On The Evolution of Male And Female Reproductive Organs
In Flies Of The Genus Drosophila

By

GIOVANNI HANNA
DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Genetics

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

David Begun, Chair

Susan Lott

Bruce Draper

Committee in Charge

2022

ACKNOWLEDGEMENTS

I would like to thank my advisors David Begun and Artyom Kopp for our stimulating discussions of current developments in the fields of evolutionary and developmental biology, and for preparing me to become a better scientist and science communicator. I truly enjoyed writing several grants with both of my advisors and those skills are invaluable.

Thank you so much to Julie Cridland and Olga Barmina for helping me navigate challenging PhD projects. Julie was instrumental to the bioinformatic analyses that made these transcriptome studies possible, and Olga was always ready to help me troubleshoot research issues along the way. I also want to thank David Luecke and José Miguel Uribe for their help in interpreting some of my results over the years. I appreciate all the support from my fellow graduate students Gavin Rice, Yige Luo, Alex Majane, Kae Lombardo, and Logan Blair who continue to be wonderful colleagues. Two brilliant undergraduate students at the time, Romina Hashemizadeh and Sabrina Lazar, were instrumental to pushing through very difficult projects and validating some great observations. I am thankful to you all.

I also thank Susan Lott and Bruce Draper for serving as dissertation committee members and for their immense moral support over the years. Susan and Bruce are some of the kindest people that I have ever met and we can all learn so much from their emotional intelligence and self-awareness. They are invaluable scientists and members of our campus and it was my pleasure to discuss science and more with them.

I also thank Teri Markow, Ignacio, and Maxi from my Masters lab for properly introducing me to the very big world of fruit fly genetics and ecology.

I would love to thank thousands of undergraduate students that I have taught over the years. You continue to inspire me to reach for my biggest goals and I look forward to teaching your children at the university level someday. You and I have learned so much from my classes and I thank you all for our engaging times in the classroom. You are my legacy.

This degree is dedicated to both of my parents. Since I was a child, Mom and Dad were honest when they couldn't answer my difficult questions and their inability to answer many questions pushed me to investigate Life in scientific ways during my more than 15 years of graduate and undergraduate schooling. Dad would have loved to hear these stories about the evolution of male and female reproductive organs in fruit flies and in animals in general, and he would have asked me all kinds of questions about it. Mom still gets the heebie-jeebies when I talk about insects and that will always entertain me endlessly.

I cannot thank my siblings Fabiola, Clarissa, and Sebastian enough, as well as Toni Rouhana, my niece Chloe, and Robert Scobie enough for holding on. Life wouldn't be as great as it is without you. I also want to thank my therapist for positively changing my life in countless and immeasurable ways.

Lastly, and very importantly, I got engaged during my PhD to the fabulous Erica Verde, who has made my life even more enjoyable, so much richer, and who sometimes calls me on my bad ideas before I spend so much time and energy on them. We are building a big life together and I wish everyone the courage to get out there and to find their partner.

ABSTRACT

The evolution of biological sexes is a fascinating field of study with much that remains to be learned. Decades of studies in Drosophila and in other animal models have made it clear that reproductive tissues in both sexes are some of the most rapidly evolving organs: They diverge more rapidly in gene expression relative to non-reproductive tissues, the genes that are biased in expression to reproductive tissues often show rapid divergence relative to non reproductive biased genes, and the reproductive organs themselves sometimes even show their capacity for physiological and behavioral changes, such as the ability of certain female reproductive organs to store sperm.

To better understand the genetic and genomic mechanisms that underlie the rapid evolution of reproductive tissues in response to local adaptation, and to study the extent of parallelism in gene expression in these tissues across species, I study the gene expression changes that underlie the accessory glands and testes across two geographic regions in three species, as seen in chapter one. Since much less is known about the female contributions to reproduction, and to begin to understand the rapid evolution of some of their tissues, I have begun to characterize the similarities and differences in gene expression across the seminal receptacle, the spermathecae, and the parovaria in eight species. This second chapter will help us better understand how certain female reproductive organs can gain and lose sperm storage functions across evolutionary time. Lastly, and to better understand how stem cell niches are maintained across species, I begin to characterize some of the gene expression differences across such niches in various species of Drosophila, and ask the question of whether such

niches use similar or different gene products for the maintenance of healthy stem cells in these niches.

Studying the rapid evolution of reproductive organs helps us characterize the shared gene expression patterns that are core to reproduction across distantly related species, while also elucidating the divergence in gene expression that allows for the rise of new molecules, physiologies, and behaviors. My three chapters of studies contribute to our understanding of male and female reproductive system function and elaborate on our limited knowledge of the evolution of female reproductive tissues. Studying the diversity of genetic and phenotypic features of closely and distantly related species will eventually allow us to better understand the co-evolution of male and female sexes across species, and will allow us to study the extent of gene expression plasticity across sexes: If different gene products related to successful fertilization are made by both males and females, how amenable are such reproductive genes to switching expression across sexes, and how common are such switches? And what effects do sex-specific selection and local adaptation have on shaping content and expression of a genome?

Chapter 1

Parallel Evolution of Reproductive Tissues across Geographic Regions in Males of the Genus *Drosophila*

INTRODUCTION

Connecting genotypes to phenotypes is one of the most outstanding contributions of biology to modern science. Given the current and sophisticated methods to better understand these connections, evolutionary biologists have gone back to studying older and unanswered questions in the field: How is it that different species can independently converge onto the same solution for a given problem? When species migrate to the same geographic region, how is it that populations of different species respond to the same general environmental differences and pressures, and what proportion of their responses depends on shared underlying genotypes? While morphological convergence does not necessarily imply developmental parallelism, many studies have shown support for the repeatability in use of gene regulatory elements, gene products, and developmental pathways in the building of the same phenotype across species, and sometimes even across distant relatives (Rokas and Carroll 2008, Castoe et al. 2009, Zhen et al. 2012, Gallant et a. 2014, Pankey et al. 2014, Pfenning et al. 2014, Jha et al. 2016, Witt et al. 2019, Young et al. 2019). However, the extent of gene re-use remains poorly understood across both short and long timescales, leaving open the question of independent gene reuse across time and space.

At the population level and in *D. melanogaster,* a recent study by Cridland et al. (2020) continues to support that male reproductive tissues exhibit more novel expression patterns in the accessory glands and express these novel phenotypes more consistently in the testes than in any other tissues studied. Such studies support the use of the testes and accessory glands for the study of rapidly evolving genotypes and phenotypes in the genus Drosophila, and can help

determine the extent of parallel gene expression changes associated with tissue differentiation across species. Above the population level, various Drosophila species show phenotypic latitudinal clines, and these include *D. melanogaster*, *D. simulans, and D. hydei,* which are often collected simultaneously in the field. While chromosomal and phenotypic differences do exist between these 3 species, similarities between them include generalist feeding habits throughout most of their geographic distribution, the recent spread of *D. simulans* and *D. hydei* across North America that coincides with that of *D. melanogaster* (David et Capy 1988, Singh et Long 1992, Fabian et al. 2012, Reinhardt et al. 2014), and their collective abilities to adapt to life at higher latitude climates such as those of Maine.

At the species level and across the large genus Drosophila, recent evidence gathered from whole male studies supports parallel latitudinal expression differentiation between *D. melanogaster* and *D. simulans* (Zhao et al. 2015). In a follow up study of the same two species as well as of *D. hydei,* Zhao and Begun (2017) were able to support parallelism for genes exhibiting latitudinal allele frequency differentiation within species as well as parallelism for genes exhibiting recurrent adaptive protein divergence between species. To further build on these studies and to better understand gene expression divergence associated with low and high latitude regions, we take a comparative transcriptomics approach where we characterize gene expression in the testes and in the accessory glands across males in populations of these 3 species and in 2 regions, Panama and Maine. This study allows us to study the extent of novel and potentially parallel gene expression changes in these rapidly evolving tissues, and to examine if adaptive processes that act at the population level can be re-used across species to respond to the same latitudinal differentiation pressures.

MATERIALS AND METHODS

Drosophila Sample Preparation

D. melanogaster, D. simulans, and *D. hydei* females were collected from Fairfield, Maine (September 2011, Latitude: 44°37'N) and Panama City, Panama (January 2012, Latitude: 8°58'N), placed individually in vials and then shipped to the laboratory where they were maintained as isofemale lines. These lines were maintained at 25°C on a 12-hour light:dark schedule and on a standard yeast-cornmeal-agar food medium, and virgin males were collected from all lines within 3 hours of eclosion. Because males in these species reach sexual maturity and are ready to court females at different ages, *D. melanogaster* males were aged to 4-5 days, *D. simulans* were aged to 4-6 days, and *D. hydei* males were aged to 14-16 days before dissection of their accessory glands and testes. I maximized the number of isofemale lines used to contribute to each biological replicate as follows: I used 12 Panama strains and 12 Maine strains for *D. melanogaster*, 13 Panama strains and 13 Maine strains for *D. simulans, and* 13 Panama strains and 12 Maine strains for *D. hydei.* Three males were collected per strain, aged after emergence to sexual maturity in low density vials, and then pooled prior to dissection to generate each RNA replicate. Three replicates were generated for each species and tissue, resulting in a total of 36 biological samples generated across the species studied.

RNA Extraction, Library Construction, and Sequencing

Within 10 minutes of dissection into cold PBS, the testes, accessory glands + anterior ejaculatory duct (henceforth referred to as accessory gland or AG) were transferred into cold Trizol and stored at -80°C before RNA was extracted following a standard Trizol-Chloroform extraction protocol, as follows. Tissues were carefully homogenized in 200 ul Trizol using a

sterile pestle and the Trizol volume was later adjusted to 1 ml. Aterwards, 200 ul of chloroform was added and the tube was shaken for 20 sec, followed by incubation for 5 min at room temperature. Samples were then centrifuged at 4°C and 13,000 rpm for 15 min and the upper phase was slowly collected. After addition of 1 ul glycogen, 500 ul isopropanol was added and the sample was mixed by gentle inversion. Samples were left at -20°C for 1 hr, after which nucleic acids were pelleted and then washed with 70% ethanol, followed by drying and resuspension in nuclease-free water. All samples were subjected to DNase digestion using the TURBO DNA-free kit (Ambion) following the manufacturer's protocol and the qualities of the resulting RNAs were studied using RNA Nano chip on a Bioanalyzer (Agilent).

Libraries were prepared using the NEBNext Ultra RNA Library Prep Kit for Illumina (New England Biolabs, Beverly, MA) with 1 ug total RNA as input for each replicate. The manufacturer's protocol was used with minor modifications. All AMPure bead elution steps were performed with PCRClean DX beads (Aline Biosciences), and qualities of libraries were estimated using the High Sensitivity DNA chip on a Bioanalyzer (Agilent). Libraries were sequenced on an Illumina HiSeq2000 to generate 90-bp paired-end reads. The focal set of genes used in all expression analyses is composed of 10628 1:1:1 orthologs identified in *D. melanogaster, D. simulans*, and *D. hydei* using MCL (van Dongen and Abreu-Goodger 2012) which uses a Markov cluster algorithm for assigning genes into families. Di- and polycistronic genes were omitted from the analysis, as were genes with one or more exons that were shared perfectly with another gene.

Sequence Alignment and Gene Expression Measures

Reads were aligned to the appropriate reference genome; *D. melanogaster* genome version 6.41 (FlyBase), *D. simulans* genome *v*ersion 3 (GCF_016746395.1), and *D. hydei* GCF_003285905.1 using Hisat2 (Kim *et al.* 2015) with default parameters. We used StringTie (Pertea *et al*. 2016) to calculate transcripts per million (TPM) for each gene in each transcriptome and measured differential expression with DESeq2 (Love et al. 2014). We also examined differential expression between 1:1:1 orthologs between *D. melanogaster, D. simulans,* and *D. hydei* populations*.* To do that, this collection of shared genes were identified using a table of *D. melanogaster* orthologs identified in other Drosophila (FlyBase, downloaded May 12, 2021). We generated a featureCounts table for both AG and testes to compare all *D. melanogaster* populations to *D. simulans* and *D. hydei,* and ran DESeq2 to identify differentially expressed genes as mentioned above. Genes were considered to show latitudinal DE for either tissue if we observed differential expression between the geographic regions, using a FDR adjusted P-value < 0.05 as the cutoff for differential expression, and also calculated absolute fold changes for all genes to estimate genome-wide expression variation.

Characterizing Tissue-Biased Genes

To identify accessory gland and testes-biased genes, we used the male data for several tissues from FlyAtlas2 (Leader et al. 2018). We defined tissue biased genes as those that A) had FPKM ≥ 1 in each tissue, B) exhibited the highest expression in the tissue of interest (accessory glands or testes) relative to other male tissues, and C) exhibited strong tissue bias with estimated *tau* (Yanai et al. 2005) ≥ 0.9.

Gene Ontology Analysis

Using the genes showing latitudinal DE in the AG and testes, we performed gene analyses using Gorilla (Eden et al. 2007, 2009) to identify processes that were elevated in these lists compared to a background of all genes expressed in the tissues of interest. We used the implementation of GOrilla on (http://cbl-gorilla.cs.technion.ac.il/) and the default p-value threshold.

RESULTS

Transcriptome Overview of Low and High Latitude Populations of *D. melanogaster*, *D. simulans*, and *D. hydei*

After filtering, we obtained 278 million paired reads for *D. melanogaster*, 604.5 million paired reads for *D. simulans*, and 609.7 million paired reads for D. hydei (S1 and S2 Tables). The significant difference in read numbers between *D. melanogaster* and the other two species is due to the fact that the *D. melanogaster* reads were sequenced in 2017 when fewer reads were obtained during sequencing runs, whereas reads from the other species are more than double because the sequencing was done in 2021. However, given that differential expression comparisons are within species, the normalization happens in inter-populations comparisons before any comparisons were made across species, and we therefore feel confident in comparing differentially expressed gene lists between geographic regions across the species studied.

Differential Gene Expression in High vs. Low Latitude Populations

In *D. melanogaster*, 813 (4.5%) and 100 (0.5%) genes were significantly differentially expressed between populations (False discovery rate (FDR) < 0.05) in the AGs and in the testes, respectively (Table 1, Fig 1). The magnitude of expression differences was significantly greater in the testes than in the AGs (Fig 2), a consistent pattern seen across tissues in all three species. To determine whether accessory gland and testes genes were more likely than other genes to be differentially expressed between geographic populations, we characterized tissue biased genes by estimating Tau (Yanai et al. 2005, Zhao et al. 2014), an index of tissue specificity, for each gene using FlyAtlas2 data (Leader et al. 2018), and found that a greater proportion of AG-biased genes (13.9% in *D. melanogaster*, 19.1% in *D. simulans*, and 33.2% in *D. hydei)* were differentially expressed between geographic populations compared to other genes (4.5% in *D. melanogaster*, 8.1% in *D. simulans*, and 11.9% in *D. hydei,* hypergeomoteric test, P < 0.0001, Table 1). This is consistent with previous results that AG-biased genes tend to exhibit greater expression variation than do other genes (Wagstaff & Begun 2005, Wagstaff & Begun 2007, Almeida & DeSalle 2008), and this observation is further supported by the expression fold changes that suggest that the testes are relatively more constrained in gene expression relative to the AGs in all three species (Panels A and C, Fig 1).

In *D. simulans*, 1071 (6.8%) and 684 (4.3%) genes were significantly differentially expressed between populations (False discovery rate (FDR) < 0.05) in the accessory glands and the testes, respectively (Table 1, Fig 1). Compared to *D. melanogaster*, *D. simulans* shows evidence of substantially more geographic expression differentiation in both tissues, and

especially in the testes, with almost seven times as many genes exhibiting differential expression (χ^2 test, P <0.001). Similarly to *D. melanogaster,* the proportion of genes differentially expressed between populations was significantly different between the AGs and the testes (Table 1). With respect to the differentially expressed orthologs across tissues and species, there is substantial variation in both the overall number of orthologs and the proportion of these orthologs expressed in both tissues. In both *D. simulans* and *D. hydei,* slight variation is observed when comparing the proportion of differentially expressed orthologs to the proportion of differentially expressed genes overall in the testes, a pattern of change that is not observed in the accessory glands of these species. It is worth noting that this pattern of change is not seen when comparing the proportions of these genes in *D. melanogaster* tissues. In *D. hydei*, 1464 (10.2%) and 408 (2.8%) genes were significantly differentially expressed between populations (False discovery rate (FDR) < 0.05) in the AGs and in the testes, respectively (Table 1, Fig 1). Compared to *D. melanogaster* and *D. simulans, D. hydei* exhibits the greatest number of both differentially expressed genes and of differentially expressed orthologs in the AGs. However, a smaller count of testes-biased genes are found to be differentially expressed in *D. hydei* relative to those in *D. simulans.*

The most striking observation in comparing 1:1 ortholog expression variation between Maine and Panama is the surprising number of orthologs expressed in parallel in the AGs of *D. simulans* and *D. hydei,* with 269 expressed orthologs that are clinal in both species ($p = 6.88\ e^{-19}$, Table 2). The next most significant result from these ortholog expression comparisons is the number of orthologs expressed in parallel in the testes of *D. simulans* and *D. hydei,* with 22 expressed orthologs that are also clinal in both species ($p = 7.57\ e^{-04}$, Table 2). The remaining

pairwise comparisons of ortholog expression variation do not yield as robust of a difference between observed and expected ortholog expression numbers in both tissues. Further analysis of the differentially expressed orthologs in all species pairwise comparisons shows that 250 out of 269 DE orthologs share directionality in expression in the AGs of *D. simulans* and *D. hydei* ($p = 5.39e^{-54}$, Table 3). Out of these 250 genes, 157 orthologs show shared higher expression in Maine and 93 orthologs share higher expression in Panama. To better understand the degree of relationship in these shared patterns of gene expression variation in the AGs of *D. simulans* and *D. hydei* across Maine and Panama, a regression analysis was done (Fig 2) and resulted in F-value of 283, an adjusted R^2 value of 0.51, and a p-value < $2.2\ e^{-16}$. Together, these data support that 1) the majority of genes showing parallel expression differentiation in the AGs of *D. simulans* and *D. hydei* show the same direction of DE in these species and that 2) the magnitudes of expression differences between high and low latitude populations are correlated across species. Lastly, we plotted log2 fold expression correlations between species for all genes showing expression differentiation in the AGs and testes, as seen in Fig S1, and validated our previous observation that the most significant numbers of parallel expression differentiation changes are observed in the accessory glands of *D. simulans* and *D. hydei.*

Gene Ontology

To better understand the functions of the 250 genes showing latitudinal DE in the AG and testes of both *D. simulans* and *D. hydei*, two gene ontologies were carried out and the results are as follows. With respect to the 157 genes that share higher expression in Maine, a 10 to 90 fold enrichment is seen in various processes, as seen in Fig 3, and some processes that

stand out include epithelial structure maintenance, somatic stem cell population maintenance, positive regulation of both insulin and hippo receptor signaling, and protein retention in the Golgi apparatus. With respect to the 93 genes that share higher expression in Panama, a 5 to 30 fold enrichment is seen in processes that include ribonucleotide complex assembly, oxidative phosphorylation, ribosomal large subunit biogenesis, proton transmembrane transport, proton motive force-driven ATP synthesis, and cytoplamsic translation, as seen in Fig 4.

