UC Davis
UC Davis Electronic Theses and Dissertations

Title

Natural Variation and Evolution of Early Gene Expression and Transcript Localization in Drosophila

Permalink

https://escholarship.org/uc/item/4dn4n67v

Author

Feitzinger, Anna

Publication Date

2022

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

Natural Variation and Evolution of Early Gene Expression and Transcript Localization in Drosophila

By

ANNA FEITZINGER
DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Biochemistry, Molecular, Cellular and Developmental Biology

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

Susan Lott, Chair

Celina Juliano

Artyom Kopp

Committee in Charge

2022

ABSTRACT

Gene expression is a multi-step process that is fundamental to all cellular activities. In *Drosophila* oogenesis and embryogenesis both RNA localization and transcription are highly regulated and critical for proper development of the embryo. During oogenesis, maternal RNAs spatially localized within the oocyte. A few localized maternal mRNAs that are required for the proper axial patterning of the embryo have been well-studied. However, a large proportion of the genome is expressed in the ovary and deposited into the embryo. These maternally derived RNAs and proteins drive the earliest events of embryogenesis before the zygotic genome is activated and gene products derived from the zygotic genome take over the control of development. Both maternal RNA localization and early embryo gene expression have undergone substantial changes over the course of evolution in *Diptera* between species of different genera. However, little was previously known about the evolutionary dynamics of mRNA localization and early gene expression on shorter timescales within a genus or within a species. In this dissertation I determine the extent of natural variation in maternal and zygotic RNA complements within and between populations of a single species, *Drosophila melanogaster*. Additionally, I examine how maternal RNA localization changes have occurred within the *Drosophila* genus.

In the first chapter of this dissertation, I utilize two geographically distinct populations of *Drosophila melanogaster* with known differences in genomic variation to examine the natural variation of maternal and zygotic transcripts within and between populations of a single species. I find that maternal gene expression has higher conservation than zygotic gene expression, similar to what has been found when comparing different species of *Drosophila*. Overall, there is more expression variation within populations than fixed expression differences between them. Furthermore, I find an increase in differential expression of maternal and zygotic transcriptomes within the Zambia population, a population with higher genomic variation, than the Raleigh population. I find that some of the most differentially expressed genes between populations are genes with known selection signatures corresponding to the out-of-Africa expansion of *D. melanogaster*. Additionally, I find an enrichment of differentially deposited mRNAs between populations on the X chromosome at stage 2, when all transcripts are a product of an XX genotype. In summary, I find that the demographic history of *D.melanogaster* has shaped the maternal and zygotic mRNA complements between populations of the species.

In the second chapter, I present a study on differences in RNA localization along the Drosophila phylogeny. To determine the identity of localized maternal mRNAs, I bisect stage 14 oocytes from five species of *Drosophila* and sequenced mRNAs from the anterior and posterior halves and then performed differential expression analysis. I find several differences in localization between species. I categorize changes in

localization in two categories. First, transcripts that are localized in a subset of species, while having no maternal deposition in others. This indicates lineage specific transcription as well as localization. And second, genes that are maternally deposited in all five species with enrichment to the anterior or posterior in only a subset of these species. I hypothesize that gains and losses of mRNAs could be due to new deposition of transcripts already equipped with cis-regulatory elements able to associate with localization machinery or a gene with pre-existing maternal expression that undergoes changes in regulatory elements that confer the ability to be localized. Overall, this dissertation shows that even within short evolutionary timescales there exists variation in early embryonic gene expression and mRNA localization, both critical to early development.

ACKNOWLEDGMENTS

From the very first year of graduate school, to the completion of this dissertation I have had the support of many. Firstly, I'd like to thank my advisor, Dr. Susan Lott who was instrumental to my research path with her guidance and encouragement through out my projects and the writing process. I'd like to thank my dissertation committee members and graduate advisor, Dr. Celina Juliano, Dr. Artyom Kopp and Dr. Bruce Draper for all of their insightful feedback, discussion and guidance. I would like to thank members of the Lott lab; Emily Cartwright, Dr. Gizem Kalay, Charlie Omura and Sherri Wykoff-Clary as well all past members for their help, guidance and support.

The education and training I received before graduate school was critical for my decision to pursue a PhD. I'd like to acknowledge all of my colleagues in the lab of Dr. Holger Knaut and others in the Developmental Genetics department at NYU for their training and encouragement to pursue graduate school. Additionally, I'd like to acknowledge all those at CUNY Hunter College that initially inspired my passion for science. Completion of this dissertation was made possible with the encouragement and unending support from my parents and family members Alicja, Frank, Paul, Joanne and Swami Feitzinger. I'd like to acknowledge all those friends and loved ones who offered emotional and personal support including Sara Aleman, Sean and Anna Forlenza, Chris Czubay, Denise Chelini, Jonathan Friedman, Tyler Baldwin and Sean Deehan among others.

TABLE OF CONTENTS

ABSTRACT ii

ACKNOWLEDGMENTS v

INTRODUCTION 1

CHAPTER 1 10

Natural variation in the maternal and zygotic mRNA complements of the early embryo in Drosophila melanogaster

Abstract 11

Background 12

Results 15

Discussion 26

Conclusion 30

Methods 32

References 38

Figures 43

CHAPTER 2 52

Evolution of mRNA Localization in the Drosophila Oocyte

Abstract 53

Background 54

Results 62

Discussion 70

Conclusion 74

Methods 75

References 79

Figures 86

Supplemental Material 92

Conclusion 112

INTRODUCTION

A fundamental question in biology is how diverse forms of life can arise during the course of evolution. It has long been suggested that variation in gene expression is a major driver of evolution, even more so than differences in variant forms of proteins [1]. Differences in the regulatory regions of genes drive differential expression which can result in phenotypic diversity [2]. Thus, natural variation of genomes in populations can lead to differences in gene expression which serve as the means in which natural selection can act.

It has been argued that natural selection can act on any stage in the life cycle of multicellular organisms [3]. Development occurs in sequential steps, each of which is dependent on the previous step, in which gene regulatory networks dictate the spatio-temporal dynamics of gene expression [4,5]. In the earliest stages of metazoan development, the zygotic genome is transcriptionally silent, relying entirely on maternal gene products to drive the earliest developmental processes [6,7]. Additionally, in many

species maternal transcripts are required for axis patterning of the developing embryo [8,9]. This process depends on the subcellular localization of maternal RNAs in the oocyte during oogenesis [10,11]. In the fruit fly Drosophila melanogaster, ~75% of the genome is maternally deposited [7]. Maternal genes drive the first nuclear divisions which occur every 8-10 minutes. After the 14th nuclear cycle, these divisions slow dramatically and cellularization of approximately 6,000 nuclei occurs [12]. Genome-wide zygotic expression does not occur until after the 14th nuclear cell division [13]. Thus, these early steps of development require the proper maternal inputs to precisely drive embryogenesis. The handing off of control from the maternal to zygotic genome, termed the maternal to zygotic transition (MZT), is a conserved feature of metazoans [7]. This transition of control requires the orchestration of two entirely different genomes, and thus is a unique and critical process in early development that requires precise regulation.

Natural variation in the maternal and zygotic mRNA complements of the early embryo in *Drosophila melanogaster*

How the MZT, a fundamental transition in development, evolves over time has been a long standing question. Previous work has focused on the evolution of maternal and zygotic mRNA complements along the Drosophila phylogeny [14]. Given that the maternal to zygotic transition is such a critical time in early development, it is expected to be under strong functional constraint. While the maternal and zygotic complements are largely conserved, some evolved differences have recently been identified in both transcriptomes over 50 million years of evolution, including gains and losses in maternal deposition [14].

Over the last century, the demographic history of *Drosophila melanogaster* has been elucidated. *D. melanogaster* has its origins in Sub-Saharan Africa, and approximately 10,000 years ago began an out-of-Africa expansion [15–17]. Coupled with this expansion, *D. melanogaster* experienced a loss of genetic diversity., whichhas been characterized by comparative genomic studies showing that African lines from the native range of *D. melanogaster* have higher genetic diversity than non-African

populations [18]. While a number of studies have focused on the population level natural variation of *D. melanogaster* genomes, there has been a gap of knowledge about how this underlying genetic variation may be correlated with natural variation in gene expression during development. In chapter 1 of this dissertation, I asked the question of how much gene expression variation exists at the critical developmental stages before and after the maternal to zygotic transition within and between populations of the same species. In particular, I determine how the demographic history of a species shapes the maternal and zygotic transcriptomes of present day populations.

