

Vita

The author was born in New Orleans, Louisiana as the second of four sisters. She obtained her diploma from Lusher Charter High School in 2012 and her bachelor's degree in Neuroscience with minors in Chemistry and Women's and Gender Studies from Agnes Scott College in December 2015. Following her mother and sister, she joined the University of New Orleans Graduate Program. She pursued an MS in Biological Sciences, joining Dr. Joel Atallah's lab in 2016. She is the proud parent of two dogs and thousands of fruit flies.