We were also curious about the functions of 12 genes showing parallel latitudinal DE in the AGs in all three species. Five of these orthologs are more highly expressed in the accessory glands of males in Panama relative to those in Maine, and four of them encode ribosomal proteins RpS18, RpL11, RpLP2, and RpL30 while one of them encodes CG12384, which is predicted to enable death domain binding activity and has not been associated with accessory glands until now. The remaining seven genes are more highly expressed in the accessory glands of all three species in Maine relative to those in Panama: These genes are knockout (ko) which is predicted to enable DNA-binding transcription activator activity and is known to be associated with Drosophila sperm (Wasbrough et al. 2010), Trehalase (Treh) which encodes an enzyme that hydrolyzes trehalose to glucose molecules in the cytoplasm and extracellular space, previously identified as an SFP in the sperm proteome (Garlovsky et al. 2022), eIF4G1 which encodes a protein involved in regulation of translation initiation within male germ cells (Franklin-Dumont et al. 2007, Wasbrough et al. 2010), Gigyf (Gyf) which encodes a protein that regulates translation and insulin/IGF signaling and previously unreported in the accessory glands, alan shepard (shep) which encodes an evolutionary conserved RNA/DNA binding protein that regulates alternative splicing and gypsy insulator activities and also previously

unreported in its association with the male reproductive tract, PMCA which encodes an evolutionary conserved ion pump in the plasma membrane and is required for maintaining resting Ca[2+] levels in all cells of tunicates while also mediating chemotaxis in their sperm (Yoshida et al. 2018), and timeless (tim) which is required for the production of circadian rhythms and is also involved in mating behavior (Nishinokubi et al. 2006, Wasbrough et al. 2010). Lastly, it is important to note that there were 0 genes that showed parallel DE differentiation across all species in the testes.

Tissue-Biased Genes

When studying the 1:1:1 orthologs that are also tissue-biased, I found 232 orthologs that are accessory gland biased (Panel B, Fig 5) and 1024 orthologs that are testes-biased across all three species (Panel B, Fig 6). Our data support the closer relationship between *D. melanogaster* and *D. simulans* relative to *D. hydei*: Pairwise comparisons of tissue-biased orthologs shows 610 shared AG-biased orthologs (Fig S2) and 2033 shared testes-biased orthologs (Fig S3) in these two species, whereas significantly fewer genes are found to be shared between either of these species and *D. hydei.* Our finding of relatively similar ortholog numbers shared between the more closely related species and their distant relative *D. hydei* (Fig S2 and Fig S3) validates that we were able to capture the majority of orthologs across all three species in both tissues of interest.

Seminal Fluid Proteins Across Species

To date, Wigby et al. (2020) have characterized the most comprehensive suite of genes whose products are transferred to females during mating in *D. melanogaster*. These 292 genes are expressed in various male reproductive tissues of *D. melanogaster* and their gene products

are referred to as “high-confidence seminal fluid proteins” or HC-SFPs. In my study of *D. melanogaster* tissues, 278 HC-SFP genes were expressed in the accessory glands, 264 HC-SFP genes were expressed in the testes, and 263 HC-SFP genes were shared in expression between both tissues (Table 4). In *D. simulans,* 224 HC-SFP genes were expressed in the accessory glands, 203 HC-SFP genes were expressed in the testes, and 200 HC-SFP genes were shared in expression between both tissues. In *D. hydei,* 120 HC-SFP genes were expressed in the accessory glands, 121 HC-SFP genes were expressed in the testes, and 117 HC-SFP genes were shared in expression between both tissues.

With regards to the number of tissue-biased HC-SFPs in *D. melanogaster, D. simulans*, and *D. hydei*, respectively, 189 SFPs, 157 SFPs, and 57 SFPs are accessory gland-biased, 29 SFPs, 17 SFPs, and 23 SFPs are testes-biased, and 12 SFPs, 5 SFPs, and 16 SFPs are both accessory gland and testes-biased (Table S3). There are 41 shared AG-biased SFP genes and 6 shared testes-biased SFP genes across species (Fig 7). We also noticed that there are significantly greater AG-biased SFPs shared between *D. melanogaster* and *D. simulans* relative to those shared with *D. hydei*: 97 AG-biased SFPs are shared between *D. melanogaster* and *D. simulans,* 6 AG-biased SFPs are shared between *D. melanogaster* and *D. hydei,* and 4 AG-biased SFPs are shared between *D. simulans* and *D. hydei.*

Functionally Important SFPs Across Species

To better understand the relative abundance of functionally important SFPs produced by males in the AGs and testes, I chose to compare 11 functionally important SFPs across dissected tissues in all three species. These SFPs are Acp26Aa, Acp36DE, Sex Peptide (SP), and

eight characterized sex peptide network genes in *D. melanogaster:* CG17575, Lectin-46Ca, Lectin-46Cb, Sems, CG9997, Aqrs, Antr, and Intr.

In the AGs, 11/11 of these functionally important SFPs are highly expressed in *D. melanogaster* AGs in Maine and in Panama (Range of 192 to 13436 TPM, Table 5), 10/11 of these SFP genes are highly expressed in *D. simulans* AGs in Maine and in Panama (Range of 210 to 11901 TPM, Table 5), and 7/11 of these SFP genes are highly expressed in *D. hydei* AGs in Maine and in Panama (Range of 1310 to 10936 TPM, Table 5). In the testes, 11/11 of these functionally important SFPs are lowly to moderately expressed in *D. melanogaster* testes in Maine and in Panama (Range of 3 to 283 TPM, Table 5), only 1/11 (CG17575) of these SFP genes is lowly expressed in *D. simulans* testes in Maine and in Panama (Range of 1-3 TPM, Table 5), and 1/11(Sems) of these SFP genes is lowly expressed in *D. hydei* testes in Maine and in Panama (Range of 10-13 TPM, Table 5). In the testes of *D. simulans* and *D. hydei* across both regions, 10/11 genes show less than 1 TPM expression. Of course, this absence of expression cannot be confirmed unless validated by in-vivo knockdown of these genes in the testes of *D. simulans* and *D. hydei,* which was not carried out in this study. While these functionally important SFPs show a large spread in abundance in the accessory glands across Maine and Panama (Fig. 8), their relative position with respect to one another is mostly conserved between geographic populations within a species and even across all three species (Fig 9, Fig 10, Fig S4, Fig S5, Fig S6, Fig S7). Given that only one of these 11 selected genes is expressed over 1 TPM in the testes of *D. simulans* and *D. hydei* (Table 5), boxplot figures that depict the relative abundance of these 11 genes in the testes are omitted from this analysis, and the relative position of these functionally important SFP genes are not discussed in the testes of these two species.

DISCUSSION

We were motivated to carry out this study due to recent findings by Zhao et al. (2015) where the authors observed a significant excess of genes exhibiting parallel differential expression in whole males of *D. melanogaster* and *D. simulans* sampled from Panama and Maine. We therefore carried out this study by examining two reproductive tissues in males across populations and species to better understand their gene expression similarities and differences, while asking the question of whether any one of these organs is more predictable in their expression differentiation across species relative to the other organ.

Perhaps the most interesting finding from this study is that the majority of genes showing parallel expression differentiation showed the same direction of DE in two out of the three species studied and especially in their accessory glands, and that the magnitude of expression differences between high and low latitude populations were correlated across species. Together, these observations support our conclusion that parallelism for expression phenotypes is the result of spatially varying selection. The simplest explanation for the quantitative parallelism observed is that the relationship between transcript abundance and fitness variation is similar between these two species across variable environments. Our data suggest that selection on standing variation likely underlies geographic expression differences in *D. simulans* and in *D. hydei.* Despite major cytological differences between all three species studied here, their shared colonization histories and general population genetic similarities have likely influenced some of their shared evolutionary responses to recent colonization of novel habitats, which include the shared suite of genes independently expressed as well as the shared magnitude of expression differentiation across regions. It remains to be seen if the

observed expression differentiation is associated with spatially varying selection in the ancestral geographic ranges of these species.

Our gene ontology analyses have characterized a list of genes that respond similarly in the accessory glands across species, and especially between *D. simulans* and *D. hydei*. Interestingly, the four ribosomal proteins RpS18, RpL11, RpLP2, and RpL30 are more highly expressed in the AGs of males of all species in Panama, and additional work is required to elucidate the mechanisms by which they promote adaptability in high vs. low latitudes. Many DE gene products have previously been associated with animal reproduction and were found in our gene ontology analyses. These include knockout (ko) which has been previously associated with Drosophila sperm (Wasbrough et al. 2010), Trehalase (Treh) which encodes an enzyme that hydrolyzes trehalose to glucose molecules in the cytoplasm and extracellular space, previously identified as an SFP in the sperm proteome (Garlovsky et al. 2022), elF4G1 which encodes a protein involved in regulation of translation initiation within male germ cells (Franklin-Dumont et al. 2007, Wasbrough et al. 2010), PMCA which encodes an evolutionary conserved ion pump in the plasma membrane and is required for maintaining resting Ca[2+] levels in all cells of tunicates while also mediating chemotaxis in their sperm (Yoshida et al. 2018), and timeless (tim) which is required for the production of circadian rhythms and is also involved in mating behavior (Nishinokubi et al. 2006, Wasbrough et al. 2010). Some gene products were previously not associated with reproductive tissues and include CG12384, which is predicted to enable death domain binding activity, Gigyf (Gyf) which encodes a protein that regulates translation and insulin/IGF signaling and previously unreported in these tissues, and alan shepard (shep) which encodes an evolutionary conserved RNA/DNA binding protein that

regulates alternative splicing and gypsy insulator activities. Our modern ability to study gene knockdown in species outside of *D. melanogaster* will help investigators 1) Study the roles played by these reproductive molecules across males, 2) Ask whether such molecules are transferred to females during mating, and 3) Study the function of these molecules in females if transferred during mating. It is important to remember that many of these reproductive molecules are not necessarily secreted by these organs and transferred to females during mating, so future studies will hopefully elaborate on the spatial context of these molecules and their evolutionary histories.

While evidence for parallelism was observed between tissues across species, there are many differences between these species with respect to their differentially expressed genes. For instance, we find significant variation in the number of tissue-biased genes across species, and many of these tissue-biased genes are also species-biased, and this observation extends to the SFPs made by male reproductive tissues across species. So far, most studies on SFPs have focused on those made by male reproductive tissues in *D. melanogaster,* and future studies will shed some light on the nature of these SFPs and their variability across species tissues. Also, and while the differentiation in expression of transcription factors was not discussed in this analysis, some variation in transcription factor differentiation is observed across our pairwise tissue comparisons. Given the ability of transcription factors to regulate gene expression, geographic differences in the qualitative abundance of these factors would certainly have a role in shaping some of the gene expression plasticity in these tissues across populations and species.

Given our findings on the rapid and parallel evolution of accessory glands in some flies with shared colonization histories in the genus Drosophila, and given that recent studies have justified the use of Drosophila accessory glands as a model for the study of prostate cancer and other pathologies (Wilson et al. 2016), I cannot help but wonder if any parallel changes in gene expression have taken place in the prostate between human populations with shared migration histories. Some studies indicate that selection for hypoxia tolerance can act on standing variation in similar genes and pathways in both *D. melanogaster* and in humans (Jha et al. 2016), which have diverged hundreds of millions of years ago, so it is not inconceivable to hypothesize that some human populations have orthologs of reproductive genes and pathways in Drosophila that may also be under positive selection. Whether such shared gene expression changes have taken place between human populations or even across diverged species remains to be studied.

In conclusion, I have used natural variation across closely- and distantly-related Drosophila species to support that the accessory glands in males of *D. simulans* and *D. hydei* have independently reused the expression of 250 shared genes in response to their change in latitudinal environment, which is a striking difference that separates the gene expression patterns of the accessory glands from those in the testes. These genes expressed in parallel are under positive selection in these species and indicate that fundamental genetic mechanisms that respond to latitudinal migrations have remained somewhat conserved in some extant species. Surprisingly, results from the accessory glands specifically indicate that convergent evolution in behavior of this organ can occur between species roughly 50 million years diverged, with spatially varying selection contributing to the maintenance of gene expression

variation in these species. While this study has been able to narrow down on one male reproductive organ that seems to be somewhat predictable in its genetic response across some extant species, it remains to be seen if such parallel changes in gene expression across species can be similarly attributed to at least one reproductive organ in females, such as the paired spermathecae or parovaria. These are two paired glandular organs of the female reproductive tract that, among other functions, secrete RNA and protein products necessary for proper fertilization of the egg, and future studies will determine if these female organs or others behave in parallel across regions and species like we have seen in the accessory glands of males. The co-occurrence of *D. melanogaster, D. simulans,* and *D. hydei* and their recent colonization of novel geographic regions allows investigators to use these models to study the effects of sex-specific selection and local adaptation on phenotypic plasticity, to explain how these migration scenarios can maintain stable genetic variation for fitness, and to study the genotypic consequences such as allele linkage potentially associated with given habitats or sexes.

FIGURES AND TABLES

Species	Tissue	Mean fold change, all genes	Mean fold change, differentially expressed genes	Differentially expressed genes (FDR 0.05)	Differentially expressed one-to-one orthologs (FDR 0.05)	Differentially expressed tissue-biased genes	P-value
D. melanogaster	Accessory Glands	97.2%	272%	813 (4.5%)	476 (4.5%)	205 (13.9%)	5.82E-51
D. melanogaster	Testes	62.3%	645%	100 (0.5%)	48 (0.45%)	15 (0.44%)	0.89164078
D. simulans	Accessory Glands	87.3%	80%	1071 (6.8%)	856 (8.1%)	256 (19.1%)	4.37E-56
D. simulans	Testes	40.1%	129%	684 (4.3%)	479 (4.5%)	284 (6.8%)	7.47E-18
D. hydei	Accessory Glands	68.6%	63%	1464 (10.2%)	1269 (11.9%)	361 (33.2%)	4.18E-105
D. hydei	Testes	31.3%	150%	408 (2.8%)	295 (2.8%)	111 (3.2%)	0.07202612

Table 1. Panama vs. Maine gene differential expression. Numbers of expressed genes were 13904 in the accessory glands and 15405 in the testes of *D. melanogaster*, 13627 in the accessory glands and 14502 in the testes of in *D. simulans*, and 12418 in the accessory glands and 13246 in the testes of *D. hydei*. 10628 expressed genes were one-to-one orthologs between species. Fold change corresponds to mean absolute fold change. Tissue biased genes were characterized by selection of those genes with a tau value of 0.9 or greater and if they were most highly expressed in these tissues of interest, and these lists were compared with similar lists generated from the FlyAtlas data. P-values derive from hypergeometric test for overrepresentation of tissue-biased genes.

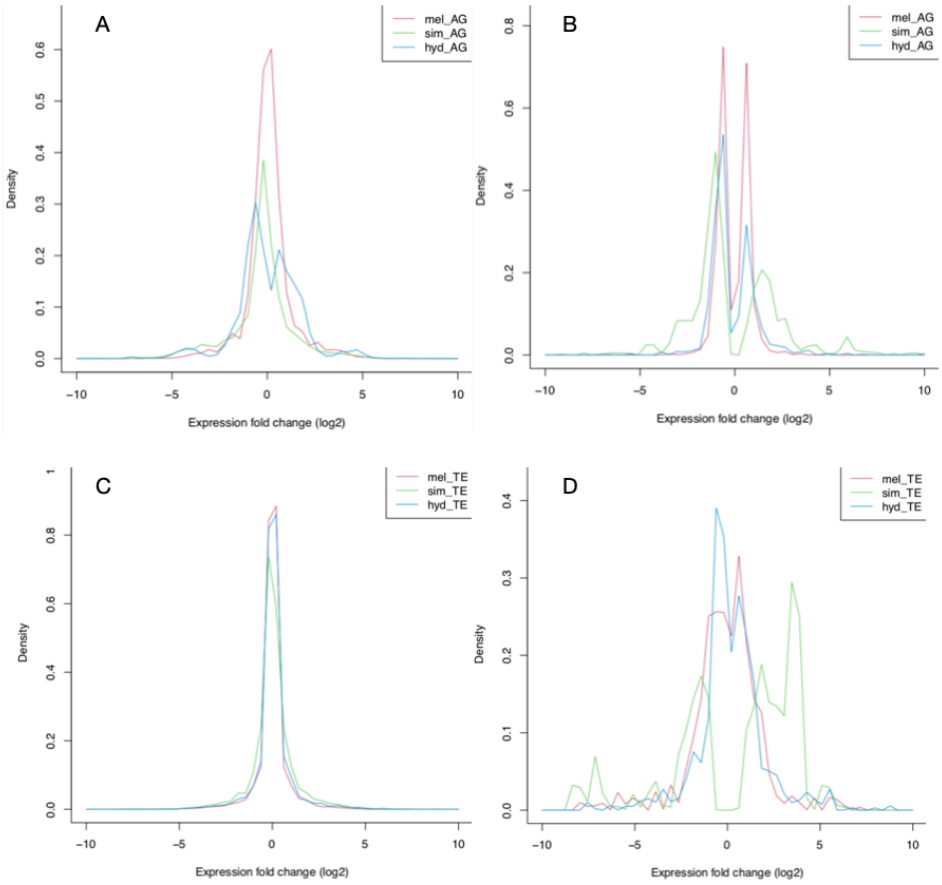

Figure 1. Expression fold changes for each comparison in *D. melanogaster, D. simulans,* and *D. hydei.* A) Fold changes (log2) for all genes expressed in the accessory glands in Maine vs. Panama. B) Fold changes (log2) for all differentially expressed genes in the accessory glands in Maine vs. Panama. C) Fold changes (log2) for all genes expressed in the testes in Maine vs. Panama. D) Fold changes (log2) for all differentially expressed genes in the testes in Maine vs. Panama.