Evolution of mRNA Localization in the Drosophila Oocyte

Maternal mRNA localization during oogenesis is a conserved feature of many species. Localized mRNAs in the oocyte serve critical roles in early development such as specifying the body axis and determination of germ cells. In *Drosophila* the accumulation of maternal mRNAs into the oocyte occurs over several stages during oogenesis [11]. Maternal mRNAs are transcribed in support cells, called nurse cells, that are intracellularly connected to the oocyte which is transcriptionally silent during oogenesis [6]. Active transport of RNAs into the oocyte from the nurse cells occurs over

the course of two days in mid-stage oogenesis (stages 2 to 10A)[]. This transport is mediated by *cis*-acting elements, often in the 3'UTR of localized RNAs, that associate with an adaptor protein, Egalitarian (Egl), that links them to the dynein motor protein and the microtubulue cytoskeleton [19,20]. After active transport into the oocyte, the microtubule cytoskeleton undergoes rearrangement and mRNAs are then further localized to the anterior and posterior within the oocyte [21]. Subsequently, nurse cells undergo nurse cell dumping and empty their contents into the oocyte, followed by cytoplasmic streaming which facilitates further entrapment of RNAs in subcellular locations [21]. By stage 14, localized transcripts are anchored to subcellular locations in the oocyte.

Localized mRNAs critical to development, such as the anterior localized transcripts that drive anterior axis patterning, have changed at several points over the course of evolution in Diptera [22]. For instance, one of the most well characterized maternally deposited genes in *Drosophila melanogaster, bicoid*, which acts as the anterior determinant in cyclorrhaphan flies, is the result of a recent gene duplication. A recent study determined that four different anteriorly localized maternal transcript are responsible for anterior fates in four species belonging to different genera of basal

Diptera. It has been hypothesized that gains in localization of maternally deposited transcripts is required for these drastic changes in development to take place. However, how these gains in localization can occur requires examination of localization changes in less evolutionarily diverged species. A systematic study examining localization changes within a genus has not previously been done.

Evidence of the rapid changes in maternal expression on smaller evolutionary timescales comes from experiments done within the Drosophila genus. Single embryo RNA-seq of 14 Drosophila species, spanning 50 million years of divergence, was performed at a stage before and after zygotic genome activation [14]. Analysis of this data has revealed that hundreds of maternally supplied transcripts are differentially expressed between even the most closely related species in this data set. However, the extent and rate of gains or losses in localization of maternal transcripts in the egg between species is unknown. Given the lack of understanding of the frequency of transcript localization changes within a genus, in chapter 2 of this dissertation I characterize changes in anteriorly and posteriorly localized transcripts from the oocytes of five species of Drosophila.

In summary, these two chapters represent my work to investigate how two fundamental processes, gene expression in the early embryo and subcellular

localization in the oocyte, evolve over different timescales. Even within the short timescales represented by populations of the same species, I find considerable variation in the transcripts present in highly critical stages of early development. On the longer timescale represented by species spanning the *Drosophila* genus, I have identified small numbers of dramatic changes in the localization of transcripts in oogenesis.

REFERENCES

1. King M, Wilson a C. Humans and Chimpanze es. Science (80-). 1975;188:107–16.

2. Carroll SB. Perspective Evo-Devo and an Expanding Evolutionary Synthesis : A Genetic Theory of Morphological Evolution. 2008;25–36.

3. Hall BK. Evolutionary Developmental Biology (Evo-Devo): Past, Present, and Future. Evol Educ Outreach. 2012;5:184–93.

4. Briscoe J, Small S. Morphogen rules: Design principles of gradient-mediated embryo patterning. Dev. 2015;142:3996–4009.

5. Koops K, Brakefield PM. - Beldade 2002. Developmental constraints versus. 2002;416.

6. Mische S, Li M, Serr M, Hays TS. Direct Observation of Regulated Ribonucleoprotein Transport Across the Nurse Cell / Oocyte Boundary □. 2007;18:2254–63.

7. Vastenhouw NL, Cao WX, Lipshitz HD. The maternal-to-zygotic transition revisited. Development. 2019;146.

8. Kugler JM, Lasko P. Localization, anchoring and translational control of oskar, gurken, bicoid and nanos mRNA during drosophila oogenesis. Fly (Austin). 2009;3:15–28.

9. Holt CE, Bullock SL. Europe PMC Funders Group Subcellular mRNA Localization in Animal Cells and Why It Matters Mechanisms of mRNA Localization : Illuminating a Multi-Step Process. 2013;326:1212–6.

10. Vazquez-Pianzola P, Suter B. Conservation of the RNA transport machineries and their coupling to translation control across eukaryotes. Comp Funct Genomics. 2012;2012.

11. Lasko P. mRNA localization and translational control in Drosophila oogenesis. Cold Spring Harb Perspect Biol. 2012;4:1–15.

12. Farrell JA, O'Farrell PH. From egg to gastrula: How the cell cycle is remodeled during the drosophila mid-blastula transition. Annu Rev Genet. 2014;48:269–94.

13. Tadros W, Lipshitz HD. The maternal-to-zygotic transition: a play in two acts. Development [Internet]. 2009;136:3033–42. Available from: http://dev.biologists.org/cgi/doi/10.1242/dev.033183

14. Atallah J, Lott SE. Conservation and evolution of maternally deposited and zygotic transcribed mRNAs in the early Drosophila embryo. PLOS Genet [Internet]. 2018;14:1–27. Available from: http://dx.plos.org/10.1371/journal.pgen.1007838

15. Pool JE, Aquadro CF. History and structure of sub-saharan populations of Drosophila melanogaster. Genetics. 2006;174:915–29.

16. Pool JE, Corbett-Detig RB, Sugino RP, Stevens KA, Cardeno CM, Crepeau MW, et al. Population Genomics of Sub-Saharan Drosophila melanogaster: African Diversity and Non-African Admixture. PLoS Genet. 2012;8.

17. Baudry E, Viginier B, Veuille M. Non-African populations of Drosophila melanogaster have a unique origin. Mol Biol Evol. 2004;21:1482–91.

18. Begun DJ, Aquadro CF. African and North American populations of Drosophila melanogaster are very different at the DNA level. Nature [Internet]. 1993;365:548–50. Available from: http://www.nature.com/doifinder/10.1038/365548a0

19. Dienstbier M, Boehl F, Li X, Bullock SL. Egalitarian is a selective RNA-binding protein linking mRNA localization signals to the dynein motor. Genes Dev. 2009;23:1546–58.

20. Vazquez-Pianzola P, Schaller B, Colombo M, Beuchle D, Neuenschwander S, Marcil A, et al. The mRNA transportome of the BicD/Egl transport machinery. RNA Biol [Internet]. Taylor & Francis; 2017;14:73–89. Available from: http://dx.doi.org/10.1080/15476286.2016.1251542

21. Lasko P. Patterning the Drosophila embryo: A paradigm for RNA-based developmental genetic regulation. Wiley Interdiscip Rev RNA. 2020;11:1–19.

22. Yoon Y, Klomp J, Martin-Martin I, Criscione F, Calvo E, Ribeiro J, et al. Embryo polarity in moth flies and mosquitoes relies on distinct old genes with localized transcript isoforms. Elife. 2019;8:1–30.

Chapter 1

Natural variation in the maternal and zygotic mRNA complements of the early embryo in *Drosophila melanogaster*

Anna A. Feitzinger*, Anthony Le, Ammon Thompson, Mehnoor Haseeb, Mohan Koumar Murugesan, Austin M. Tang, Susan E. Lott

Department of Evolution and Ecology, University of California, Davis CA 95616

*Corresponding author: afeitzinger@ucdavis.edu (AAF)

Abstract

Background Maternal gene products supplied to the egg during oogenesis drive the earliest events of development in all metazoans. After the initial stages of embryogenesis, maternal transcripts are degraded as zygotic transcription is activated; this is known as the maternal to zygotic transition (MZT). Recently, it has been shown that the expression of maternal and zygotic transcripts has evolved in the Drosophila genus over the course of 50 million years. However, the extent of natural variation of maternal and zygotic transcripts within a species has yet to be determined. We asked how the maternal and zygotic pools of mRNA vary within and between populations of *D. melanogaster.* In order to maximize sampling of genetic diversity, African lines of *D. melanogaster* originating from Zambia as well as DGRP lines originating from North America were chosen for transcriptomic analysis. **Results** Generally, we find that maternal transcripts are more highly conserved, and zygotic transcripts evolve at a higher rate. We find that there is more within-population variation in transcript abundance than between-populations and that expression variation is highest post-MZT between African lines. **Conclusions** Determining the natural variation of gene expression surrounding the MZT in natural populations of *D. melanogaster* gives insight into the extent of how a tightly regulated process may vary within a species, the extent of developmental constraint on both the maternal and zygotic genomes, and reveals expression changes allowing this species to adapt as it spread across the world.