Tissue	Species 1	Species 2	1 to 1 Orthologs Expressed in Both	Clinal in Species 1	Clinal in Species 2	Clinal in Both Observed	Clinal in Both Expected	Pbinom (lower.tail=FALSE)
AG	Mel	Sim	6726	433	868	68	55.9	4.37E-02
AG	Mel	Hyd	6234	277	1281	57	56.9	4.61E-01
AG	Sim	Hyd	5749	771	1179	269	158.1	6.88E-19
Testis	Mel	Sim	9504	45	460	4	2.2	6.97E-02
Testis	Mel	Hyd	8974	40	264	3	1.2	3.14E-02
Testis	Sim	Hyd	8659	397	240	22	11.0	7.67E-04

Table 2. Panama vs. Maine parallel gene expression in the accessory glands and testes of sexually mature males in *D. melanogaster*, *D. simulans*, and *D. hydei*

Tissue	Species 1	Species 2	1 to 1 Orthologs Clinal in Both	Shared Directionality	Pbinom (lower.tail=FALSE)	Shared Expression higher in Maine	Shared Expression higher in Panama
AG	Mel	Sim	68	46	1.09E-03	34	12
AG	Mel	Hyd	57	38	3.75E-03	29	9
AG	Sim	Hyd	269	250	5.39E-54	157	93
Testis	Mel	Sim	4	0	NA	0	0
Testis	Mel	Hyd	3	2	1.25E-01	0	2
Testis	Sim	Hyd	22	11	4.16E-01	6	5

Table 3. Shared directionality of Orthologs expression in the accessory glands and testes of sexually mature males in *D. melanogaster*, *D. simulans*, and *D. hydei*

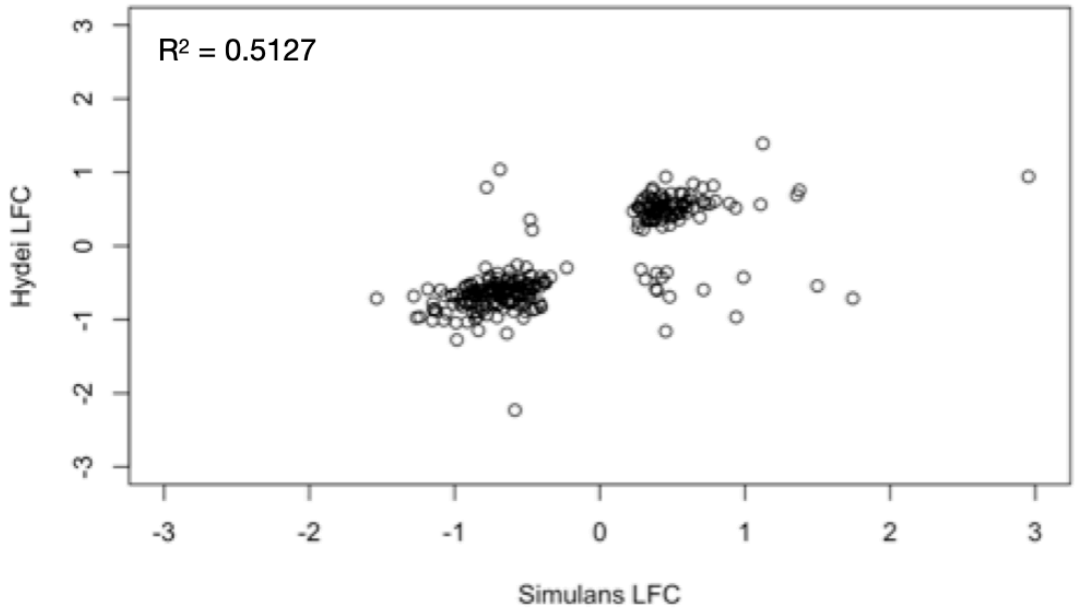

Figure 2. Linear regression analysis of log fold changes calculated from DESeq2 analyses of accessory glands in *D. simulans* and *D. hydei*. Dots represent log transformed fold change values of differential gene expression in *D. simulans* (X-axis) and *D. hydei* (Y-axis) across Maine and Panama. R^2: adjusted R-squared. p-value < $2.2e^{-16}$ and F- statistic = 283.

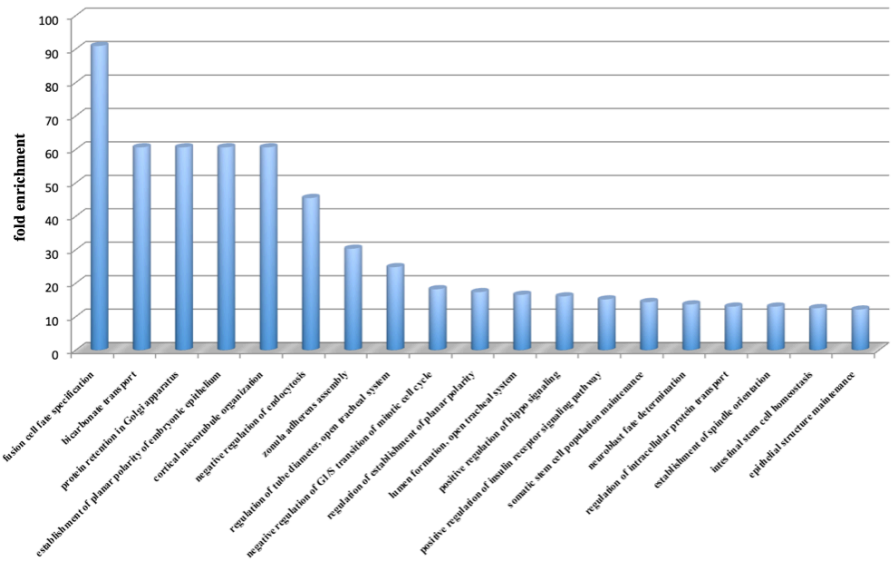

Figure 3. Enrichment of GO annotation terms for orthologous gene in *D. simulans* and *D. hydei* where shared expression is higher in Maine. Bars indicate fold enrichment of the respective GO terms in the phenotypic classes.

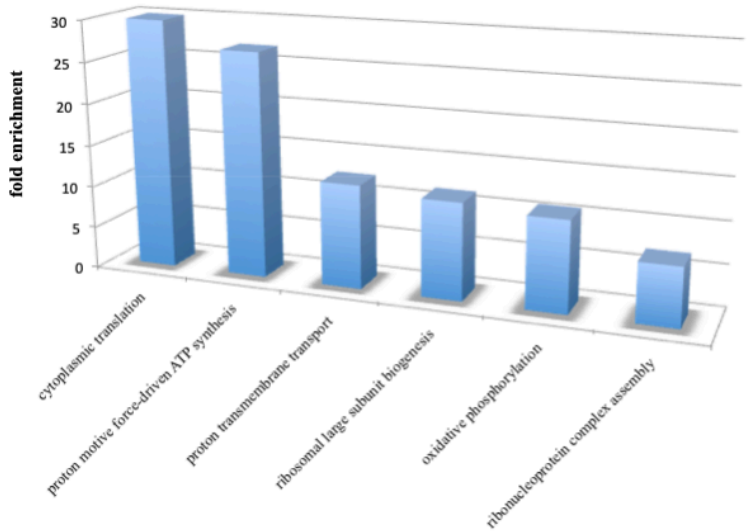

Figure 4. Enrichment of GO annotation terms for orthologous gene in *D. simulans* and *D. hydei* where shared expression is higher in Panama. Bars indicate fold enrichment of the respective GO terms in the phenotypic classes.

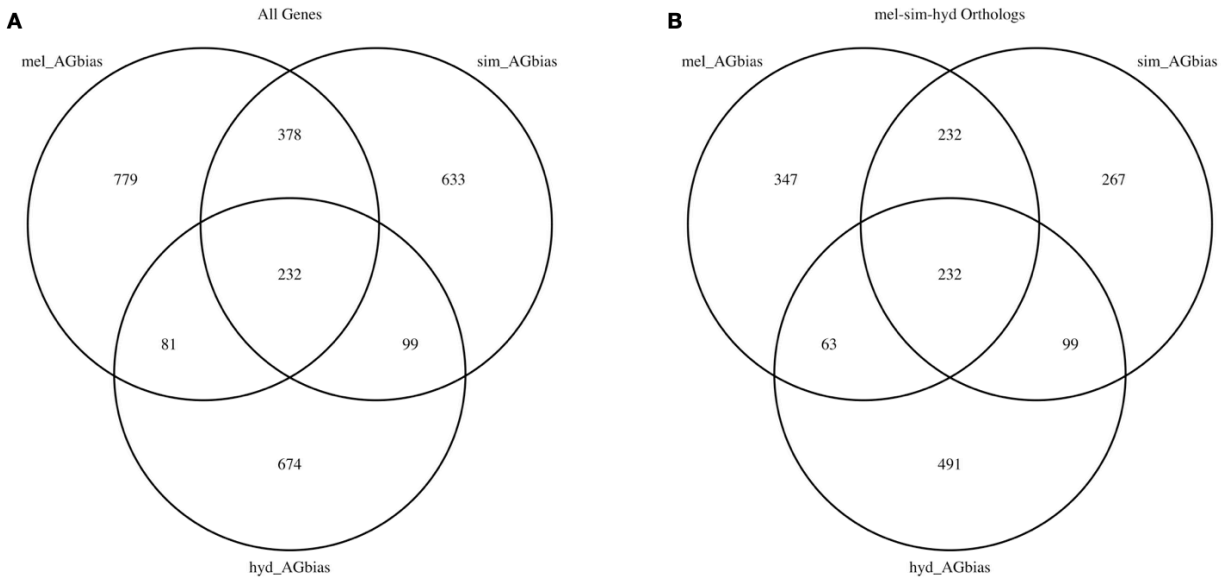

Figure 5. Venn diagrams depicting all shared accessory gland biased genes (Panel A) and shared accessory gland biased orthologs (Panel B) in *D. melanogaster, D. simulans,* and *D. hydei.*

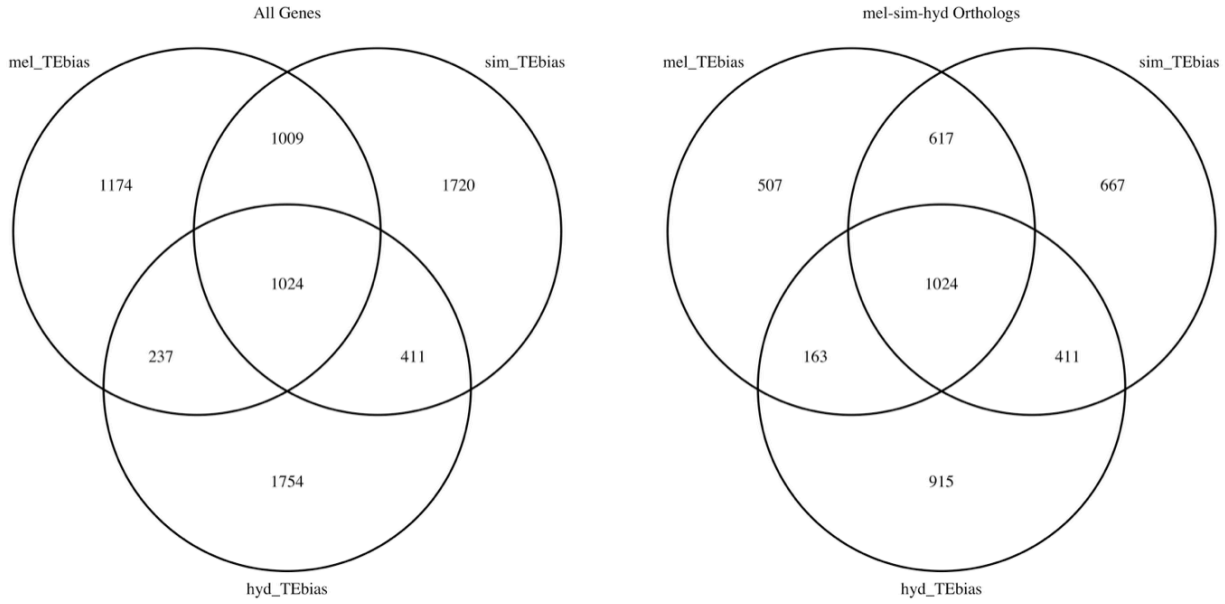

Figure 6. Venn diagrams depicting all shared testes biased genes (Panel A) and shared testes biased orthologs (Panel B) in *D. melanogaster, D. simulans,* and *D. hydei.*

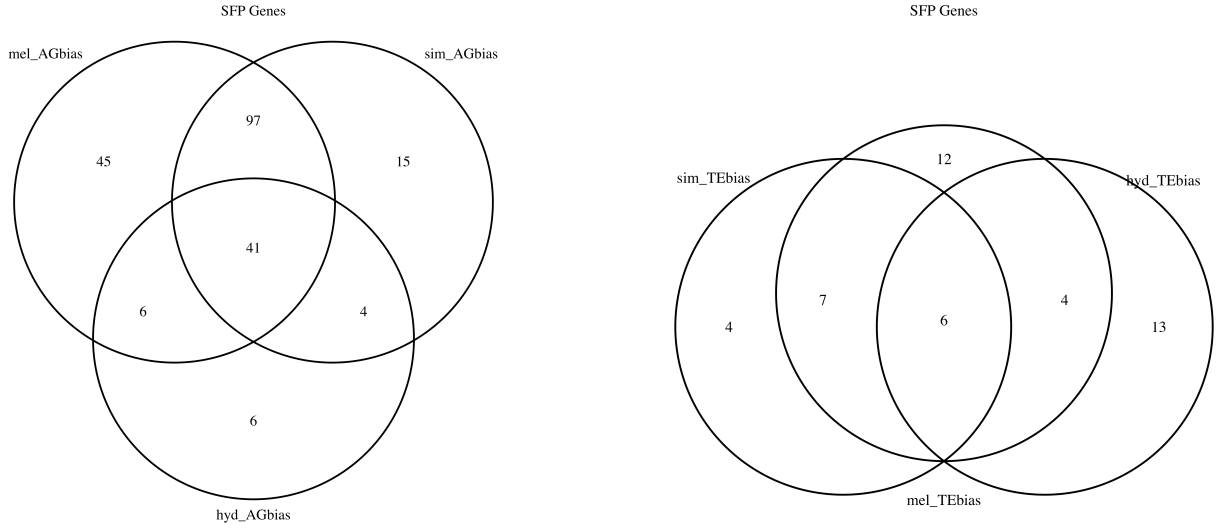

Figure 7. Venn diagram showing the number of high confidence SFP genes that are also A) accessory gland biased and B) Testes-biased in expression in *D. melanogaster, D. simulans,* and *D. hydei*

Species	HC-SFP genes expressed in AGs	HC-SFP genes expressed in Testes	HC-SFP genes expressed in AGs & Testes
D. melanogaster	278	264	263
D. simulans	224	203	200
D. hydei	120	121	117

Table 4. High confidence SFPs expressed in the accessory glands, testes, or in both tissues in *D. melanogaster, D. simulans*, and *D. hydei.*

Gene.Name	mel-AG-M	mel-AG-P	mel-Te-M	mel-Te-P	sim-AG-M	sim-AG-P	sim-Te-M	sim-Te-P	hyd-AG-M	hyd-AG-P	hyd-Te-M	hyd-Te-P
CG17575	3154.1897	4656.52555	21.4187333	77.0560703	3289.18815	3394.09863	2.067931	1.675954	8295.49121	10936.9674	0.59378133	0.521018
lectin-46Ca	2637.07902	3874.77865	13.3822923	53.2977117	784.886841	856.383341	0.000116	0	9750.78906	7823.15462	0.82815333	0.78109
lectin-46Cb	2329.53044	3492.47607	13.6282433	49.259593	582.983378	701.801412	6.6667E-07	0	7412.75602	7308.87142	0.768678	0.30231833
CG9997	969.810648	793.205221	2.72586467	23.3673163	1485.01115	1321.7159	0.319514	0.693256	1792.05497	1310.33293	0.42637467	0.14409
aqrs	1212.96741	1303.15165	3.837039	38.379221	1894.81083	1670.83793	0.21659133	0.30676633	3872.30111	2962.63118	0.857237	1.142705
antr	631.606364	843.151164	3.61375267	17.7643187	730.555542	706.928731	0.37703933	0.18928833	2366.76904	2581.96598	0.82635733	1.61789967
intrepid	192.010305	211.663234	0.454753	3.03166867	289.94456	210.007416	0.296749	0.34263033	0	0	0	0
Sems	665.093974	906.181946	3.370887	26.3561323	0	0	0	0	6725.59147	6860.85986	10.8985143	12.8335057
sex peptide	6052.53744	5806.68099	17.1378373	187.592553	4584.83911	3784.77987	0.374112	0.42492333	0	0	0	0
Acp36DE	13436.4089	6531.47689	23.5112137	283.228902	11901.0977	9214.90169	0.012697	0.002437	0	0	0	0
Acp26Aa	3440.4235	1662.08101	3.131304	29.8429457	7172.92253	6285.69694	0.664667	0.29455767	0	0	0	0

Table 5. Differential expression of selected functionally important SFPs.

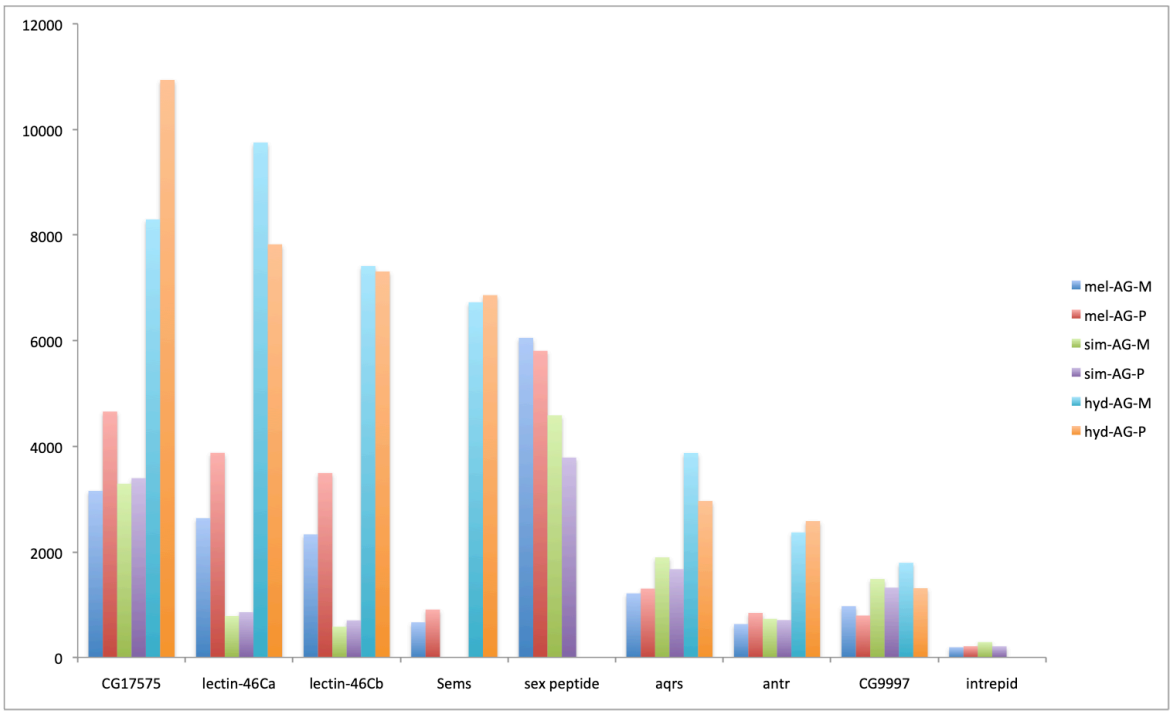

Figure 8. Expression of 9 functionally important SFPs in the accessory glands across all three species studied. Sems shows no expression in the accessory glands of *D. simulans*

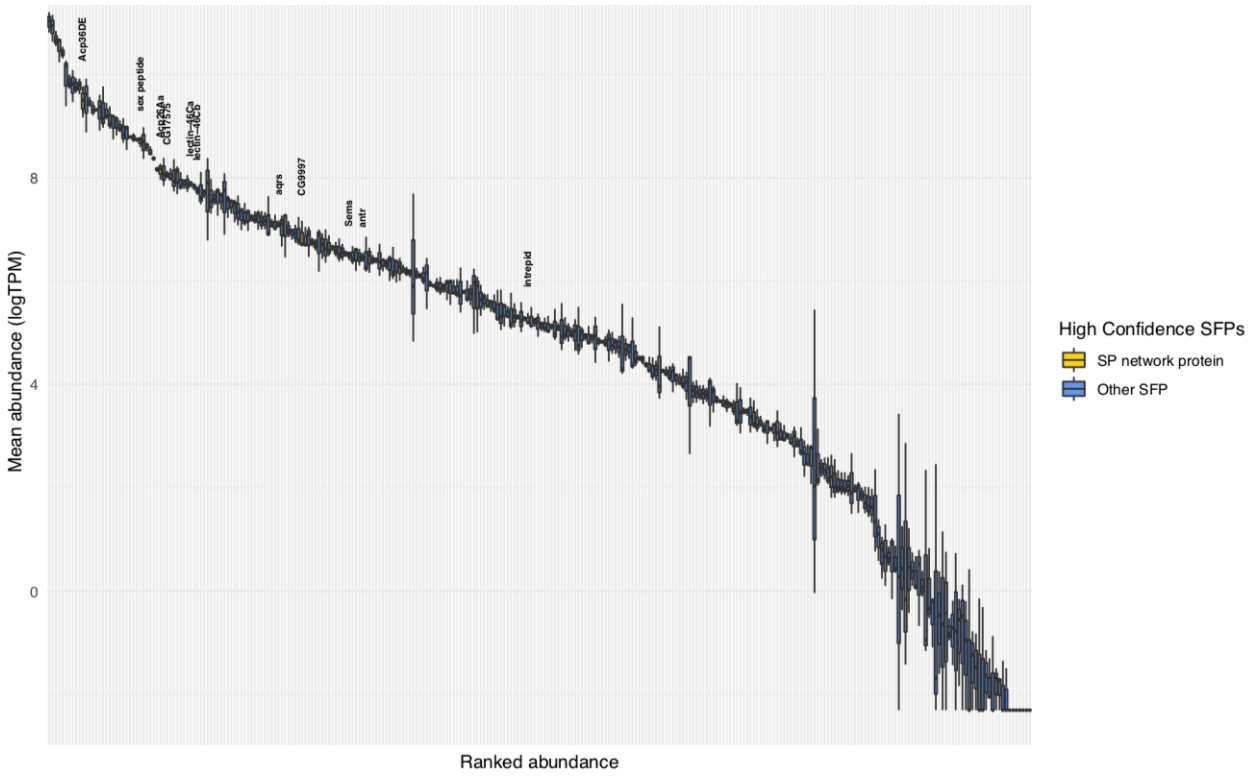

Figure 9. Boxplot of the relative abundance of functionally important SFPs and the rest of the known SFPs found in *D. melanogaster* accessory glands in Maine. Transcript abundances were averaged across all three biological replicates and were sorted by decreasing order. Selected functionally important SFPs are colored in yellow and the rest of the known SFPs are colored in blue.

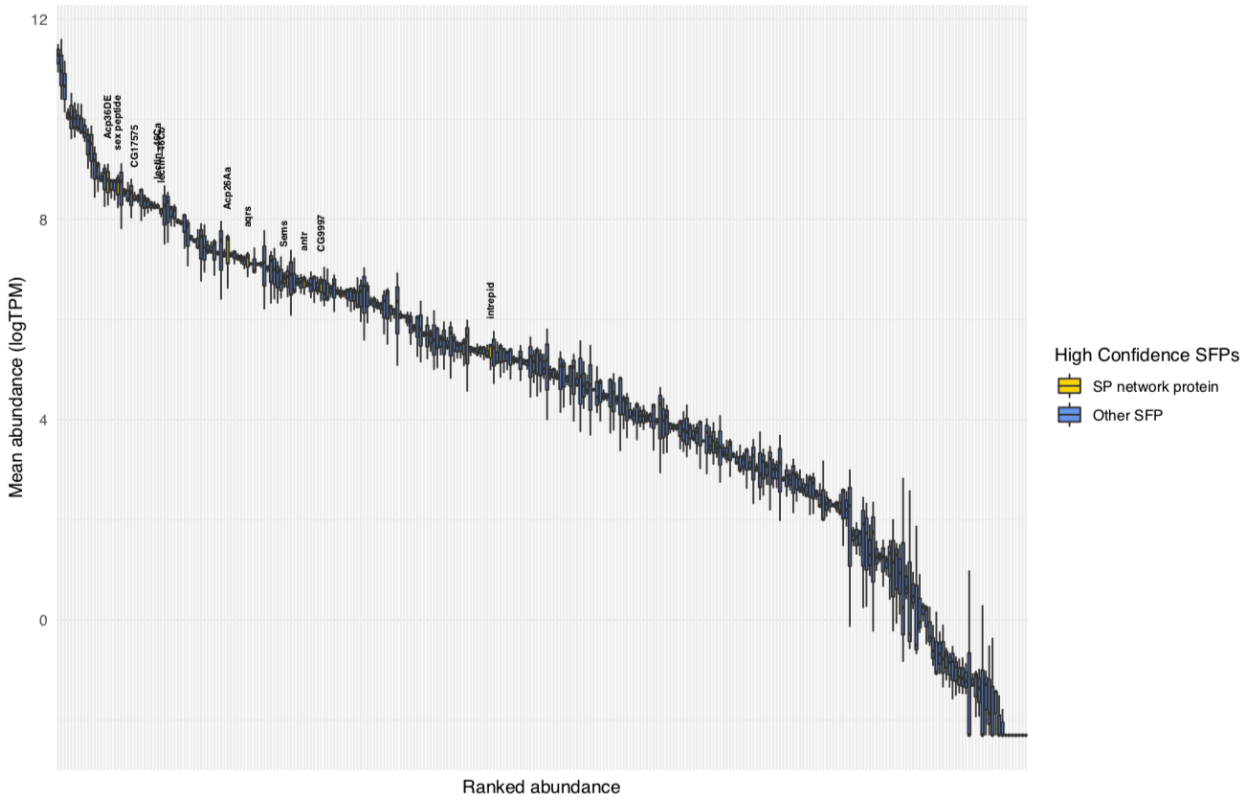

Figure 10. Boxplot of the relative abundance of functionally important SFPs and the rest of the known SFPs found in *D. melanogaster* accessory glands in Panama. Transcript abundances were averaged across all three biological replicates and were sorted by decreasing order. Selected functionally important SFPs are colored in yellow and the rest of the known SFPs are colored in blue.

REFERENCES

Almeida FC, Rob DeSalle. (2008). Evidence of Adaptive Evolution of Accessory Gland Proteins in Closely Related Species of the *Drosophila repleta* Group, *Molecular Biology and Evolution*, Volume 25, Issue 9, Pages 2043–2053.

Castoe T.A. (2009). Evidence for an ancient adaptive episode of convergent molecular evolution. *Proc. Natl. Acad. Sci.* 106, 8986 LP – 8991.

Cridland JM, Majane AC, Sheehy HK, Begun DJ. (2020). Polymorphism and Divergence of Novel Gene Expression Patterns in *Drosophila melanogaster*. Genetics. 216(1): 79-93.

David J, Capy P. (1988). Genetic variation of Drosophila melanogaster natural populations. Trends Genet.; 4:106–111.

Eden, E, R Navon, I Steinfeld, D Lipson, and Z Yakhini. (2009). GOrilla: A Tool For Discovery And Visualization of Enriched GO Terms in Ranked Gene Lists. BMC Bioinformatics 10:48.

Eden, E, D Lipson, S Yogev, and Z Yakhini. (2007). Discovering Motifs in Ranked Lists of DNA sequences. PLoS Computational Biology, 3(3):e39.

Fabian DK, Kapun M, Nolte V, Kofler R, Schmidt PS, Schlötterer C, et al. (2012). Genome-wide patterns of latitudinal differentiation among populations of Drosophila melanogaster from North America. Mol Ecol.; 21: 4748–4769.

Tina M. Franklin-Dumont, Chandrima Chatterjee, Steven A. Wasserman, Stephen DiNardo. (2007). A novel eIF4G homolog, Off-schedule, couples translational control to meiosis and differentiation in *Drosophila* spermatocytes. *Development* 1; 134 (15): 2851–2861.

J. R. Gallant, *et al.* (2014). Genomic basis for the convergent evolution of electric organs. *Sci.* 344, 1522–1525.

Garlovsky MD, Sandler JA, Karr TL. (2022). Functional Diversity and Evolution of the Drosophila Sperm Proteome. Mol Cell Proteomics. 21(10):100281

Jha AR, Zhou D, Brown CD, Kreitman M, Haddad GG, White KP. (2016). Shared Genetic Signals of Hypoxia Adaptation in Drosophila and in High-Altitude Human Populations. Mol Biol Evol. 33(2):501-17.

Kim, D, B Langmead, and S L Salzberg. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* 12: 357–360.

Leader, D. P., S. A. Krause, A. Pandit, S. A. Davies, and J. A. T. Dow. (2018). FlyAtlas 2: a new version of the Drosophila melanogaster expression atlas with RNA-Seq, miRNA-Seq and sex-specific data. Nucleic Acids Res. 46: D809–D815.

Love MI, Huber W, Anders S. (2014). Moderated estimation of fold change and dispersion for RNA-Seq data with DESeq2. Genome Biol. 15:550.

Nishinokubi I, Shimoda M, Ishida N. (2006). Mating rhythms of Drosophila: rescue of tim01 mutants by D. ananassae timeless. J Circadian Rhythms. 8; 4:4.

Pankey, M.S., V. N. Minin, G. C. Imholte, M. A. Suchard, T. H. Oakley. (2014). Predictable transcriptome evolution in the convergent and complex bioluminescent organs of squid. *Proc. Natl. Acad. Sci.* 111, E4736 LP-E4742.

Pertea, M., D. Kim, G. M. Pertea, J. T. Leek, and S. L. Salzberg. (2016). Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. Nat. Protoc. 11: 1650–1667.

A. R. Pfenning, *et al.* (2014). Convergent transcriptional specializations in the brains of humans and song-learning birds. *Science (80).* 346, 1256846.

Reinhardt JA, Kolaczkowski B, Jones CD, Begun DJ, Kern AD. (2014). Parallel geographic variation in Dro- sophila melanogaster. 197: 361–373.

A. Rokas, S. B. Carroll. (2008). Frequent and Widespread Parallel Evolution of Protein Sequences. *Mol. Biol. Evol.* 25, 1943–1953.

Singh RS, Long AD. (1992). Geographic variation in Drosophila: From molecules to morphology and back. Trends Ecol Evol. 7: 340–345.

Van Dongen, S., and C. Abreu-Goodger. (2012). Using MCL to extract clusters from networks. Methods Mol. Biol. 804: 281–295.

Von Heijne, G. (1990). The signal peptide. J. Membr. Biol. 115: 195–201.

Wagstaff BJ, Begun DJ. (2005). Molecular population genetics of accessory gland protein genes and testis-expressed genes in Drosophila mojavensis and D. arizonae. Genetics. 171(3):1083-101.

Wagstaff BJ, David J Begun. (2007). Adaptive Evolution of Recently Duplicated Accessory Gland Protein Genes in Desert Drosophila, *Genetics*, Volume 177, Issue 2, Pages 1023–1030.

Wasbrough, E.R., Steve Dorus, Svenja Hester, Julie Howard-Murkin, Kathryn Lilley, Elaine Wilkin, Ashoka Polpitiya, Konstantinos Petritis, Timothy L. Karr. (2010). The Drosophila melanogaster sperm proteome-II (DmSP-II), Journal of Proteomics, Volume 73, Issue 11.

Wilson C, Leiblich A, Goberdhan DC, Hamdy F. (2017). The Drosophila Accessory Gland as a Model for Prostate Cancer and Other Pathologies. Curr Top Dev Biol. 121:339-375.

Witt K.E., E. Huerta-Sánchez. (2019). Convergent evolution in human and domesticate adaptation to high-altitude environments. *Philos. Trans. R. Soc. B Biol. Sci.* 374, 20180235.

Yanai I, Benjamin H, Shmoish M, Chalifa-Caspi V, Shklar M, Ophir R, et al. (2005). Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. Bioinformatics. 21: 650–659.

R. L. Young, *et al.* (2019). Conserved transcriptomic profiles underpin monogamy across vertebrates. *Proc. Natl. Acad. Sci.* 116, 1331 LP – 1336.

Yoshida, K., Shiba, K., Sakamoto, A. *et al.* (2018). Ca^{2+} efflux via plasma membrane Ca^{2+}-ATPase mediates chemotaxis in ascidian sperm. *Sci Rep* 8, 16622.

Zhao L, Saelao P, Jones CD, Begun DJ. (2014). Origin and spread of de novo genes in Drosophila melanogaster populations. Science. 343: 769–772.

Zhao L, Wit J, Svetec N, Begun DJ. (2015). Parallel Gene Expression Differences between Low and High Latitude Populations of *Drosophila melanogaster* and *D*. *simulans*. PLoS Genet 11(5).

Zhao L, Begun DJ. (2017). Genomics of parallel adaptation at two timescales in *Drosophila*. PLoS Genet 13(10).

Zhen Y., M. L. Aardema, E. M. Medina, S. Schumer, P. Andolfatto. (2012). Parallel Molecular Evolution in an Herbivore Community. *Science (80).* 337, 1634–1637.

Chapter 2

Evolution of Female Sperm Storage Organs

in Flies of the Genus *Drosophila*

INTRODUCTION

Can an organ compensate for the loss of a specific function in a neighboring one? It has long been known that changes in the developmental regulation of evolutionarily conserved genes are a major force in phenotypic evolution, supporting that the origin of new traits does not necessarily require the evolution of new genes (King and Wilson 1975, Carroll 2005, Davidson 2006). These studies suggest that the origin of a novel trait such as sperm storage might not require the evolution of a new organ, but may instead rely on the modification of an existing organ to satisfy this function. The goal of this research is to test this hypothesis.