Background

Over the course of the development of multicellular organisms, an embryo that starts with a single nucleus undergoes divisions with dynamic changes in gene expression to give rise to a functional organism. This can require tight temporal and spatial control of gene expression throughout development, which is complicated by the fact that early development requires the coordination of gene expression across two different genomes. The earliest steps of embryonic development are under complete control of gene products supplied by the maternal genome before developmental control is transferred to the zygote [1]. This process, where control of development is handed off between the maternal and zygotic genomes, is known as the maternal to zygotic transition (MZT) and has been the subject of study of many model organisms [2]. In *Drosophila melanogaster*, maternal RNAs are transcribed during oogenesis in specialized cells called nurse cells and then supplied to the oocyte [3]. During the MZT, these maternal RNAs are degraded as the zygotic genome is activated, ~ 2.5 hours after fertilization [4]. Levels of many transcripts produced by both the maternal and zygotic genomes appear invariant across the MZT, indicating precise coordination of maternal degradation and zygotic transcription [5].

Given the importance of early development to organism survival and its dependence on precise regulation and coordination across the maternal and zygotic genomes, it may be

unsurprising that a previous study found a high level of conservation of transcript levels at these stages across *Drosophila* species [6]. However, the same study [6] also identified changes in transcript representation and abundance across the 50 million years of divergence time of *Drosophila* at both the maternal and zygotic stages and map these changes on the phylogeny. Given that these species have significant differences in the environments in which they develop, some of these changes may be functionally critical to developing under different conditions. Correlations of maternal and zygotic transcript levels decreased with evolutionary divergence, and changes in transcript representation were found even between closely related species [6]. Yet, a significant question remains: do differences in maternal and zygotic transcript levels evolve in the comparatively short evolutionary timescales represented by different populations within a species? Understanding the extent of changes in transcript levels in these critical developmental stages of populations within a species can inform us about the timescale of evolutionary change. Exploring the types of genes that change in the context of different populations may also be a promising avenue for understanding the functions and potential adaptive value of these changes.

In this study, we sought to determine the extent of variation in maternal and zygotic embryonic transcriptomes between populations. To maximize the probability of observing differences, we chose populations of *D. melanogaster* from Africa and North America, as these were likely to be highly genetically diverged. As a species, there is

evidence that *D. melanogaster* has its origins in Sub-Saharan Africa [7,8]. Approximately 10,000 years ago, it is likely that *D. melanogaster* began to expand beyond Sub-Saharan Africa [9,10] and eventually into northern Africa, Asia, and Europe. Only within the past few hundred years were North American populations of *D. melanogaster* founded [11]. With the expansion of *D. melanogaster* out of Sub-Saharan Africa, there was likely a significant loss in genetic diversity [12]. Efforts to sequence genomes from different lines and geographic populations of *D. melanogaster*, including African populations, is ongoing and allows us to understand underlying genetic variation and the demographic history of the species [8]. Taking advantage of the large number of sequenced genomes and RNA sequencing technology, it has more recently become possible to interrogate correlations between genetic variation and transcriptome diversity. For instance, a previous study found that for adult flies, the greater genetic diversity of African populations of *D. melanogaster* did not result in a significantly higher level of gene expression differences within an African population as compared to within a European population [13]. This has brought to light the extent of differential gene expression between these populations within the same species.

Here, we address how the maternal and zygotic transcriptomes controlling the critical processes in early embryogenesis differ between populations of *D. melanogaster*. We performed RNA-Seq on embryos from four lines from Zambia and four lines from North America, from two developmental stages, one stage where all transcripts present are

maternal in origin and the other after zygotic genome activation. Transcript level variation was quantified within two populations as well as putative fixed differences in gene expression between them. We discovered that variation of both maternal and zygotic transcript levels is higher within populations than between populations. We find that there is more expression variation within the Zambia population at both stages relative to the Raleigh population. We observe an enrichment on the X chromosome for maternally deposited mRNAs that are differentially deposited between the two populations. Additionally, we find less transcript level variation between any two of our *D. melanogaster* lines than between species of *Drosophila* ranging 250,000 - 8 million years divergence time. Overall, our results demonstrate that expression level variation at these two stages is consistent with what is known about the differences in genetic variation between these populations. Furthermore, differences in transcript levels at these two stages between populations of *D. melanogaster* recapitulate what is known between species of *Drosophila.*

Results

To investigate the natural variation of RNA levels within a species at stages of embryogenesis controlled by maternal and zygotic genomes, we sequenced embryonic transcriptomes from different *D. melanogaster* populations. Single embryos were collected at a stage in which all RNA has been maternally provided (Bownes' stage 2,

[14]), and another stage after zygotic genome activation (late stage 5; or end of blastoderm stage). To maximize genetic diversity, we chose four lines from Siavonga, Zambia and four Drosophila Genetics Reference Panel (DGRP) [15] lines from Raleigh, North Carolina. Three biological replicates were sequenced per line and stage. An average of 2.83 and 2.89 million high-quality 100 bp paired-end reads were mapped to the *D. melanogaster* genome from the Zambia and Raleigh lines, respectively. Hierarchical clustering of the transcriptomes resulted in samples clustering initially by stage then by population, with the exception of one Raleigh line whose stage 5 sample fell outside the three other stage 5 Raleigh samples (Figure 1A). When we included transcriptomes from an outgroup, *D. simulans,* which share a common ancestor ~2.5 MYA with *D. melanogaster* [16], to the clustering, the *D. simulans* samples clustered by stage with, but outside of, the *D. melanogaster* transcriptomes (Figure 1A). Principal component analysis also separates individual lines by stage with the corresponding principal component (PC1) representing nearly 80% of the variation (Figure 1B).

Expression Variation Differs Within Populations

To explore the patterns of variation in the maternal and zygotic embryonic transcriptomes within and between populations of *D. melanogaster*, we performed differential expression (DE) analysis on our transcriptomic dataset. First, we asked how many genes are differentially expressed within each population, Zambia or Raleigh, at maternal and zygotic stages of development. To do this we implemented a likelihood

ratio test in DESeq2. We normalized our differential expression results to numbers of genes expressed (see Methods) at each stage in order to compare proportions of genes differentially expressed (DE) between stages. We found that overall, there are more DE genes at stage 5 than at stage 2 within both populations (Fig 2A). This is consistent with previous findings between species that zygotic gene expression evolves faster than maternal gene expression [6]. Strikingly, there are many more differentially expressed genes at both stage 2 and stage 5 within the Zambia population than with the Raleigh population.

We asked if there were similarities in the identity of genes with differential expression within populations at the two stages. A proportion of genes were found to be differentially expressed within both populations at stage 2 and stage 5 (Figure 2A). Of all the DE genes at stage 2 combined from both populations, 43% were only DE within the Zambia lines and 28% within only the Raleigh lines, while 29% of genes were DE in both populations (Supplemental Figure S1). At stage 5 the percent of genes only DE within the Zambia lines stayed relatively similar at 39% whereas the percentage of genes only varying expression within the Raleigh population was lower at 20% and the percentage of genes differentially expressed in both was higher at 41% (Supplemental Figure S1). Thus, the percentage of genes varying in expression levels in both populations is higher in stage 5 than stage 2. There is a common set of genes that vary in transcript levels within both populations in addition to a unique set of genes that vary

only within the respective populations, and these vary by stage, with more shared differences at stage 5.

Differences in the Magnitude of Expression Variation Within Populations

With more genes differentially expressed within the Zambia population than the Raleigh population, we asked if the magnitude of expression changes were similar between populations. To do this, we found the maximum and minimum expression value for each differentially expressed gene within the populations. From this, we computed the log ratio of the fold change for each DE gene. We then asked if the distribution of the log ratio of fold changes for DE genes were different between the two populations at either stage (Figure 2C). There is no significant difference between the means of log ratio of fold changes when comparing stage 2 between populations (t-test, $p = 0.9109$), thus there is no evidence that the magnitude of transcript abundance changes is different between populations. There is, however, a significant difference between the means of the log ratio of fold changes between the two populations at stage 5 (t-test, $p= 7.278e-06$) with a higher magnitude of fold changes within the Raleigh population. Therefore, although there are fewer genes differentially expressed within the Raleigh population at stage 5, the magnitude of these differences is on average higher than the genes differentially expressed within the Zambia population at this stage.

More Differences within Populations Than Between Populations at Maternal and Zygotic Stages

Next, we asked if there were fixed expression differences between the populations. We define fixed expression differences as genes that are on average higher, or lower, in one population than the other (i.e. have similar levels in all lines from a population, that are significantly different than all the lines in the other population; see Figure 3A for examples). We used the Raleigh lines and the Zambia lines as replicates in DE analysis. Similar to the expression variation within populations, the percentage of genes that were differentially expressed between populations increased from stage 2 to stage 5 (Figure 2A). We find that there are more genes differentially expressed within populations than fixed expression differences between the populations at both stages (Figure 2A).

In addition to finding fixed expression differences, we asked how many genes were differentially expressed between individual lines. Genes differentially expressed between lines from different populations in the pairwise analysis represent differences only between the two lines in the comparison, rather than fixed expression differences between the two populations as in the previous analysis. This resulted in DE analysis between every pair of lines resulting in 28 of total comparisons. 12 DE tests between lines of the same population (RR and ZZ), and 16 DE tests between lines of different populations (RZ). Since there are fewer tests between lines of the same population than

between lines of different populations, we used bootstrapping in order to compare the average number of DE genes between these categories. Similar to the previous within population analysis, there are fewer DE genes between individual Raleigh lines (RR) than Zambia lines (ZZ), at both stages (Figure 2B). Interestingly, we find that the average pairwise differences between lines (RZ) of different populations at stage 2 was not significantly different ($p = 0.06972$; Wilcoxon rank sum test) than the average pairwise differences between Zambia lines (ZZ) at this stage (Figure 2B). However, at stage 5, the average number of differences between lines of different populations are higher relative to the number of differences between Zambia lines ($p < 2.2e-16$; Wilcoxon rank sum test). Therefore, there is as much variation of expression between individual Zambia lines at stage 2 as between individual lines from different populations at this stage. In contrast, variation between individual lines from different populations at stage 5 surpasses the differences between individual Zambia lines at this stage.

More Expression Variation Between than Within Species

Expanding our analysis, we investigated gene expression variation within and between species of *Drosophila* at maternal and zygotic stages. In a previous study, we generated RNA-seq data from *D. simulans, D. sechellia, D. yakuba,* and *D. erecta* from stage 2 and stage 5 embryos using the same single embryo RNA extraction method implemented here. We chose these two pairs of sister species as they are closely

related, but one pair (*D. simulans* and *D. sechellia*) diverged more recently (~250,000 years ago, [16] than the other pair (*D. yakuba* and *D. erecta,* estimated 8 MYA divergence time [17,18]. RNAseq reads from these species were processed identically to the *D. melanogaster* reads for this analysis (see Methods). Genes considered in this analysis were limited to one-to-one orthologs across the 5 species, a total of 12,110 genes. As we had only one line for each of the other species, we performed the DE analysis pairwise for each of our *D. melanogaster* lines, as well as between each pair of sister species. The number of DE genes in each population or species was normalized to the number of genes transcribed at each stage to compare the percentage of DE genes at both stages and across species. From these comparisons, within and between species there are more DE genes at stage 5 than stage 2 (Figure 4). For maternal genes, the more closely related species pair *D. simulans* and *D. sechellia* have the highest proportion of DE genes. While most *D. melanogaster* lines have fewer differences than either of the species comparisons at this stage, two of the Raleigh vs. Zambia comparisons have as high of a proportion of their maternal genome differentially expressed as the more distantly related species pair, *D. yakuba* and *D. erecta*. For stage 5, both species pairs have a larger proportion of their transcripts differentially expressed than any of the within-species comparisons of *D. melanogaster.* Both stages have, on average, fewer genes differentially expressed for within-species comparisons than between species, but this pattern is much stronger for stage 5, a stage with more DE genes in all comparisons.

Enrichment of DE Genes at Maternal Stage on the X Chromosome

Before the zygotic genome is activated, embryonic development is entirely under control of maternal gene products. Therefore, all stage 2 transcriptomes are supplied entirely by XX genomes and the zygotic genome is transcribed by either XX or XY genomes. Given the possibility of different evolutionary pressures, we asked whether there is a difference of enrichment of DE genes on the autosomes or X chromosome across maternal and zygotic stages. For our stage 5 transcriptomes, these were collected from XY embryos, so they are directly comparable. As onset of Drosophila dosage compensation [19,20] occurs sometime after stage 5 [21], collecting a single sex is necessary at this stage. We normalized the number of DE genes per chromosome by the number of genes expressed on each chromosome.

Interestingly, we found that DE genes at stage 2 between populations were enriched on the X chromosome compared to the autosomes (Figure 3B). However, enrichment of DE genes on the X chromosome is absent at stage 5 in our samples. Maternal transcripts are not completely degraded by stage 5, so we also asked if the trend seen for all of stage 5 transcripts were the same for transcripts that are zygotic only. As expected, fixed expression differences between zygotic-only genes were not enriched on the X chromosome (Fisher's exact test, $p < 0.05$) having the same result as all genes at stage 5.

The most differentially expressed genes have known selection signatures

Several of the most differentially deposited transcripts between populations are genes that have signatures of selection at the level of the genome under different conditions. For example, a previous study found that genes within the chemosensory system have undergone local adaptation following *D. melanogaster*'s global expansion out of Africa [22]. This study was based on the genomes of five different geographically distinct populations of *D. melanogaster* including both North American and African populations. Notable within the top ten most DE maternally deposited genes between populations is Gstd9, a glutothione-S-transferase, which belongs to a gene family that was found to have signals of selective sweeps upon global expansion [22]. In total, seven glutathione-S-transferases were found to be differentially deposited between the Raleigh and Zambia populations. In the same study [22] the zinc finger protein family was shown to have strong population differentiation. Zcchc7, a zinc-finger protein, is also among the top ten most differentially deposited transcripts. These two genes both have undergone dramatic qualitative changes in maternal deposition (Figure 3A and Figure 6A).

The second most significantly differentially deposited transcript is the actin binding protein Unc-115a. The paralog of this gene, Unc-115b, was also found to be differentially expressed between populations. Both genes have higher expression levels in the Raleigh population. Interestingly, Unc-115b was found in a previous study to be

the most highly upregulated gene in a *D. melanogaster* strain resistant to the insecticide DDT 91-R compared to a DDT compromised strain, 91-C [23]. Unc-11b was one of two genes found in this study to be highly upregulated across all stages of development that were assayed [23]. This gene was found to be in one of six selective sweeps that coincided with constitutive expression differences between DDT resistant and compromised lines.

Several of the most differentially expressed genes are annotated as pseudogenes

The most differentially maternally deposited gene between the Zambia and Raleigh populations in our analysis is the gene CR40354 which is annotated in the *D. melanogaster* genome as a pseudogene with unknown function. This prompted us to investigate other genes annotated as pseudogenes in our dataset because previous annotations that identified these genes as pseudogenes were more likely to have been done in non-African populations. We asked how many pseudogenes were maternally deposited and zygotically expressed within and between populations. A total of 69 and 70 genes labeled as pseudogenes were found to be maternally deposited within the Raleigh and Zambia populations, respectively. A total of 16 and 8 genes labeled as pseudogenes were found to be expressed from the zygotic genome but not the maternal genome (zygotic-only, see Methods) in the Raleigh and Zambia populations. Between the populations, 18 pseudogenes were found to be differentially maternally deposited and 16 of the zygotic-only pseudogenes were found to be differentially

expressed at stage 5. One pseudogene which caught our attention was the *swallow Ψ* (*swaΨ*) pseudogene which is differentially expressed within the Zambia population in our analysis. *swaΨ* is a result of a recent genome duplication of *swallow* and is only found in *D. melanogaster* [24]. *swallow* is a critical gene to early development and is required for proper Bicoid positioning in the embryo [25]. Previous studies [26] have suggested that *swaΨ* not transcribed in *D. melanogaster*. We found it to be very lowly expressed in the Raleigh lines, but variably expressed within the Zambia lines with one line, ZI160, showing relatively high expression levels (Figure 5D). To investigate further, we sequenced the swaΨ locus in each of the lines. We discovered a 15bp population-specific deletion present. All Raleigh lines have a 15bp deletion in the annotated exon 3 of *swaΨ*, which is not present in all four Zambian lines. This sequence is part of the fully functional exon 3 of the *swallow* gene.

Variation in Heat shock Proteins

Modifying maternal RNAs and proteins in the embryo can have effects on development, phenotypes and ultimately fitness [27,28]. One gene family that is critical to survival is heat shock proteins [29,30]. In total, 17 and 19 heat shock proteins were found to be differentially deposited within both the Raleigh population and Zambia population, respectively. By contrast, 6 and 8 zygotic-only heat shock proteins were found to be differentially expressed from the zygotic genome within the two populations. Previous work by Lockwood et al. has shown evidence that higher levels of maternal deposition

of a heat shock protein increases embryo thermal tolerance in *D. melanogaster* [31]. Interestingly, Hsp23 was found to be differentially deposited in the lines that we examined (Figure 6A, bottom panel). Specifically, the levels of Hsp23 mRNA in ZI094 is between 4-14X higher than the other three Zambia lines and 11-600X higher levels than the Raleigh lines, all which have variable expression. This overall trend persists at stage 5, with mean levels of Hsp23 increasing in ZI094 and maintaining higher expression levels compared to all other lines. Based on this observation, we performed heat shock experiments on all lines to assay differences in embryo survival after heat stress (see Methods). In general, we found that heat shock tolerance does not correspond in a predictive way with levels of heat shock transcripts (Figure 6B).

Discussion

While previous studies have shown that the maternal and early zygotic transcriptomes are highly conserved across species [6,32,33], here we show that that there is variation present in gene expression on the shorter evolutionary timescale represented within a species, *D. melanogaster*. We chose lines from Siavonga, Zambia and Raleigh, North Carolina, USA to encompass a broad span of genetic diversity within and among populations.