The spermathecae, parovaria, and seminal receptacle of *Drosophila* females are excellent models for the study of gain and loss of organ functions. Ancestrally, both spermathecae and seminal receptacle stored sperm in the female (Pitnick et al. 1999). In females of *Drosophila melanogaster*, the spermathecae and parovaria are major exocrine glands that are required for normal fertility and sperm storage (Sun et al. 2013), yet 34% of *Drosophila* species show a) loss of sperm storage function and degeneration of tissues in the spermathecae, and b) serious morphological divergence in their seminal receptacle, which behaves as the primary sperm-storage organ in these species (Pitnick et al. 1999). These observations suggest that neighboring organs in the female's reproductive tract, the parovaria and seminal receptacle, might have potentially acquired glandular functions previously restricted to the spermathecae.

A comparative investigation of these organ-specific transcriptomes could point to the evolution of some level of parallel gene expression changes that underlie gain of sperm storage

functions. If the transcriptomes underlying independent evolution of sperm storage are not themselves convergent, then I will identify the various genomic mechanisms of expression gain and loss by organ and by species, which will help elucidate the relative contributions of gene origin and regulatory circuit rewiring to the evolution of organ complexity. These genomic mechanisms include gene duplication, gene co-option, and de novo gene origin. In this study, I take a comparative transcriptomics approach to ask whether there is substantial overlap in gene expression that is potentially associated with the novel ability of the parovaria to store sperm in some species, and I also investigate gene expression differences that are potentially associated with the loss of sperm storage in the spermathecae of other species.

MATERIALS AND METHODS

Experimental Design

Strong phylogenetic support exists for the independent gain of sperm storage function in the parovaria of *D. nigricruria* and *D. takahashii,* and for the independent loss of sperm storage function in the spermathecae of *D. bifurca* and *D. serrata* (Pitnick et al. 1999). My dissections as well as those of my colleague (CE McDonough-Goldstein, personal communication) confirmed the presence of sperm in the parovaria of *D. takahashii* and *D. nigricruria* females after mating, and my dissections have also confirmed the absence of sperm in the spermathecae of *D. serrata* and *D. bifurca* females after mating. These observations were made at several timepoints, including immediately after and hours after mating.

To better understand the independent gain of sperm storage in the parovaria of *D. takahashii* and *D. nigricruria,* I dissected the seminal receptacles, spermathecae, and

parovaria from these two species as well as those in *D. melanogaster, D. greeni, D. repleta,* and *D. hydei,* with the goal of comparing the transcriptome outputs from these tissues across species. Because the overall pattern of sperm storage across Drosophila examined supports that females in ancestral Drosophilid species stored sperm in their seminal receptacle and spermathecae (Pitnick et al. 1999), I chose to compare the derived tissue transcriptomes of *D. takahashii and D. nigricruria* to those of Drosophila species that only store sperm in the seminal receptacle and spermathecae, as seen in Fig.1. This study will allow me to examine the independent evolution of sperm storage ability in two distinct groups of Drosophila relatives.

To better understand the independent loss of sperm storage in the spermathecae of *D. serrata* and *D. bifurca,* I dissected the seminal receptacle, spermathecae, and parovaria from these two species as well as those in *D. melanogaster, D. greeni, D. repleta,* and *D. hydei,* with the goal of comparing the transcriptome outputs from these tissues across species. Here, I chose to compare the derived tissue transcriptomes of *D. serrata* and *D. bifurca* to those that store sperm in the seminal receptacle and spermathecae, as seen in Fig. 2.

Genomic Sequencing

A high quality genome assembly (Adams et al. 2000, Dros. Consort. 2007) exists for *D. melanogaster,* which is a species that retains sperm storage function in both the spermathecae and seminal receptacle (Pitnick et al. 1999). The relatively compact size of *Drosophila* genomes (Dros. Consort. 2007, Bosco et al. 2007, Song et al. 2011) has allowed for the rapid sequencing of 27 fly genomes in the past fifteen years. I will therefore rely on well-established methods in the field, including those recently used by the Begun lab (Song et al. 2011) to sequence the

genomes of *D. bifurca, D. nigricruria,* and *D. greeni*. Together with previous sequences for the remaining species (Adams et al. 2000, Dros. Consort. 2007), I will have access to complete genomes of eight Drosophila species for my studies.

RNA extraction and Sequencing

Virgin females were collected within 3 hours of eclosion and raised at 25°C in low density vials (up to 5 females per vial) until sexual maturity. All tissue dissections included 100-200 seminal receptacles, 100-200 pairs of spermathecae, and 100-200 pairs of parovaria. Dissected tissues were immediately transferred into Trizol and stored at -80°C before RNA was extracted following a standard Trizol-Chloroform extraction protocol. Three replicates were generated for each tissue, for a total of 72 biological samples across eight species. Poly(A)+ RNA was prepared from an aliquot of each total RNA sample. Due to the limited availability of RNA for some of the samples and due to the variation in RNA integrity between some of these samples, I decided to profile gene expression using the 3' TAG-Seq protocol. For typical experiments, at least 2 million reads per sample are required for DGE analysis of highly- and moderately-expressed genes. Three biological replicates were sequenced for each tissue to allow for meaningful DGE analysis. Barcoded sequencing libraries were prepared each from 1.3 ng total RNA sample using the QuantSeq FWD kit (Lexogen, Vienna, Austria) for multiplexed sequencing according to the recommendations of the manufacturer using the UMI Second-Strand Synthesis module (Lexogen). The fragment size distribution of the libraries was verified via micro-capillary gel electrophoresis on a LabChip GX system (PerkinElmer, Waltham, MA).

The libraries were quantified by fluorometry on a Qubit fluorometer (LifeTechnologies, Carlsbad, CA), and pooled in equimolar ratios. The library pool was Exonuclease VII (NEB, Ipswich, MA) treated, SPRI-bead purified with KapaPure beads (Kapa Biosystems / Roche, Basel, Switzerland), and quantified via qPCR with a Kapa Library Quant kit (Kapa Biosystems) on a QuantStudio 5 RT-PCR system (Applied Biosystems, Foster City, CA). Libraries were sequenced on a NextSeq 500 sequencer (Illumina, San Diego, CA) with single-end 80 bp reads. Expression of each gene in each organ and species will be quantified in FPKM. I will use 1 FPKM as my detection threshold, since current research (Hebenstreit et al. 2011, Wagner et al. 2013) supports that gene expression levels below this threshold most likely correspond to non-functional transcription. This threshold can be modified based on the expression data generated, and several indices of transcriptome similarity across organs and species will be estimated.

Characterizing gain and loss of gene expression

In each organ, I am interested in identifying the genes that are expressed in some species but not in others. For each expressed/unexpressed gene, I will determine the branch of the *Drosophila* phylogeny where its expression in the parovaria, spermathecae, or seminal receptacle was gained or lost in a fully hierarchical Bayesian approach implemented in BEAST (Drummond et al. 2012). This will allow me to identify those genes for which the gain or loss of expression can be mapped with greater than 0.95 posterior probability to a specific branch of the phylogeny despite missing data, if any. Using these inferences, I will determine if there are consistent patterns of expression gain or loss across organs and species with respect to number

and types of genes/proteins. These data, in concert with transcriptome and genome assemblies, will be used to identify de novo genes that may evolve in the context of novel organ function.

Evolution of transcriptomes

Here, I will take a broader quantitative analysis of gene expression evolution. I will determine whether organs that have recently changed sperm storage functions show rapid transcriptome divergence using FPKM estimates of 1-to-1 orthologs. The relative importance of up-regulation and down–regulation will be investigated, and the particular gene functions associated with the largest expression changes will be determined with gene ontologies. I will also use reliable paralog data to determine whether organ function evolution is associated with novel gene duplications, perhaps associated with neofunctionalization. Using FPKM estimates, I will also determine whether parallel changes of organ function are associated with evidence of parallelism at the transcriptome level.

RESULTS

Tables 1 and 2 present the total number of 3’ TAG seq reads for the 72 biological replicates, ranging between 4,000,000 and 7,000,000 reads per replicate. Tables 3 and 4 present the RNA concentrations for the various biological replicates.

DISCUSSION

Decades of studies in Drosophila reproductive biology have characterized the various transcriptomic components of the testes and associated male tissues, and especially those made by the Drosophila organ equivalent to the human prostate: the accessory glands. However, studies of these reproductive transcriptomic components have remained more limited in females, in part due to the difficulty of accessing some of these tissues, and in part due to the limited amounts of RNA that could historically be sequenced from some of these tissues, and especially the spermathecae and parovaria. To begin to achieve a similar level of understanding of the molecular and biological roles carried out by the collection of RNAs in female seminal fluids, I utilized transcriptomics to study the composition of three female reproductive tissues dissected from females across 8 Drosophila species.

Why study sperm storage in females of Drosophila?

Little remains known with respect to the roles played by the diverse reproductive organs and tissues outside of ovaries in females. In the few and far away studies that describe the molecular components of these organs in Drosophila, the seminal receptacle, the spermathecae, and the parovaria are female reproductive organs known to be important sites for the secretion of products associated with future and successful fertilization. However, and with respect to sperm storage molecules made by females specifically, only glucose dehydrogenase (Gld) has been demonstrated so far to have an active role in recruiting sperm to the spermathecae (Schiff et al. 1992, Lida and Cavener 2004). While little is known about the

active role played by females in sperm storage, much more is known about the tissue of origins, protein classes, and degree of involvement of male-transferred proteins in this process. These proteins include glucose dehydrogenase (Gld), which in addition to being produced by the spermathecae, is made by the male's ejaculatory duct and ejaculatory bulb (Cavener 1983, Schiff et al. 1992, Lida and Cavener 2004, McGraw et al. 2004, Ram & Wolfner 2009). Another protein made in the ejaculatory duct and responsible for both sperm storage and sperm release from storage is Esterase-6 (Gilbert 1981, Gilbert and Richmond 1982). Otherwise, there has been a bias in the literature towards the characterization of accessory gland proteins involved in female sperm storage, which include but are not limited to the prohormone Acp36DE (Avila & Wolfner 2009, Neubaum & Wolfner 1999, Bloch & Wolfner 2003, Chapman et al. 2000), the lectins Acp29AB (Wong et al. 2008), CG1652 and CG1656 (Ravi & Wolfner 2007, Ram & Wolfner 2009), the serine proteases Seminase (LaFlamme et al. 2012) and CG9997 (Ravi & Wolfner 2007, Ram & Wolfner 2009), and the metalloprotease CG11864 (LaFlamme et al. 2012, Ravi Ram et al. 2006). Although only one female-produced product (Gld) has been characterized to be involved in sperm storage, ablation of the spermathecal secretory cells results in a decrease of sperm recruitment into this organ and in the rapid loss of sperm motility in the seminal receptacle, implying the ability of certain products to act at a distance from their production site in the female reproductive tract (Schnakenberg et al. 2011). In addition to such observations, the impressive variation in physiology of these female reproductive tissues across species and especially in the volume of the cuboidal secretory cells that make up the spermathecae and parovaria, as well as the variation in the sperm storage abilities of these three tissues (Pitnick et al. 1999) suggest that these organs vary in molecular and biological

functions across species, which motivates us to study the transcriptome bases of these varied sperm storage phenotypes across tissues and species.

In addition to better understanding the functions of these female organs across species, it is important to remember that sperm maturation rarely ends in the testes of animals. As Pitnick et al. (2020) have pointed out, sperm activation in animals often takes place in the female reproductive tract. Modification of sperm after transfer to the female prolongs the ability of these sperm to be retained for future use, which predicts an evolutionary dynamic relationship of the reproductive gene products made by both males and females to support future and successful fertilization. It has been previously shown that male-biased gene expression evolves rapidly (Haerty *et al.*, 2007; Zhang *et al.*, 2007; Harrison *et al.*, 2015), and it is suspected that such changes also take place in the evolution of female-biased gene expression. In support of the dynamic behavior of gene expression across male and female reproductive organs, recent work by Sirot et al. (2014) supports the evolution of a novel serine endopeptidase through duplication of an existing FRT gene, indicating a switch in sex-biased expression of one of the duplicates from the ancestral pattern of female reproductive tract expression to the male accessory gland. My work on the evolution of these organs in females will hopefully shed some light on the frequency of gene expression changes in females across species, and will also serve as a foundation for characterizing female contributions to the male-female interactions that are important to animal fertility.

Why study the evolution of sperm storage in females of Drosophila?

So far, few studies have examined the co-evolution of male and female reproductive organs and products. By characterizing the transcriptomic contents of these dynamic female organs in Drosophila, these studies begin to compliment the large body of knowledge that exists regarding male contributions to sex. Also, better understanding the co-evolution of sexes has the potential to characterize the genetic bases of reproductive isolation between divergent populations and divergent species. The diverse male-female interactions that take place in female tissues are likely to be involved in speciation, and one cannot understand such interactions without first categorizing the female-specific contributions to reproduction.

By examining the content of these female tissue transcriptomes, I hypothesize that gene co-option may explain some of the novel abilities of sperm storage seen in the parovaria of *D. takahashii* and *D. nigricruria.* In a recent study by Meslin et al. (2015), the authors revealed a suite of highly expressed and secreted gene products in the female bursa that confer stomach-like traits for mechanical and enzymatic digestion of the spermatophore. By studying these bursa genes in an evolutionary framework, the authors found that the majority of these digestive genes were co-opted in expression into the bursa, while they also still remain expressed in their ancestral non-reproductive tissues. Other examples that illustrate the extent by which a pre-existing phenotype can change due to gene co-option include the strong association between the expression of dsx in Drosophila male forelegs and the presence of sex brushes in flies of the immigrans species group (Rice et al. 2018), and the association between

the expression of ebony and male-specific abdominal pigmentation (Signor et al. 2016), among others.

While the results from these tissue transcritpomes have not yet been studied, the underlying genomic mechanisms for the gain and loss of sperm storage in these female organs will begin to compliment the wealth of species data that exists for males in these species. While the female tissue transcriptomes have not yet been characterized, I am curious about the overall strategies used by females across Drosophila species in the storage of sperm. I am also intrigued about the amenability of gene expression switches between the sexes. When we compliment the male reproductive products to those made by females, do we find mostly find evidence for the continuity in expression of the same genes and by the same sex? Do we find evidence for gene expression switches between the sexes, given that these molecules are ultimately optimized to function in the female reproductive tract?

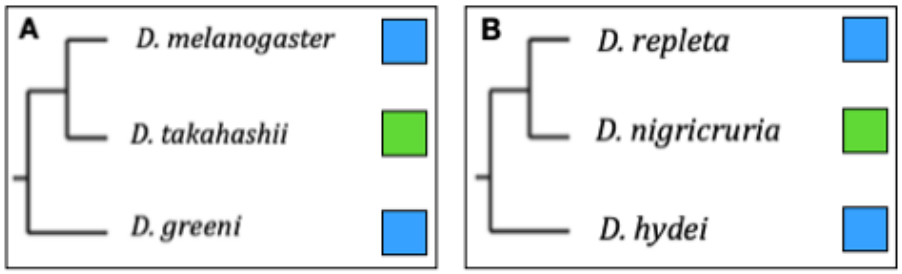

Figure 1. Gain of sperm storage in females. Panels A and B depict the Drosophila species chosen for the tissue comparisons. Blue colors indicate species where females only store sperm in the seminal receptacle and the spermathecae, whereas green colors indicate species where females store sperm in the seminal receptacle, spermathecae, and parovaria.

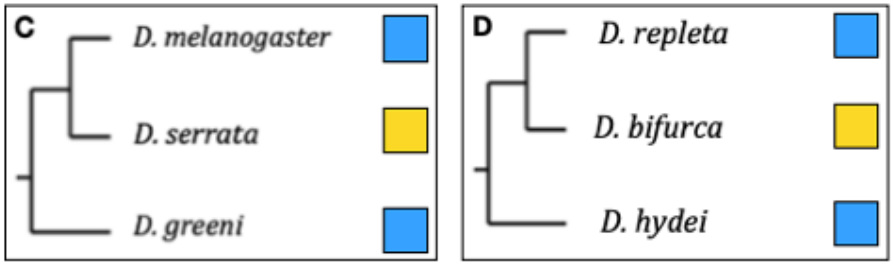

Figure 2. Loss of sperm storage in females. Panels C and D depict the Drosophila species chosen for the tissue comparisons. Blue colors indicate species where females store sperm in both the seminal receptacle and the spermathecae, whereas yellow colors indicate species where females only store sperm in the seminal receptacle, having lost the ability to store sperm in the spermathecae.

Species	Tissue	Replicate1 (Million)	Replicate2 (Million)	Replicate3 (Million)
D. melanogaster	Seminal Receptacle	4,251,173	4,197,622	4,108,152
D. melanogaster	Spermathecae	4,202,434	4,159,211	4,124,396
D. melanogaster	Parovaria	4,154,996	4,061,629	4,032,752
D. serrata	Seminal Receptacle	4,382,185	4,454,906	4,479,418
D. serrata	Spermathecae	4,162,191	4,478,506	4,511,651
D. serrata	Parovaria	4,451,787	4,461,173	4,450,532
D. takahashii	Seminal Receptacle	4,451,170	4,101,486	6,592,187
D. takahashii	Spermathecae	4,553,976	5,743,592	4,083,356
D. takahashii	Parovaria	4,123,693	4,057,115	6,119,308
D. greeni	Seminal Receptacle	4,180,173	4,090,413	4,406,926
D. greeni	Spermathecae	4,155,264	4,152,353	4,209,395
D. greeni	Parovaria	4,304,716	4,421,533	4,103,207

Table 1. Total RNA sequencing read numbers for the seminal receptacle, spermathecae, and parovaria of females in the melanogaster species subgroup. The numbers presented are in millions of reads per sample. These are single-end 80 bp reads.

Species	Tissue	Replicate 1 (Million)	Replicate 2 (Million)	Replicate 3 (Million)
D. repleta	Seminal Receptacle	Sequencing in progress	Sequencing in progress	Sequencing in progress
D. repleta	Spermathecae	Sequencing in progress	Sequencing in progress	Sequencing in progress
D. repleta	Parovaria	Sequencing in progress	Sequencing in progress	Sequencing in progress
D. nigricruria	Seminal Receptacle	4,611,617	4,353,412	4,560,164
D. nigricruria	Spermathecae	4,443,953	4,253,698	4,048,971
D. nigricruria	Parovaria	4,382,755	4,606,667	4,101,152
D. bifurca	Seminal Receptacle	4,022,522	4,084,433	4,162,980
D. bifurca	Spermathecae	4,061,723	4,093,491	4,133,574
D. bifurca	Parovaria	4,064,102	4,156,349	4,190,445
D. hydei	Seminal Receptacle	4,082,994	4,164,898	4,286,627
D. hydei	Spermathecae	4,078,058	4,038,672	4,104,426
D. hydei	Parovaria	4,512,175	4,308,182	4,802,648

Table 2. Total RNA sequencing read numbers for the seminal receptacle, spermathecae, and parovaria of females in the repleta species subgroup. The numbers presented are in millions of reads per sample. These are single-end 80 bp reads.

Species	Tissue	Replicate 1 concentration (ng/ul)	Replicate 2 concentration (ng/ul)	Replicate 3 concentration (ng/ul)
D. melanogaster	Seminal Receptacle	4.92	8.08	2.57
D. melanogaster	Spermathecae	8.24	9.18	0.90
D. melanogaster	Parovaria	2.20	3.71	2.30
D. serrata	Seminal Receptacle	2.60	2.92	2.68
D. serrata	Spermathecae	11.0	4.79	11.3
D. serrata	Parovaria	4.78	1.50	2.00
D. takahashii	Seminal Receptacle	3.23	2.40	1.80
D. takahashii	Spermathecae	24.1	8.08	6.96
D. takahashii	Parovaria	5.05	3.00	1.70
D. greeni	Seminal Receptacle	5.89	6.37	1.90
D. greeni	Spermathecae	2.20	4.78	0.66
D. greeni	Parovaria	0.90	0.18	0.45

Table 3. RNA concentrations for the seminal receptacle, spermathecae, and parovaria samples from females dissected in the melanogaster species subgroup.

Species	Tissue	Replicate 1 concentration (ng/ul)	Replicate 2 concentration (ng/ul)	Replicate 3 concentration (ng/ul)
D. repleta	Seminal Receptacle	18.6	17.4	16.4
D. repleta	Spermathecae	19.7	21.5	25.8
D. repleta	Parovaria	4.78	2.10	1.20
D. nigricruria	Seminal Receptacle	3.70	3.59	2.93
D. nigricruria	Spermathecae	5.86	3.94	3.79
D. nigricruria	Parovaria	2.50	0.28	1.70
D. bifurca	Seminal Receptacle	1.10	4.80	14.8
D. bifurca	Spermathecae	43.4	43.3	26.6
D. bifurca	Parovaria	13.4	7.59	6.93
D. hydei	Seminal Receptacle	7.45	13.8	11.3
D. hydei	Spermathecae	8.75	7.87	7.64
D. hydei	Parovaria	6.77	4.84	7.39

Table 4. RNA concentrations for the seminal receptacle, spermathecae, and parovaria samples from females dissected in the repleta species subgroup.

REFERENCES

Adams et al. (2000). The genome sequence of *Drosophila melanogaster*. Science. 287(5461): 2185-95.

Avila FW, Wolfner MF. (2009). Acp36DE is required for uterine conformational changes in mated Drosophila females. Proc Natl Acad Sci USA. 106 (37) 15796-15800.

Bloch Qazi MC, Wolfner MF. (2003). An early role for the *Drosophila melanogaster* male seminal protein Acp36DE in female sperm storage. J Exp Biol; 206:3521- 8.

Bosco et al. (2007). Analysis of Drosophila Species Genome Size and Satellite DNA Content Reveals Significant Differences Among Strains as Well as Between Species. Genetics 177: 1277-1290.

Carroll, S.B. (2005). Evolution at two levels: on genes and form. PLoS Biol *3*, e245.

Cavener DR. (1983). The response of enzyme polymorphisms to developmental rate selection in *Drosophila melanogaster.* Genetics; 105:105-13.

Chapman T, Neubaum DM, Wolfner MF, Partridge L. (2000). The role of male accessory gland protein Acp36DE in sperm competition in *Drosophila melanogaster.* Proc Biol Sci; 267:1097-105.

Davidson, E.H. (2006). The Regulatory Genome: Gene Regulatory Networks In Development And Evolution. Academic Press, 1st Edition.

Drosophila 12 Genomes Consortium (2007). Evolution of genes and genomes on the Drosophila phylogeny. *Nature* 450: 203-218.

Drummond AJ, Suchard MA, Xie D, Rambaut A. (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol Biol Evol. 29(8):1969-73.

Gilbert DG. (1981). Ejaculate esterase 6 and initial sperm use by female *Drosophila melanogaster.* J Insect Physiol ; 27:641-50.

Gilbert DG, Richmond RC. (1982). Esterase 6 in *Drosophila melanogaster*: reproductive function of active and null males at low temperature. Proc Natl Acad Sci USA; 79:2962-6.

Haerty, W., Jagadeeshan, S., Kulathinal, R. J., Wong, A., Ravi Ram, K., Sirot, L. K., Levesque, L., Artieri, C. G., Wolfner, M. F., Civetta, A. & Singh, R. S. (2007). Evolution in the fast lane: rapidly evolving sex-related genes in *Drosophila*. *Genetics* 177, 1321–1335.

Harrison, P. W., Wright, A. E., Zimmer, F., Dean, R., Montgomery, S. H., Pointer, M. A. & Mank, J. E. (2015). Sexual selection drives evolution and rapid turnover of male gene expression. *Proceedings of the National Academy of Sciences, USA* 112, 4393–4398.

Hebenstreit, D. et al. (2011). NA sequencing reveals two major classes of gene expression levels in metazoan cells. Molecular Systems Biology Vol. 7: 497.

King, M.C., and Wilson, A.C. (1975). Evolution at two levels in humans and chimpanzees. Science *188*, 107- 116.

LaFlamme BA, Ram KR, Wolfner MF. (2012). The *Drosophila melanogaster* seminal fluid protease "seminase" regulates proteolytic and post-mating reproductive processes. PLoS Genet. 8(1): e1002435.

Iida K, Cavener DR. (2004). Glucose dehydrogenase is required for normal sperm storage and utilization in female *Drosophila melanogaster.* J Exp Biol; 207:675- 81.

McGraw LA, Gibson G, Clark AG, Wolfner MF. (2004). Genes regulated by mating, sperm, or seminal proteins in mated female *Drosophila melanogaster.* Curr Biol; 14:1509-14.

Meslin, C, Melissa S. Plakke, Aaron B. Deutsch, Brandon S. Small, Nathan I. Morehouse, Nathan L. Clark. (2015). Digestive Organ in the Female Reproductive Tract Borrows Genes from Multiple Organ Systems to Adopt Critical Functions, *Molecular Biology and Evolution*, Volume 32, Issue 6, Pages 1567–1580.

Neubaum DM, Wolfner MF. (1999). Mated *Drosophila melanogaster* females require a seminal fluid protein, Acp36DE, to store sperm efficiently. Genetics; 153:845-57.

Pitnick, Scott, Therese Markow, and Greg S. Spicer. (1999). Evolution of Multiple Kinds of Female Sperm-Storage Organs in Drosophila. *Evolution* 53, no. 6: 1804–22.

Pitnick, S., Wolfner, M.F. and Dorus, S. (2020), Post-ejaculatory modifications to sperm (PEMS). Biol Rev, 95: 365-392.

Ram KR, Wolfner MF. (2009). A network of interactions among seminal proteins underlies the long-term post- mating response in Drosophila. Proc Natl Acad Sci USA; 106:15384-9.

Ravi KR, Wolfner MF. (2007). Sustained post-mating response in *Drosophila melanogaster* requires multiple seminal fluid proteins. PLoS Genet; 3:1-11.

Ravi Ram K, Sirot LK, Wolfner MF. (2006). Predicted seminal astacin-like protease is required for processing of repro- ductive proteins in *Drosophila melanogaster*. Proc Natl Acad Sci USA.

Rice, G, Barmina, O, Hu, K, Kopp, A. (2018). Evolving *doublesex* expression correlates with the origin and diversification of male sexual ornaments in the *Drosophila immigrans* species group. *Evolution & Development*. 20: 78– 88.

Schiff NM, Feng Y, Quine JA, Krasney PA, Cavener DR. (1992). Evolution of the expression of the Gld gene in the reproductive tract of Drosophila*.* Mol Biol Evol. 9:1029-49.

Schnakenberg SL, Matias WR, Siegal ML. (2011). Sperm- storage defects and live birth in Drosophila females lacking spermathecal secretory cells. PLoS Biol; 9:e1001192.

Signor SA, Liu Y, Rebeiz M, and Kopp A. (2016). Genetic Convergence in the Evolution of Male-Specific Color Patterns in Drosophila. *Current Biology,* pii: S0960-9822(16)30785-0.

Sirot, L. K., Findlay, G. D., Sitnik, J. L., Frasheri, D., Avila, F. W. & Wolfner, M. F. (2014). Molecular characterization and evolution of a gene family encoding both female- and male-specific reproductive porteins in *Drosophila*. *Molecular Biology and Evolution* 31, 1554–1567.

Song X, et al. (2011). The 19 genomes of Drosophila: a BAC library resource for genus-wide and genome-scale comparative evolutionary research. Genetics. Apr;187(4):1023-30.

Sun, J., and Spradling, A.C. (2013). Ovulation in *Drosophila* is controlled by secretory cells of the female reproductive tract. *eLife* 2:e00415.

Pitnick, S., Markow, T., Spicer, G.S. (1999). Evolution of multiple kinds of female sperm-storage organs in Drosophila. Evolution 53(6): 1804-1822.

Wagner GP, Kin K, Lynch VJ. (2013). A model based criterion for gene expression calls using RNA-seq data. Theory Biosci. 132(3):159-64.

Wong A, Albright SN, Giebel JD, Ram KR, Ji S, Fiumera AC, et al. (2008).A role for Acp29AB, a predicted seminal fluid lectin, in female sperm storage in *Drosophila melanogaster.* Genetics, 180:921-31

Zhang, Y., Sturgill, D., Parisi, M., Kumar, S. & Oliver, B. (2007). Constraint and turnover in sex-biased gene expression in the genus *Drosophila*. *Nature* 450, 233 – 237.

Chapter 3

Re-Wiring of Gene Regulatory Networks in the *Drosophila* Testes:

Mechanism of Gene Co-Option and Transcriptome Evolution

INTRODUCTION

The relationship between the evolution of genomes and organismal phenotypes is complex and multifaceted. It has long been appreciated that changes in gene regulation are a major force in phenotypic evolution, and that the origin of new traits does not necessarily require the origin of new genes (Carroll, 2005; Davidson, 2006; King and Wilson, 1975). In particular, gene cooption, or the exaptation of pre-existing genes for new organismal functions, is one of the major sources of evolutionary innovation and diversification, and gene cooption results in rewiring of the pre-existing gene regulatory network (Monteiro and Podlaha, 2009; Rosin and Kramer, 2009; Schlosser and Wagner, 2004; True and Carroll, 2002). A number of studies suggest that changes in the developmental regulation of evolutionarily conserved genes can give rise to new morphological structures (Gotoh et al., 2014; Kijimoto et al., 2012; Tanaka et al., 2011). However, little is known about the contribution of gene regulatory network rewiring to the remodeling of tissue-specific transcriptomes, which takes place via gains and losses of gene expression, a process known as transcriptome turnover.

Most work on transcriptome evolution has focused on quantitative variation in gene expression levels between and within species (Brawand et al., 2011; Enard et al., 2002; Gilad et al., 2006; Khaitovich et al., 2005b; McManus et al., 2010; Nuzhdin et al., 2004; Ranz et al., 2003). However, qualitative (gene presence/absence) differences between the transcriptomes of different species have the potential to produce the greatest evolutionary shifts in organ function: The recruitment of the *Prolactin* gene into the endometrial decidual cells during the

evolution of pregnancy in placental mammals is a prime example (Emera et al., 2012; Emera and Wagner, 2012; Lynch et al., 2011). In primates, the origin of corticotropin-releasing hormone expression in the placenta has been linked to the evolution of the fetal adrenal zone, a structure unique to primates that in turn stimulates estrogen production in the placenta (Power and Schulkin, 2006). Similarly, the evolution of optically complex eyes in many animal lineages involved the origin and regulatory recruitment of a variety of lens crystallin genes (Cvekl and Piatigorsky, 1996; True and Carroll, 2002).

Since transcriptome turnover is a consequence of evolutionary changes in the regulatory networks that control tissue development, I proposed to use the gain of esg expression in the testes as a model to study the effects of gene regulatory rewiring on tissue function. This is a great model for two reasons. Firstly, male reproduction evolves rapidly both in *Drosophila* and in mammals (Brawand et al., 2011; Ellegren and Parsch, 2007; Khaitovich et al., 2005a; Meiklejohn et al., 2003; Parsch and Ellegren, 2013; Voolstra et al., 2007), leading us to expect substantial transcriptome turnover over short evolutionary distances. Secondly, the gene esg is autonomously expressed in three cell types that constitute the testis stem cell niche in *D. melanogaster,* and loss of expression of this gene leads to the progressive loss of all cell types in the testis niche, which has serious consequences to the testis transcriptome and to the male's fertility in this species. However, this gene does not seem to have the same roles in the testes of other Drosophila examined, and our data support a gain of expression in the testes of *D. melanogaster* after divergence from *D. simulans.* This model will allow me to test if 1) Changes in the non-coding region led to the gain of esg expression in the testes of *D.*

melanogaster, and 2) Rewiring of gene regulatory networks in the testes led to changes in tissue function.

Preliminary data

Genome assemblies and testes transcriptomes.

In collaboration with others in the modEncode consortium, the Kopp Lab sequenced the genomes of 8 new species from the *D. melanogaster* species group (Chen, 2014). For most of these species, the resulting genome assemblies are of very high quality with scaffold N50 between 390kb and 3128kb (Table 1). Together with the previously sequenced genomes (Clark et al., 2007), we now have access to complete genome sequences of 14 representatives of the *D. melanogaster* species group (Fig. 8). The Kopp Lab then selected 11 species from the melanogaster species group for further analysis based on their phylogenetic position and the availability of high-quality genome assemblies (Table 1). For each of these species, the Kopp lab sequenced the transcriptomes of testes using pools of 30-50 mixed-age males raised under standard conditions. For each species, the strain chosen was the same strain used for the reference genome sequence (Chen, 2014). Sequencing libraries were prepared from 1 mcg of total RNA per sample. ERCC control spikes (Jiang et al., 2011; Loven et al., 2012) were added to each library to enable data normalization and presence/absence calls. All libraries were sequenced on Illumina HiSeq 3000 or 4000 with long 150bp or 100bp paired-end reads.

Identification of Esg transcriptional targets by DamID.

To better understand the gene regulatory network that includes escargot, our collaborators at UCLA dissected whole testes from flies expressing low levels of Dam::Esg fusion protein and from control flies expressing Dam alone (method visualized in figure 12), and tissues were immediately frozen on dry ice and transferred to 80°C. Genomic DNA was isolated from approximately 50 testes per genotype and processed following the protocol in Choksi et al (2006). Triplicate samples of labeled DNA were hybridized with a dye-swap to Nimble- Gen 2.1 M whole-genome tiling arrays (Roche) at the FlyChip facility (www.flychip.org.uk). DamID data were analyzed to identify Esg binding regions (EBRs) with minor modifications to the protocol in Southall and Brand (2009). Overall, these analyses suggest that 314 and 1100 target genes are potentially regulated by Esg when Esg TFBS are examined within 2kb and 5kb of target genes, respectively.

MATERIALS AND METHODS

Fly Stocks

The following stocks were used in my study: *D. melanogaster* w^{1118} (identifier FBal0018186), the GFP trap line *D. melanogaster*: esg-GFP (gift from Lynn Cooley and Leanne Jones, identifier P01986), *D. simulans* 14021-0251.011 (National Drosophila Species Stock Center), and *D. yakuba* 14021-0261.01 011 (National Drosophila Species Stock Center). For each species, the wildtype strain chosen was the same strain used for the reference genome sequence (Chen, 2014).

Immunostaining and Microscopy

Fly cultures were raised on standard *Drosophila* media at 22°C, and virgin males were collected and stored in low-density vials. Tissues were dissected in 1x PBS, processed, and immunostained by fixing tissues for 20 minutes at room temperature in 4% formaldehyde (0.1 M Pipes pH = 6.9, 1 mM EGTA pH = 7.0, 2 mM MgCl, 1% Triton X-100). This was followed by three washes in 1x PBS and three washes in 0.1 M Tris-HCl/0.3 M NaCl (pH 7.4) with 0.5% Triton X-100 (TNT) for 15 minutes each. Tissues were then blocked in Invitrogen/molecular probes solution Image-iT FX signal enhancer (Cat #136933) for 30 minutes at room temperature, and washed three times in TNT. The GFP trap line *D. melanogaster* esg-GFP did not require any additional staining and was mounted in Fluomount G to be visualized immediately afterwards under confocal microscopy. After blocking tissues dissected from wild-type flies, appropriate primary antibodies were added to the samples dissected and incubated at 4 °C overnight to allow ample time for the antibodies to bind at both tested dilutions. The primary antibody used was rat anti-esg and was used at both 1:1000 and at 1:100 dilutions in TNT as suggested by Dr. Hayashi, who could not guarantee the efficacy of this antibody more than 25 years after it was developed by his laboratory. The secondary antibody used was AlexaFluor555-conjugated goat anti-rat (1:500 dilution, Thermofisher). To minimize cross-reactivity, this secondary antibody was adsorbed against mouse IgG, mouse serum, and human serum prior to conjugation. For some of the later imaging in *D. melanogaster* testes, the secondary antibody was pre-absorbed with testes and intestines dissected from other males to reduce background staining and to produce a clearer signal. Lastly, tissues were mounted in Prolong Gold after six 15minute washes in TNT. Fluorescent images were taken on an Olympus 1000 confocal microscope at the

MCB LM Imaging Facility at the University of California Davis. Images were processed using Image J software (v2.0.0, Wayne Rasband, National Institute of Health, http://imagej.nih.gov/ij) and Adobe Photoshop, and at least eight testes were examined for each study with 20x and 40x oil-immersion objectives. Z-series projection were produced using ImageJ.