Our results show that the transcriptomic dynamics at these developmental stages reflect what is known about the population genetic history of *D. melanogaster* from genomic studies. Previous studies found more genetic variation within African populations than non-African populations [7,8,34,35] and we found the same pattern with the maternal and early zygotic transcriptomes. There are differential transcript abundances within both the Zambia and Raleigh populations, and some of the same transcripts are variable within each population, but there is more population-specific variation within the Zambia lines. We also find that with pairwise comparisons between lines, the Raleigh lines have far fewer genes identified as differentially expressed, but comparisons within Zambia have as many (stage 2) or only slightly fewer (stage 5) differentially expressed genes as when comparing lines from the two populations. The increased number of differentially expressed genes in the Zambia lines is consistent with high levels of genomic variation found in the ancestral range of this species [35]. And the reduced number of differentially expressed genes in Raleigh likely reflects the lower genetic polymorphism levels following the out-of-Africa bottleneck [7,36]. Interestingly, while consistent with the genomic variation within these lines, our results stand in contrast to microarray studies in adults which found less transcript variation within African and non-African populations than between, which has been taken as a sign of directional selection [13,37].

Also consistent with previous genomic studies are the numbers of genes highlighted by our DE analysis that have also been identified in studies performing artificial selection or population genomic studies on the global expansion of the species [22,23]. Many are used as examples throughout the manuscript and have been associated with xenobiotic metabolism (GstD9, Cyp12d1-p), possible environmental adaptation to global expansion (Zcchc7), and DDT resistance (Unc-115a, Unc-115b). Thus, many of our most significantly DE genes are also likely under selection, and their functions are consistent with adaptation to a new environment. Studies to determine the adaptive function of these genes are often carried out in adults [38,39] but our data suggests that these differences in transcript level are also present in the embryo, and thus may potentially be of adaptive value at this stage.

We find a stage 2 specific enrichment of differentially expressed genes between the Zambia and Raleigh populations on the X chromosome. Previous studies have shown a reduction in heterozygosity on the X chromosome relative to the autosomes in temperate European populations compared to populations from sub-Saharan Africa. This reduction in heterozygosity has been attributed to demographic events following the out-of-Africa expansion of *D. melanogaster* [12]. Therefore, it is possible that the decreased heterozygosity on the X chromosome has led to decreased differences in transcript levels of genes on the X within the Raleigh populations. This decrease in expression variation within Raleigh may contribute to the strong signal between

population differences in expression we find specifically on this chromosome. However, this pattern of enrichment is only seen at stage 2, where all transcripts are from the maternal, XX, genome, and therefore may be under unique selective pressures.

Genes annotated as pseudogenes were called significantly differentially expressed in our analysis both within and between populations. Most striking is the fixed expression difference of the *swallow pseudogene* (*swaΨ)* between populations at stage 2. *swaΨ* is the result of a relatively recent duplication of the *swallow* gene which is maternally expressed and required for proper anterior-posterior axis patterning. Genome-wide analysis of pseudogenes in *D. melanogaster* has shown that *D. melanogaster* have relatively low proportion of pseudogenes (110 were identified in one study [40]), with respect to their proteome, compared to other eukaryotic genomes such as human, nematode, and budding yeast [40]. It has been speculated that the low number pseudogenes suggests a high rate of DNA loss in *Drosophila* [26]. Here, we find that *swaΨ* has most likely acquired a 15bp deletion after the migration of *D. melanogaster* out of Africa. Interestingly, we also find that *swaΨ* is expressed in a number of the Zambia lines but very low to no expression was detected in the Raleigh lines. These data suggest that in addition to deletions *swaΨ* has also lost maternal expression over time.

Among the genes we found to be differentially expressed were heat shock proteins, including Hsp23. Previous work demonstrated that maternally loaded Hsp 23 increases

embryo heat tolerance when maternally loaded [31]. Here, we adapted the same heat shock and embryo lethality protocol to determine differences in thermotolerance between lines at stage 2. We did not find a linear relationship between thermotolerance and maternal Hsp23 levels at varying temperatures of heat shock. As the previous study [31] was overexpressing Hsp23 while leaving the levels of other heat shock proteins unaffected, natural variation in other genes that affect thermotolerance may explain the differences in our results. In fact, we have found differential expression in over 30 heat shock proteins at stage 2 within the two populations as well as glutathione s-transferases which both have shown to have roles in thermotolerance [41]. It is possible that more complex interactions among the genes in these networks underlie the patterns of thermotolerance we find in these lines across temperatures.

Conclusions

Previous studies have found a high degree of conservation of the maternal transcriptome across species [6,32,33]; this study provides evidence this is also true within *D. melanogaster*. Whether examining the number or proportion of differentially expressed genes within populations, between populations, between pairs of lines, or between species, there are fewer differences in transcript levels found at stage 2, when all transcripts are maternal, than at stage 5, after zygotic genome activation. The

analysis of proportions of genes DE within and between species is especially suggestive relative to these stage-specific dynamics. At stage 5, the proportion of genes DE between species is far higher than the within-*D. melanogaster* comparisons, and there is a higher proportion of DE genes overall in every comparison. In contrast, at stage 2, there are fewer genes DE in each comparison, and the between species comparisons (while still higher on average than the within-*D. melanogaster* comparisons) are only slightly higher. This suggests that relative to one another, more of the maternal transcriptome may be under stabilizing selection than the more rapidly evolving zygotic stage transcriptome [42].

In conclusion, we find that the maternal and zygotic transcriptomes, while generally conserved, do show some interesting differences in transcript abundance even in the relatively short period of evolutionary time represented by the diversity within a species. This species, *D. melanogaster,* has more variation in transcript abundance at these critical developmental stages within populations than between them. And consistent with what has been determined between Drosophila species [6], we show that the maternal transcriptome is more highly conserved than the zygotic transcriptome, and more of the maternal genome may be under purifying selection. Together, the presented data highlight how a constrained developmental trait evolves over short periods of evolutionary time.

Methods

Embryo collection and sequencing library generation

Fly populations from Siavonga, Zambia (courtesy of the Langley Lab, University of California, Davis) and Raleigh, North Carolina, USA (the DGRP lines; [15] were population controlled on cornmeal fly food at 25 degrees C. Four lines from Zambia (ZI050, ZI094, ZI160, ZI470) and four lines from Raleigh (RAL307, RAL357, RAL360, RAL517) were selected for embryo collection. Embryos were dechorionated using 50% bleach, and imaged on a Zeiss Axioimager, under halocarbon oil, to determine stage. Since embryos were collected from a large number of mothers, it is unlikely that multiple samples came from the same mother. Stage 2 and late stage 5 embryos were identified based on morphology. Stage 2 embryos were selected based on the vitelline membrane retracting from both the anterior and posterior poles, prior to when pole cells become visible. Late stage 5 embryos were chosen based on having completed cellularization, but not yet having started gastrulation. Embryos were then removed from the slide with a brush, cleaned of excess oil, placed into a drop of Trizol reagent (Ambion), and ruptured with a needle, then moved to a tube with more Trizol to be frozen at -80˚C until extraction. RNA and DNA were extracted as in the manufacturer's protocol, with the exception of extracting in an excess of reagent (1 mL was used) compared to expected mRNA and DNA concentration. Extracted DNA for stage 5 embryos was used for genotyping for sex as in Lott et al, 2011, XY embryos were selected for transcriptomic

analysis, due to the incomplete nature of X chromosomal dosage compensation in XX embryos at this stage [5].

RNA-Seq libraries were prepared using poly-A enrichment for each of the 8 lines (4 Zambia lines and 4 Raleigh lines), for both stage 2 and stage 5, with 3 replicates each, for a total of 48 libraries. These samples were sequenced 100bp, paired-end, on an Illumina HiSeq4000. The sequencing was carried out by the DNA Technologies and Expression Analysis Core at the UC Davis Genome Center, supported by NIH Shared Instrumentation Grant 1S10OD010786-01.

Data Processing

Reads were trimmed and adapters removed using Cutadapt [43], and gently (PHRED Q < 20) trimmed for quality [44]. Mapping was done with the D. melanogaster Flybase genome release 6.18 and associated annotation file using hisat2 version 2.1.0 [45] using default parameters. Gene level counts were generated using featureCounts of the subRead [46] package in R [47](R version 3.4.1). Counts were normalized to sequencing depth and RNA composition using DEseq2's median of ratios. Count data can be found in Supplemental File 1.

Data availability

All raw and processed data are available at NCBI/GEO under an accession number. Processed data (transcript level counts) is also available in Supplemental File 1.

Hierarchical Clustering and PCA Analysis

We performed hierarchical clustering analysis in R using the hclust function. A dissimilarity matrix (*dist()*) of one minus the Spearman correlation (*cor()*) was used for hierarchical clustering. Principal component analysis (PCA) was also performed in R using the *prcomp()* function.