RNA extraction and quantitative RT-PCR

After dissection from 7 day old and sexually males *in D. melanogaster, D. simulans,* and *D. yakuba*, eighty testes per species replicate were frozen at -80°C in fresh Trizol buffer (Trizol Life Technologies, 1559602659) before RNA was extracted following a standard Trizol-chloroform extraction protocol. For each replicate, first strand complementary DNA (cDNA) was generated from 1 ug of purified total RNA using Superscript III reverse transcriptase (Invitrogen) and oligo(dT)12-18 and random hexamers. The cDNA was then used to as a template for qPCR analysis using SYBR green detection on a BioRad cycler. The following reactions were carried out in triplicate and melting curves were examined to ensure single products. Lastly, results were quantified using the delta-delta-Ct method (Livak and Schmittgen 2001) to normalize to Act5C transcript levels and to control for species genotypes. The data present averages and standard deviations from at least 3 independent experiments. The primer pairs chosen were published earlier (Xing and Li 2015) and the following forward and reverse primers were used.

Act5C: GATCTGTATGCCAACACCGT, GCGGGGCAATGATCTTGATC

esg: TACCCATCATCACCATGCGCCTAT, TCCCGGCTGGCTAGTGTTTAGATT

RESULTS

Immunostaining and Microscopy

Both panels in figure 1 support the expression and presence of the escargot transcription factor in the testes of *D. melanogaster* males, as evidenced by visualizing testes dissected from the enhancer trap Esg-GFP. These images support specific presence of the esg factor in cells that make up the stem cell niche (figure 11 for reference) at the distal end of the Drosophila testis. These images also support the absence of the esg transcription factor in the remaining cell types that make up the testis in this species. Dissection of the intestines in this transgenic line also supports the expression of esg in the stem cells that make up the intestine (figure 2) but nowhere else in that organ, an expression pattern earlier characterized and described by colleagues in the field (Korzelius et al. 2014, Loza-Coll et al. 2014).

When using antibodies to stain the testes of *D. melanogaster*, the 555 Secondary Goat Anti Rat antibody alone (figure 3, left panel) produces a strong background signal. Similar images were obtained (figure 3, right panel) when imaging testes that were stained with the primary anti-esg and the secondary antibody chosen. To potentially decrease the background signal in the testes, aliquots of the secondary antibody were later pre-absorbed with testes and intestines dissected from *D. melanogaster* males, and these aliquots were used to produce the images seen in figure 4. Similarly to the earlier results, a strong background continued to be seen in these confocal images. This staining pattern was also observed when studying the expression of esg in the testes of *D. simulans* (figure 5), where the secondary antibody alone produced a strong background signal in the testes of this species.

Quantitative RT-PCR

To validate the result of differential expression of esg in our preliminary transcriptome study of the testes in *D. melanogaster, D. simulans,* and *D. yakuba* (figure 7), I performed qPCR experiments using testes from these three species to determine the relative change in expression of escargot across this tissue. I found that escargot was 27 times more highly expressed in the testes of *D. melanogaster* relative to *D. simulans,* and that escargot was absent in expression in the testes of *D. yakuba* (figure 6).

DISCUSSION

Numerous challenges were encountered while working on this study. Whether the secondary antibody was pre-absorbed or not, the strong signal could not be avoided when visualizing the testes under confocal microscopy in *D. melanogaster* and in *D. simulans*. This fact alone made the comparison between treatments of the same tissue impossible, as there was no mechanism that could reliably help compare tissues stained with a secondary antibody alone with those tissues stained with both primary and secondary antibodies. While it was comforting to know that escargot was indeed expressed in the stem cell niche of *D. melanogaster* testes, as validated by my images of the enhancer trap Esg-GFP line, I could not reproduce quality images of specific staining in the testes using this secondary antibody. To make matters more challenging, the anti-escargot antibody produced by Dr. Shigeo Hayashi was the only antibody against escargot that has ever been used successfully for the purposes of

tissue staining. Since then, a handful of other escargot antibodies were generated by other academic and non-academic labs, but none of them have been able to be used for immunostaining and have only been used in Western blots (personal communication with various colleagues).

Separately from my immunostaining studies, it was exciting to validate the differential expression of escargot in the testes of these three species using qPCR.

All of my qPCR studies independently supported the differential expression of esg seen across the testes of *D. melanogaster, D. simulans,* and *D. yakuba.* While I could not visualize the difference in expression seen for this important gene across species testes, I hope that future studies can verify the minimal expression of escargot in the testes of *D. simulans* relative to *D. melanogaster*, and to verify the absence in expression of escargot in the testes of *D. yakuba*. If that is indeed the case, then I am curious about localization of escargot products in the testes of *D. simulans*. While escargot is expressed in the hub cells, cyst stem cells, and germline stem cells of the testes in *D. melanogaster,* I wonder if esg expression is potentially limited to one or two of the stem cell types in the testes of *D. simulans.* Another hypothesis for the lower expression of esg in the testes of *D. simulans* could be related to a significantly lower expression of this gene in each of the three cell types where it is expressed in *D. melanogaster* testes. A third scenario that might explain the strong difference in expression of this gene between *D. melanogaster* and *D. simulans* testes could be related to the spatial expression of this gene, where escargot might be lowly expressed in the testes of *D. simulans* but in cells outside of the stem cell niche. No matter where the escargot gene is expressed in the testes of *D. simulans,* our testes transcriptome data (figure 7) seem to support that the expression of

escargot is a derived feature of the testes in flies of the genus Drosophila, and that this gene has quickly become an important regulator of testis somatic stem cell fate (Demarco et al. 2022, 2020) in at least one species, *D. melanogaster*, after divergence from ancestral flies in this group. Lastly, and outside of the testes, the transcription factor escargot also regulates intestinal stem cell fate in *D. melanogaster*. While there are no studies has been published that examine the expression of this gene in the intestines of species outside of *D. melanogaster*, I wonder if such qualitative differential expression could also be seen in other Drosophila tissues. Escargot is an important transcription factor in the Snail family that maintains stemness is various tissues of *D. melanogaster*, and this feature seems to be derived in the testes of *D. melanogaster* as well as other species including *D. kikkawai* and *D. ficusphila,* as seen in our tramscriptome study of the testes across species (figure 7). In the future, I encourage colleagues to study the derived expression of escargot in the testes of these species, as this gene product seems to have rapidly rewired at least one gene regulatory network underlying the cyst stem cells of the *D. melanogaster* testes, and yet nothing is known about the expression of this gene outside of *D. melanogaster*. If colleagues do pick up such studies, I encourage them to use a different method to visualize the presence of this transcription factor across species tissues, and perhaps the more modern method of in situ hybridization chain reaction could visualize the differences in expression that are suggested by our testes transcriptome and qRT-PCR studies.

TABLES AND FIGURES

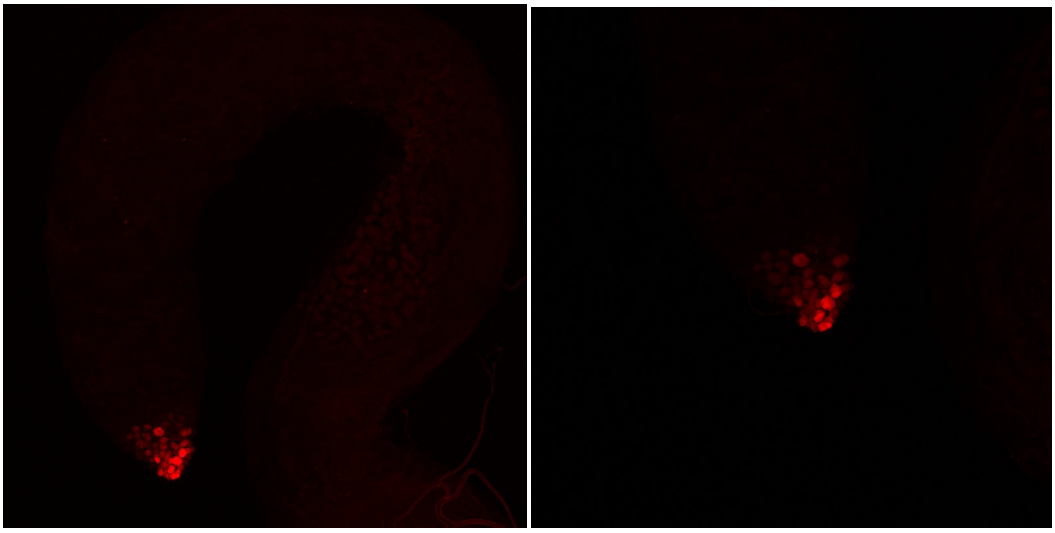

Figure 1. Left Panel: Adult testis in *D. melanogaster* showing expression of the esg gene in the stem cell niche using enhancer trap Esg-GFP. Note that esg shows little to no expression outside of the testis niche. Right Panel: Magnified image supports expression of the esg gene in the *D. melanogaster* niche.

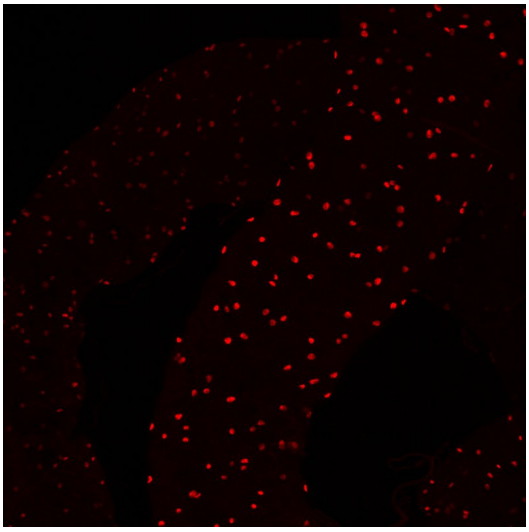

Figure 2. Adult intestines in *D. melanogaster* showing expression of the esg gene in the intestinal stem cells using enhancer trap Esg-GFP.

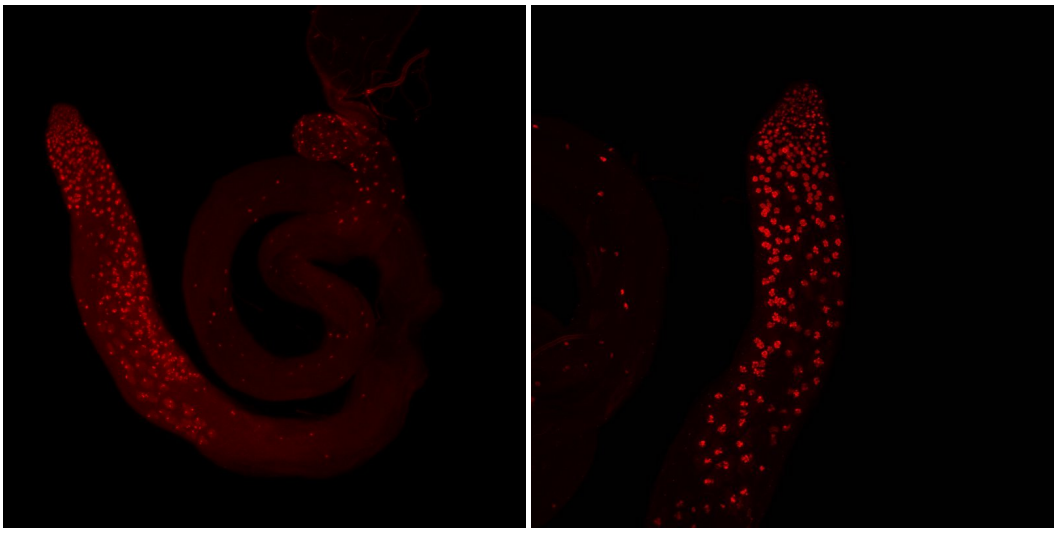

Figure 3. Left Panel: *D. melanogaster* adult testis stained with 555 Secondary Goat Anti Rat antibody only. Right Panel: *D. melanogaster* adult testis stained with Hayashi anti-esg antibody and 555 Secondary Goat Anti Rat antibody.

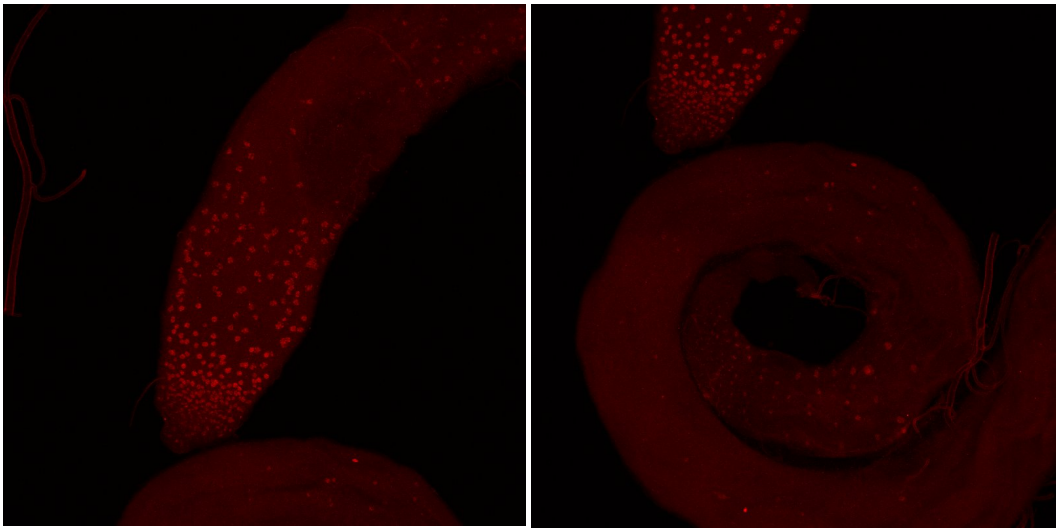

Figure 4. Left and Right Panels: *D. melanogaster* adult testes stained with a pre-absorbed 555 Secondary Goat Anti Rat antibody only.

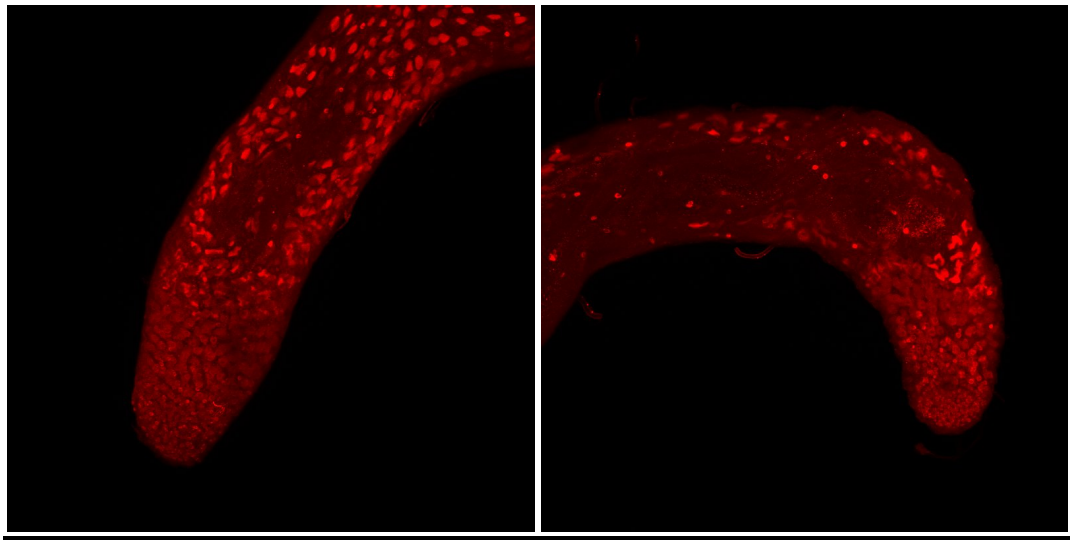

Figure 5. Left Panel: *D. simulans* adult testis stained with 555 Secondary Goat Anti Rat antibody only. Right Panel: *D. simulans* adult testis stained with Hayashi anti-esg antibody and 555 Secondary Goat Anti Rat antibody.

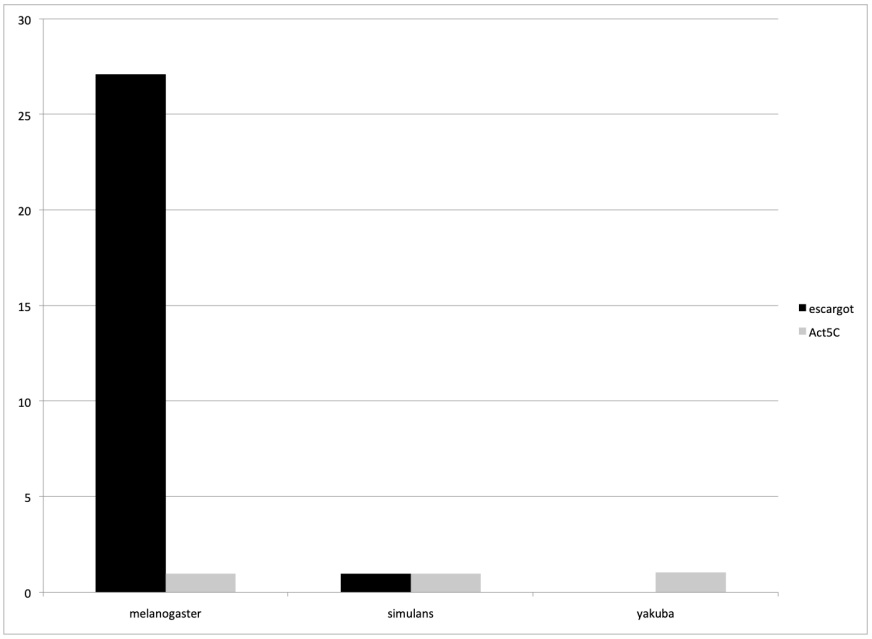

Figure 6. Relative escargot expression in the testes of *D. melanogaster, D. simulans,* and *D. yakuba.* Sexually mature males were dissected to quantify mRNA levels of both Act5C (control) and escargot. The results were normalized to Act5C.