Determining on or off State

To determine whether a gene was likely to be transcribed based on the count data, we ran Zigzag [48] on our data. A full description of how this program was utilized, see Supplemental File 2.

Differential Expression Analysis

Differential expression analysis was done using the DEseq2 [49] package in R. Using DEseq2, we implemented the LRT (likelihood ratio test). For within- population analysis the replicates for each line were given the same label for the design matrix. For determining the differences between populations, we labeled lines as either Raleigh or Zambia in the design matrix and implemented the LRT test. When comparing the

number of DE genes within and between populations, the number of DE genes is divided by the number of genes expressed in order to compare % DE genes between stages. We counted a gene as expressed in the total number of genes expressed for normalization if the gene was expressed in at least one line, as described above.

For pairwise differences between lines, DEseq2 was run on every possible combination of pairs. Since there are more between population pairs than within population pairs, we ran bootstrapping in R in order to compare the number of DE genes between lines of the same population and between lines of different populations. To test if the distributions of bootstrapped averages were significantly different from one another, we implemented a Wilcoxon rank sum test in R.

For differential expression analysis between species we used RNA-seq data previously generated in the lab [6] from *D. simulans, D. sechellia, D. erecta,* and *D. yakuba.* Reads were aligned using HISAT2 followed by FeatureCounts to generate expression levels in counts. Counts were then normalized using the *norm()* function in DESeq2. Only genes which had orthologs in all seven species were considered. An expression cut off of 3 counts was used to determine which genes were considered expressed in each line.

Test of Enrichment on Autosomes or Sex Chromosomes

To determine whether there was enrichment of DE genes on either the autosomes or sex chromosomes the chromosomal location of each DE gene was determined. Number

of DE genes per chromosome was normalized to the number of genes expressed on the chromosome. We implemented a Fisher's exact test in R to determine if there is a significant difference in how many DE genes are on autosomes compared to the X chromosome. This was performed by doing individual tests between the number of DE genes on each autosome and the X.

Heat Shock of Embryos

We adapted the heat shock and embryo survival protocols from [31]. Flies aged 3-5 days were allowed to lay on a clearance plate for one hour. Plates were then swapped with clear agar collection plates with additional yeast and flies allowed to lay for an additional hour in order to collect 0-1 hour aged embryos. Plates were then wrapped in parafilm and fully submerged in a heat bath at the given temperature for 40 minutes. Embryos were then grouped in a line of 20 embryos using a brush. Proportion of embryos hatched was assayed 48 hours after heat shock to determine embryo survival. Three temperatures were assayed.

Authors' contributions

S.E.L. and A.A.F designed the study. A.L., M.H, A.M. and M.K.M collected the data. A.T. contributed analysis for determining on and off state. A.A.F. analyzed the data. A.A.F. and S.E.L. wrote the manuscript.

Acknowledgements

We would like to thank members of the Lott Lab for discussion, especially Emily Cartwright and Charles Omura for reading and comments on the analysis and manuscript, and Sherri Wykoff-Clary for her management and fly care. We thank the UC Davis fly community, including the Superfly group, for their contributions.

References

1. Tadros W, Lipshitz HD. The maternal-to-zygotic transition: a play in two acts. Development [Internet]. 2009;136:3033–42. Available from: http://dev.biologists.org/cgi/doi/10.1242/dev.033183

2. Vastenhouw NL, Cao WX, Lipshitz HD. The maternal-to-zygotic transition revisited. Development. 2019;146.

3. Lasko P. mRNA localization and translational control in Drosophila oogenesis. Cold Spring Harb Perspect Biol. 2012;4:1–15.

4. Bashirullah A, Halsell SR, Cooperstock RL, Kloc M, Karaiskakis A, Fisher WW, et al. Joint action of two RNA degradation pathways controls the timing of maternal transcript elimination at the midblastula transition in Drosophila melanogaster. EMBO J. 1999;18:2610–20.

5. Lott SE, Villalta JE, Schroth GP, Luo S, Tonkin LA, Eisen MB. Noncanonical compensation of zygotic X transcription in early Drosophila melanogaster development revealed through single-embryo RNA-Seq. PLoS Biol. 2011;9.

6. Atallah J, Lott SE. Conservation and evolution of maternally deposited and zygotic transcribed mRNAs in the early Drosophila embryo. PLOS Genet [Internet]. 2018;14:1–27. Available from: http://dx.plos.org/10.1371/journal.pgen.1007838

7. Begun DJ, Aquadro CF. African and North American populations of Drosophila melanogaster are very different at the DNA level. Nature [Internet]. 1993;365:548–50. Available from: http://www.nature.com/doifinder/10.1038/365548a0

8. Pool JE, Aquadro CF. History and structure of sub-saharan populations of Drosophila melanogaster. Genetics. 2006;174:915–29.

9. Li H, Stephan W. Inferring the demographic history and rate of adaptive substitution in Drosophila. PLoS Genet. 2006;2:1580–9.

10. Thornton K, Andolfatto P. Approximate Bayesian inference reveals evidence for a recent, severe bottleneck in a Netherlands population of Drosophila melanogaster. Genetics. 2006;172:1607–19.

11. Brennecke J, Aravin AA, Stark A, Dus M, Kellis M, Sachidanandam R, et al. Discrete Small RNA-Generating Loci as Master Regulators of Transposon Activity in Drosophila. Cell. 2007;128:1089–103.

12. Kauer M, Dieringer D, Schlötterer C. Nonneutral admixture of immigrant genotypes in African Drosophila melanogaster populations from Zimbabwe. Mol Biol Evol. 2003;20:1329–37.

13. Hutter S, Saminadin-peter SS, Stephan W, Parsch J. Gene expression variation in African and European populations of Drosophila melanogaster. 2008;9:1–15.

14. Bownes M. Patterning and morphogenesis of the follicle cell epithelium during Drosophila oogenesis. 1998;552:541–52.

15. MacKay TFC, Richards S, Stone EA, Barbadilla A, Ayroles JF, Zhu D, et al. The Drosophila melanogaster Genetic Reference Panel. Nature. 2012;482:173–8.

16. McDermott SR, Kliman RM. Estimation of isolation times of the island species in the Drosophila simulans complex from multilocus DNA sequence data. PLoS One. 2008;3.

17. Ometto L, Cestaro A, Ramasamy S, Grassi A, Revadi S, Siozios S, et al. Linking genomics and ecology to investigate the complex evolution of an invasive Drosophila pest. Genome Biol Evol. 2013;5:745–57.

18. Wheat CW, Wahlberg N. Phylogenomic insights into the cambrian explosion, the colonization of land and the evolution of flight in Arthropoda. Syst Biol. 2013;62:93–109.

19. Lucchesi JC, Kuroda MI. Dosage compensation in drosophila. Cold Spring Harb Perspect Biol. 2015;7:1–21.

20. Conrad T, Akhtar A. Dosage compensation in Drosophila melanogaster: Epigenetic fine-tuning of chromosome-wide transcription. Nat Rev Genet. Nature Publishing Group; 2012;13:123–34.

21. Georgiev P, Chlamydas S, Akhtar A. Drosophila dosage compensation Males are from Mars, females are from Venus. Fly (Austin). 2011;5:147–54.

22. Arguello JR, Cardoso-moreira M, Grenier JK, Gottipati S, Clark AG, Benton R. Extensive local adaptation within the chemosensory system following Drosophila melanogaster ’ s global expansion. Nat Commun [Internet]. Nature Publishing Group; 2016;7:1–12. Available from: http://dx.doi.org/10.1038/ncomms11855

23. Seong KM, Coates BS, Sun W, Clark JM, Pittendrigh BR. Changes in Neuronal Signaling and Cell Stress Response Pathways are Associated with a Multigenic Response of Drosophila melanogaster to DDT Selection. 2017;9:3356–72.

24. Chao Y -C, Donahue KM, Pokrywka NJ, Stephenson EC. Sequence of swallow, a gene required for the localization of bicoid message in Drosophila eggs. Dev Genet. 1991;12:333–41.

25. Stephenson EC, Chao YC, Fackenthal JD. Molecular analysis of the swallow gene of Drosophila melanogaster. Genes Dev. 1988;2:1655–65.

26. Petrov DA, Chao YC, Stephenson EC, Hartl DL. Pseudogene evolution in Drosophila suggests a high rate of DNA loss [1]. Mol Biol Evol. 1998;15:1562–7.

27. Driever W, Nüsslein-Volhard C. The bicoid protein determines position in the Drosophila embryo in a concentration-dependent manner. Cell. 1988;

28. Zhang J, Houston DW, King M Lou, Payne C, Wylie C, Heasman J. The role of maternal VegT in establishing the primary germ layers in Xenopus embryos. Cell. 1998;94:515–24.