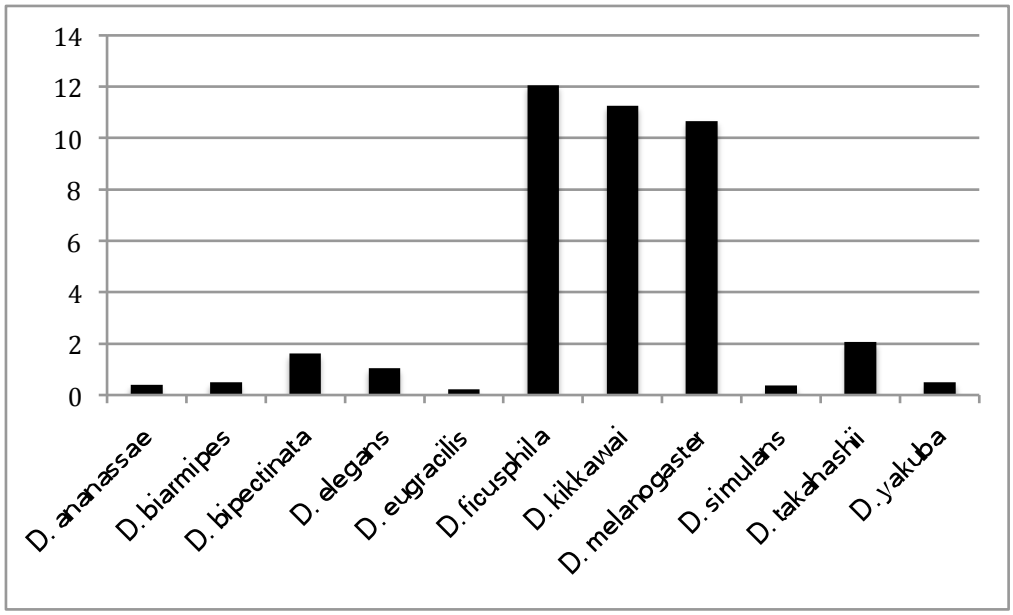

Figure 7. Escargot expression (TPM) examined in 11 closely-related species of Drosophila in the *melanogaster* group.

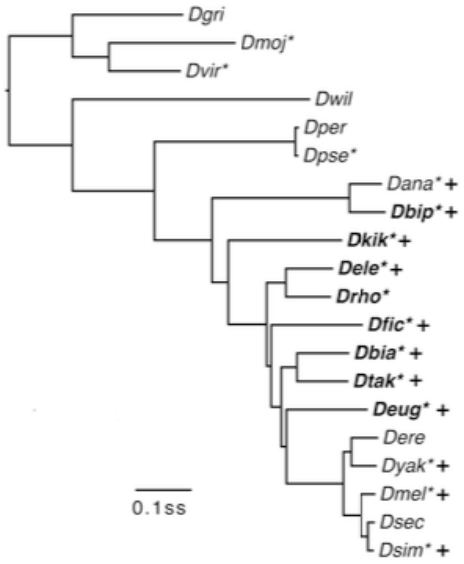

Figure 8. Phylogeny of the species used in this project. Newly sequenced genomes are in bold. (*****) RNA-seq data for adult males and females and embryos; while (**+**) represents those species with RNA-seq data for AGs and testes.

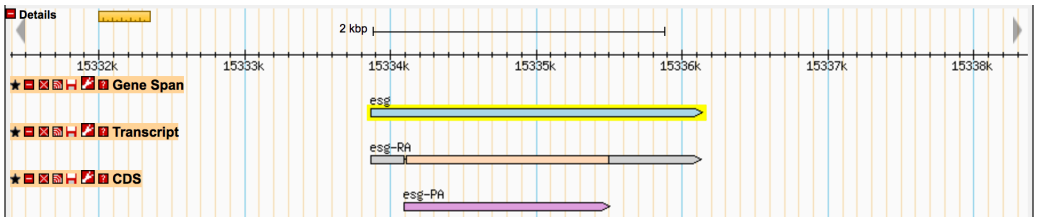

Figure 9. Gene span shows total extent of transcribed region of the esg gene in *D. melanogaster*, with direction of transcription indicated. Transcript shows the exon (1410bp, orange) and noncoding sequence (grey). CDS shows extent of sequence (470AA) encoding Escargot polypeptide.

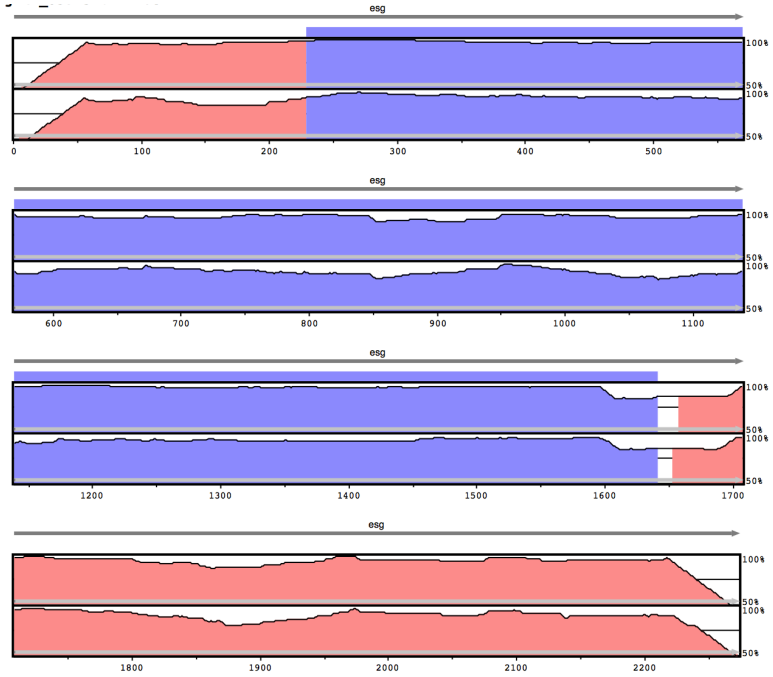

Figure 10. Extent of alignment of the escargot gene between *D. melanogaster* (reference sequence) and *D. simulans* (top track), and between *D. melanogaster* (reference) and *D. yakuba* (bottom track). Grey arrow parallels gene length. Purple color reflects exon. Pink color reflects conserved non-coding sequence.

Species	Genome coverage	Genome scaffold N50	Mapped RNA-seq reads, millions				
			Male	Female	Embryo	Testis	AG
D. biarmipes	35.2X	3,128 kb	58.8	56.9	196.8	15.1	20.6
D. bipectinata	36.3X	663 kb	45.4	37.1	258.8	26.3	34.1
D. elegans	36.7X	1,714 kb	48.8	50.2	297.8	39.1	26.1
D. eugracilis	40.7X	977 kb	51.6	60.1	201.6	219.7	34.7
D. ficusphila	36.1X	1,049 kb	78.8	81.2	200.5	29.5	27.4
D. kikkawai	34.2X	911 kb	47.2	34.2	168.6	28.1	30.4
D. rhopaloa	30.4X	45 kb	59.4	45.4	192.3		
D. takahashii	31.7X	390 kb	52.3	49.4	197.1	22.3	29.6
D. yakuba	*	*	96.5	91.8	#	24.6	23.3
D. simulans	*	*	172.1	151.7	#	172.1	50.9
D. melanogaster	*	*	103.4	96.2	^	21.2	20.2
D. ananassae	*	*	76.8	92.1		19.9	18.4

Table 1. Drosophila genome and RNAseq data. (*) Published genome (Clark et al., 2007).

(#) Published data (Paris et al., 2013). (^) modENCODE data.

Stage of Development	Region of Expression
Embryonic	Central brain primordium, dorsal ectoderm, ventral ectoderm, antennal primodrium, visual primordium, tracheal morphogenesis, peripheral nervous system, wing disc, abdominal histoblasts, gonad, genital disc, haltere disc, wing disc, prothoracic leg disc, metathoracic leg disc, dorsal ridge, mesothoracic leg disc, spiracle
Adult Male	testis stem cell niche cells, male and female intestinal stem cells, malphigian tubules, salivary glands, neuroblasts,

Table 2. Escargot expression in regions of the embryo and adult in *D. melanogaster*.

Gene	**Cell Type**	**Stocks**
unpaired	Hub cells	upd-GAL4
traffic jam	Cyst stem cells	tj-GAL4
nanos	Germline stem cells	nos-GAL4-VP16

Table 3. Gal4 drivers used to drive the expression of Gal4 in the hub, Cyst Stem Cells, and Germline Stem Cells in *D. melanogaster* adult testes.

REFERENCES

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate - a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society Series B-Statistical Methodology *57*, 289- 300.

Brawand, D., Soumillon, M., Necsulea, A., Julien, P., Csardi, G., Harrigan, P., Weier, M., Liechti, A., Aximu- Petri, A., Kircher, M*., et al.* (2011). The evolution of gene expression levels in mammalian organs. Nature *478*, 343-348.

Carroll, S.B. (2005). Evolution at two levels: on genes and form. PLoS Biol *3*, e245.

Chen, Z.-J.e.a. (2014). Comparative Validation of the *D. melanogaster* modENCODE transcriptome annotation. Genome Biol *in press*.

Choksi SP, Southall TD, Bossing T, Edoff K, de Wit E, Fischer BE, van Steensel B, Micklem G, Brand AH (2006) Prospero acts as a binary switch between self-renewal and differentiation in Drosophila neural stem cells. Dev Cell 11: 775 – 789.

Clark, A.G., Eisen, M.B., Smith, D.R., Bergman, C.M., Oliver, B., Markow, T.A., Kaufman, T.C., Kellis, M., Gelbart, W., Iyer, V.N*., et al.* (2007). Evolution of genes and genomes on the Drosophila phylogeny. Nature *450*, 203-218.

Cvekl, A., and Piatigorsky, J. (1996). Lens development and crystallin gene expression: many roles for Pax-6. Bioessays *18*, 621-630.

Davidson, E.H. (2006). The Regulatory Genome: Gene Regulatory Networks In Development And Evolution (Academic Press).

Senos Demarco, R., Uyemura, B.S., and Jones, D.L. (2020). Egfr signaling stimulates autophagy to regulate stem cell maintenance and lipid homeostasis in the *Drosophila* testis. Cell Rep *30*, 1101–1116.e5.

Sênos Demarco R, Stack BJ, Tang AM, Voog J, Sandall SL, Southall TD, Brand AH, Jones DL. (2022). Escargot controls somatic stem cell maintenance through the attenuation of the insulin receptor pathway in Drosophila. Cell Rep. 39(3):110679.

Ellegren, H., and Parsch, J. (2007). The evolution of sex-biased genes and sex- biased gene expression. Nat Rev Genet *8*, 689-698.

Emera, D., Casola, C., Lynch, V.J., Wildman, D.E., Agnew, D., and Wagner, G.P. (2012). Convergent evolution of endometrial prolactin expression in primates, mice, and elephants through the independent recruitment of transposable elements. Mol Biol Evol *29*, 239-247.

Emera, D., and Wagner, G.P. (2012). Transformation of a transposon into a derived prolactin promoter with function during human pregnancy. Proc Natl Acad Sci U S A *109*, 11246-11251.

Enard, W., Khaitovich, P., Klose, J., Zollner, S., Heissig, F., Giavalisco, P., Nieselt-Struwe, K., Muchmore, E., Varki, A., Ravid, R*., et al.* (2002). Intra- and interspecific variation in primate gene expression patterns. Science *296*, 340-343.

Fuse, N., Hirose, S. and Hayashi, S. (1994). Diploidy of Drosophila imaginal cells is maintained by a transcriptional repressor encoded by escargot. Genes & Development, 8(19), pp. 2270–2281.

Gilad, Y., Oshlack, A., Smyth, G.K., Speed, T.P., and White, K.P. (2006). Expression profiling in primates reveals a rapid evolution of human transcription factors. Nature *440*, 242-245.

Gotoh, H., Miyakawa, H., Ishikawa, A., Ishikawa, Y., Sugime, Y., Emlen, D.J., Lavine, L.C., and Miura, T. (2014). Developmental Link between Sex and Nutrition; doublesex Regulates Sex-Specific Mandible Growth via Juvenile Hormone Signaling in Stag Beetles. PLoS Genet *10*, e1004098.

Jiang, L., Schlesinger, F., Davis, C.A., Zhang, Y., Li, R., Salit, M., Gingeras, T.R., and Oliver, B. (2011). Synthetic spike-in standards for RNA-seq experiments. Genome Res *21*, 1543-1551.

Khaitovich, P., Paabo, S., and Weiss, G. (2005b). Toward a neutral evolutionary model of gene expression. Genetics *170*, 929-939.

Kijimoto, T., Moczek, A.P., and Andrews, J. (2012). Diversification of doublesex function underlies morph-, sex-, and species-specific development of beetle horns. Proc Natl Acad Sci U S A *109*, 20526-20531.

King, M.C., and Wilson, A.C. (1975). Evolution at two levels in humans and chimpanzees. Science *188*, 107- 116.

Korzelius, J., Naumann, S.K., Loza-Coll, M.A., Chan, J.S., Dutta, D., Oberheim, J., Glasser, C., Southall, T.D., Brand, A.H., Jones, D.L., and Edgar, B.A. (2014). Escargot maintains stemness and suppresses differentiation in *Drosophila* intestinal stem cells. EMBO J. *33*, 2967–2982.

Livak, K. J., and T. D. Schmittgen, 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)). *Methods* 25 402–408.

Loven, J., Orlando, D.A., Sigova, A.A., Lin, C.Y., Rahl, P.B., Burge, C.B., Levens, D.L., Lee, T.I., and Young, R.A. (2012). Revisiting global gene expression analysis. Cell *151*, 476-482.

Loza-Coll, M.A., Southall, T.D., Sandall, S.L., Brand, A.H., and Jones, D.L. (2014). Regulation of *Drosophila* intestinal stem cell maintenance and differentiation by the transcription factor escargot. EMBO J. *33*, 2983–2996.

Lynch, V.J., Leclerc, R.D., May, G., and Wagner, G.P. (2011). Transposon- mediated rewiring of gene regulatory networks contributed to the evolution of pregnancy in mammals. Nat Genet *43*, 1154-1159.

McManus, C.J., Coolon, J.D., O'Duff, M., Eipper-Mains, J., Graveley, B.R., and Wittkopp, P.J. (2010). Regulatory divergence in Drosophila revealed by mRNA-seq. Genome Res.

Meiklejohn, C.D., Parsch, J., Ranz, J.M., and Hartl, D.L. (2003). Rapid evolution of male-biased gene expression in Drosophila. Proc Natl Acad Sci U S A *100*, 9894- 9899.

Monteiro, A., and Podlaha, O. (2009). Wings, horns, and butterfly eyespots: how do complex traits evolve? PLoS Biol *7*, e37.

Nuzhdin, S.V., Wayne, M.L., Harmon, K.L., and McIntyre, L.M. (2004). Common pattern of evolution of gene expression level and protein sequence in Drosophila. Mol Biol Evol *21*, 1308-1317.

Parsch, J., and Ellegren, H. (2013). The evolutionary causes and consequences of sex-biased gene expression. Nat Rev Genet *14*, 83-87.

Power, M.L., and Schulkin, J. (2006). Functions of corticotropin-releasing hormone in anthropoid primates: from brain to placenta. Am J Hum Biol *18*, 431-447.

Ranz, J.M., Castillo-Davis, C.I., Meiklejohn, C.D., and Hartl, D.L. (2003). Sex- dependent gene expression and evolution of the Drosophila transcriptome. Science *300*, 1742-1745.

Rosin, F.M., and Kramer, E.M. (2009). Old dogs, new tricks: regulatory evolution in conserved genetic modules leads to novel morphologies in plants. Dev Biol *332*, 25- 35.

Schlosser, G., and Wagner, G.P. (2004). Modularity in development and evolution (Chicago: University of Chicago Press).

Senos Demarco, R., Uyemura, B.S., and Jones, D.L. (2020). Egfr signaling stimulates autophagy to regulate stem cell maintenance and lipid homeostasis in the Drosophila testis. Cell Rep 30, 1101–1116.e5.

Southall TD, Brand AH (2009) Neural stem cell transcriptional networks highlight genes essential for nervous system development. EMBO J 28: 3799 – 3807.

Tanaka, K., Barmina, O., Sanders, L.E., Arbeitman, M.N., and Kopp, A. (2011). Evolution of Sex-Specific Traits through Changes in HOX-Dependent doublesex Expression. PLoS Biol *9*, e1001131.

True, J.R., and Carroll, S.B. (2002). Gene co-option in physiological and morphological evolution. Annu Rev Cell Dev Biol *18*, 53-80.

Voolstra, C., Tautz, D., Farbrother, P., Eichinger, L., and Harr, B. (2007). Contrasting evolution of expression differences in the testis between species and subspecies of the house mouse. Genome Res *17*, 42-49.

Xing, Y., Li, W. Heterochromatin components in germline stem cell maintenance. (2015) *Sci Rep* 5, 17463.