29. Feder ME, Hofmann GE. Heat-shock proteins, molecular chaperones, and the stress response: Evolutionary and ecological physiology. Annu Rev Physiol. 1999;61:243–82.

30. Kregal K. Highlighted Topics. J Appl Physiol [Internet]. 2001;90:1593–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11007556

31. Lockwood BL, Julick CR, Montooth KL. Maternal loading of a small heat shock protein increases embryo thermal tolerance in Drosophila melanogaster. 2017;4492–501.

32. Preuss KM, Lopez JA, Colbourne JK, Wade MJ. Identification of maternally-loaded RNA transcripts in unfertilized eggs of Tribolium castaneum. BMC Genomics [Internet]. BMC Genomics; 2012;13:1. Available from: BMC Genomics

33. Heyn P, Kircher M, Dahl A, Kelso J, Tomancak P, Kalinka AT, et al. The earliest transcribed zygotic genes are short, newly evolved, and different across species. Cell Rep [Internet]. The Authors; 2014;6:285–92. Available from: http://dx.doi.org/10.1016/j.celrep.2013.12.030

34. Langley CH, Stevens K, Cardeno C, Lee YCG, Schrider DR, Pool JE, et al. Genomic variation in natural populations of *Drosophila melanogaster*. Genetics. 2012;192:533.

35. Baudry E, Viginier B, Veuille M. Non-African populations of Drosophila melanogaster have a unique origin. Mol Biol Evol. 2004;21:1482–91.

36. David JR, Capy P. Genetic variation of Drosophila melanogaster natural populations. Trends Genet. 1988;4:106–11.

37. Müller L, Hutter S, Stamboliyska R, Saminadin-Peter SS, Stephan W, Parsch J. Population transcriptomics of Drosophila melanogaster females. BMC Genomics. 2011;12:15–7.

38. Yang J, McCart C, Woods DJ, Terhzaz S, Greenwood KG, Ffrench-Constant RH, et al. A Drosophila systems approach to xenobiotic metabolism. Physiol Genomics. 2007;30:223–31.

39. Strycharz JP, Lao A, Li H, Qiu X, Hyeock S, Sun W, et al. Resistance in the highly DDT-resistant 91-R strain of Drosophila melanogaster involves decreased penetration , increased metabolism , and direct excretion. Pestic Biochem Physiol [Internet]. Elsevier Inc.; 2013;107:207–17. Available from: http://dx.doi.org/10.1016/j.pestbp.2013.06.010

40. Harrison PM, Milburn D, Zhang Z, Bertone P, Gerstein M. Identification of pseudogenes in the Drosophila melanogaster genome. Nucleic Acids Res. 2003;31:1033–7.

41. Riveron J, Boto T, Alcorta E. Transcriptional basis of the acclimation to high environmental temperature at the olfactory receptor organs of Drosophila melanogaster. BMC Genomics. 2013;14.

42. Nuzhdin S V., Wayne ML, Harmon KL, McIntyre LM. Common pattern of evolution of gene expression level and protein sequence in Drosophila. Mol Biol Evol. 2004;21:1308–17.

43. Wang J, Yin Y, Lau S, Sankaran J, Rothenberg E, Wohland T, et al. Anosmin1 Shuttles Fgf to Facilitate Its Diffusion, Increase Its Local Concentration, and Induce Sensory Organs. Dev Cell [Internet]. Elsevier Inc.; 2018;46:751-766.e12. Available from: https://doi.org/10.1016/j.devcel.2018.07.015

44. MacManes MD. On the optimal trimming of high-throughput mRNA sequence data. Front Genet. 2014;5:1–7.

45. Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. Nat Protoc [Internet]. Nature Publishing Group; 2016;11:1650–67. Available from: http://dx.doi.org/10.1038/nprot.2016-095

46. Liao Y, Smyth GK, Shi W. FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2014;30:923–30.

47. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: Ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29:15–21.

48. Thompson A, May MR, Moore BR, Kopp A. A hierarchical Bayesian mixture model for inferring the expression state of genes in transcriptomes. Proc Natl Acad Sci U S A. 2020;117:19339–46.

49. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15:1–21.

Figures

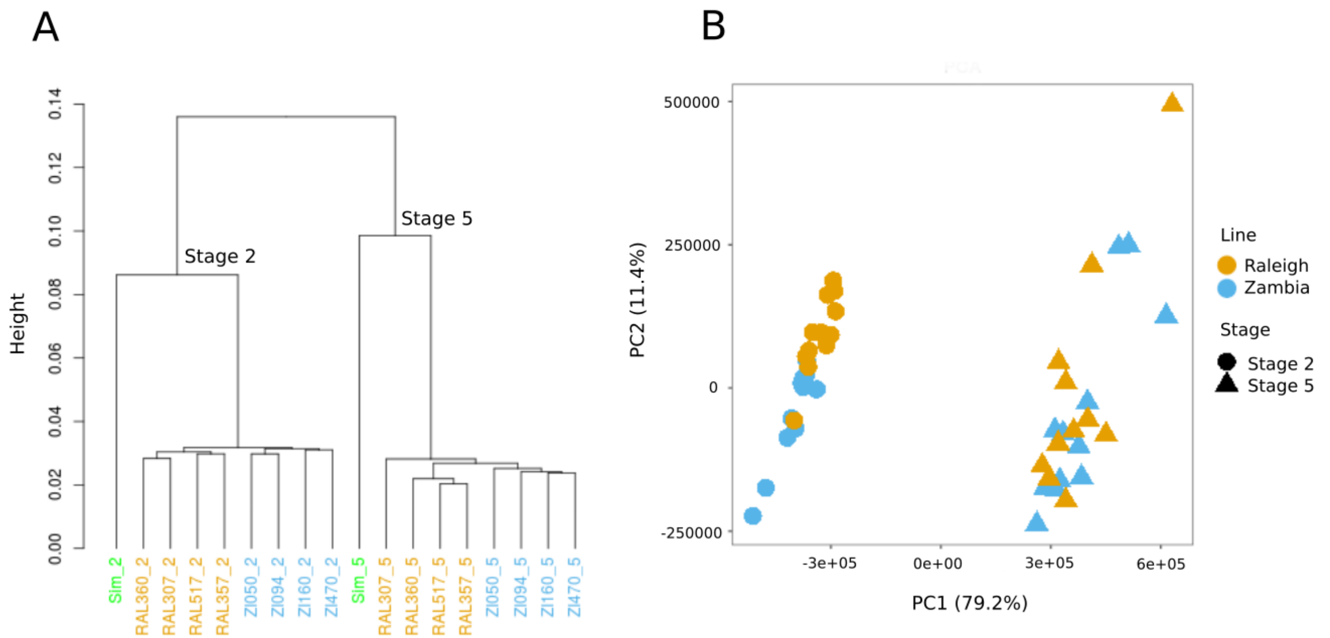


Figure 1. Populations are distinct at each developmental stage. A) Hierarchical clustering of transcriptomes from stage 2 (labels ending with _2) and stage 5 (labels ending with _5) embryos, from 8 lines of *D. melanogaster*, four from Raleigh (RAL, orange) and four from Zambia (ZI, blue), with closely related species *D. simulans* (Sim, in green text) as an outgroup. Samples cluster first by stage, then by species, then by population. B) PCA shows that these same samples separate first by stage (PC1, which explains a large proportion of the variance at 79.2%), then by population (PC2, 11.4% of the variance), though more distinctly at stage 2 than stage 5.

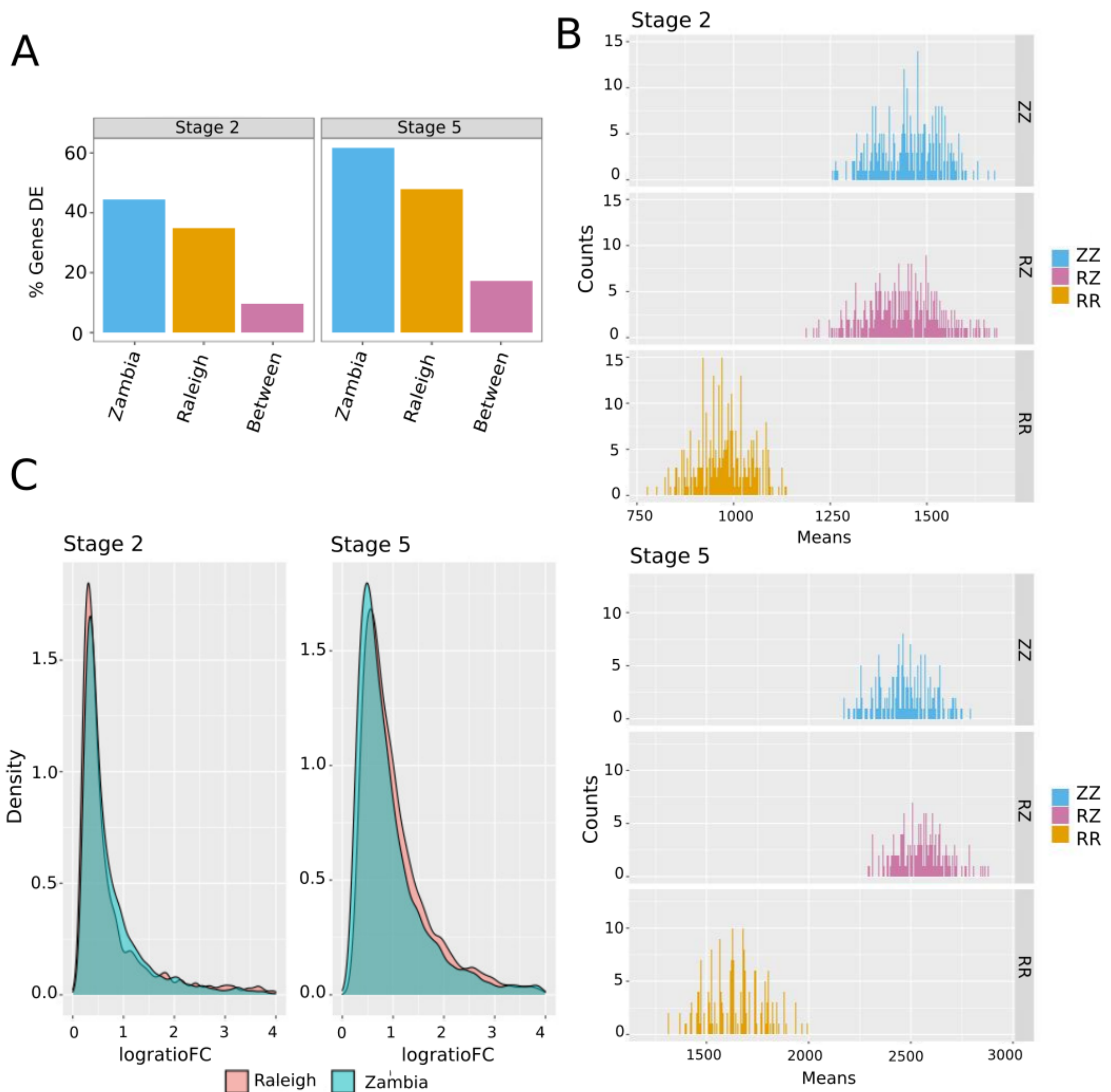


Figure 2. Differential expression within and between populations (Higher number of differential expression within Zambia. Larger magnitude of changes within Raleigh at Stage 5) A) Percent of genes differentially expressed within and between the Zambia and Raleigh populations at stage 2 and stage 5. More differences are found within populations (blue, orange) than between populations (pink). (B) To control for the

number of comparisons within and between lines, we also examined pairwise differences between lines at each stage (stage 2 top, stage 5 bottom). When compared in this way, at both stages, the distributions of DE genes within the Zambia population and between the Zambia and Raleigh populations are similar, with fewer DE genes within the Raleigh lines. (C) Distributions of the magnitudes of differences in expression in DE genes, which shows that the magnitude of changes between differentially expressed genes is greater within the Raleigh population at stage 5 than the Zambia population at this stage.

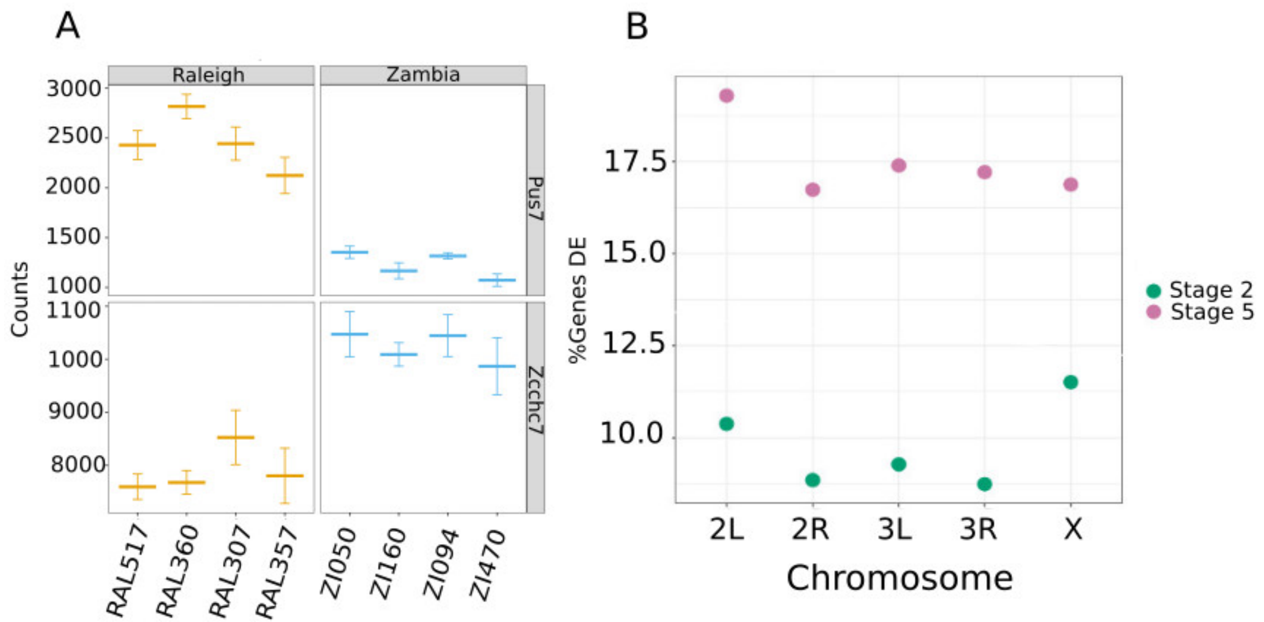


Figure 3. Examination of putative fixed differences between populations. A) Expression levels in counts for two example genes, showing what we categorize as fixed differences in transcript levels between populations. B) Percentage of genes that are differentially expressed as compared to the number of genes on the chromosome at each stage. At stage 2, where all transcripts are maternal in origin, there is a significant enrichment of DE genes on the X chromosome.

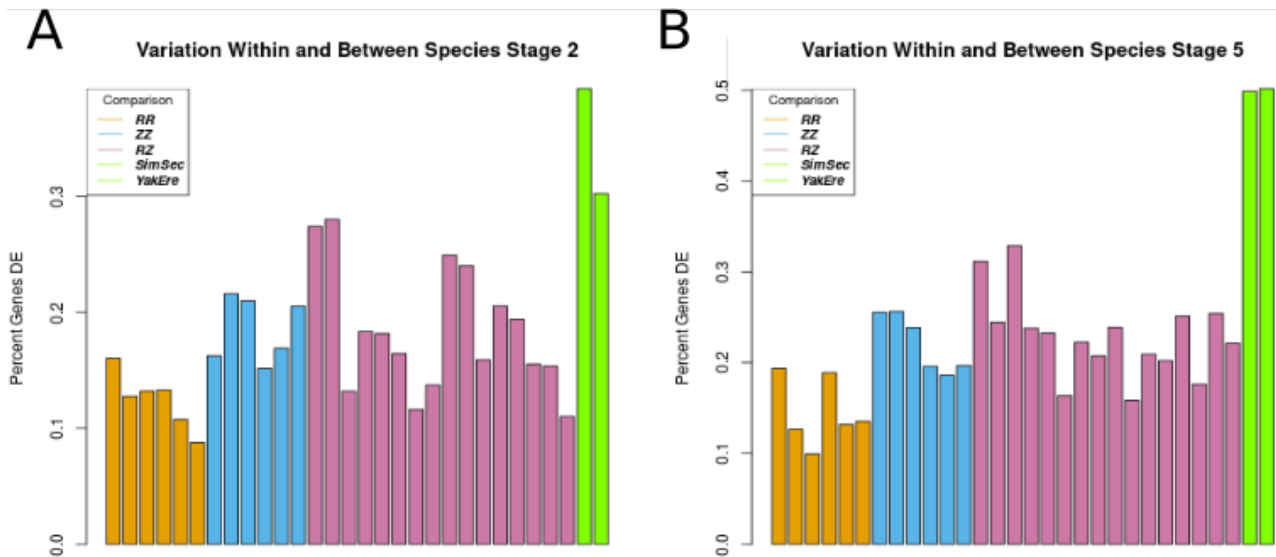


Figure 4. Differential Expression within and Between Species. DE analysis was done between individual Raleigh lines (orange), individual Zambia lines (blue), between lines of the two populations (purple) and between species pairs (green) from stage 2 and stage 5 embryos. The between species DE analysis was done between *D. simulans* and *D. sechellia* as well as *D. yakuba* and *D.erecta.* It was found that there were on average fewer DE genes between lines of *D. melanogaster* than between species pairs at both stages.

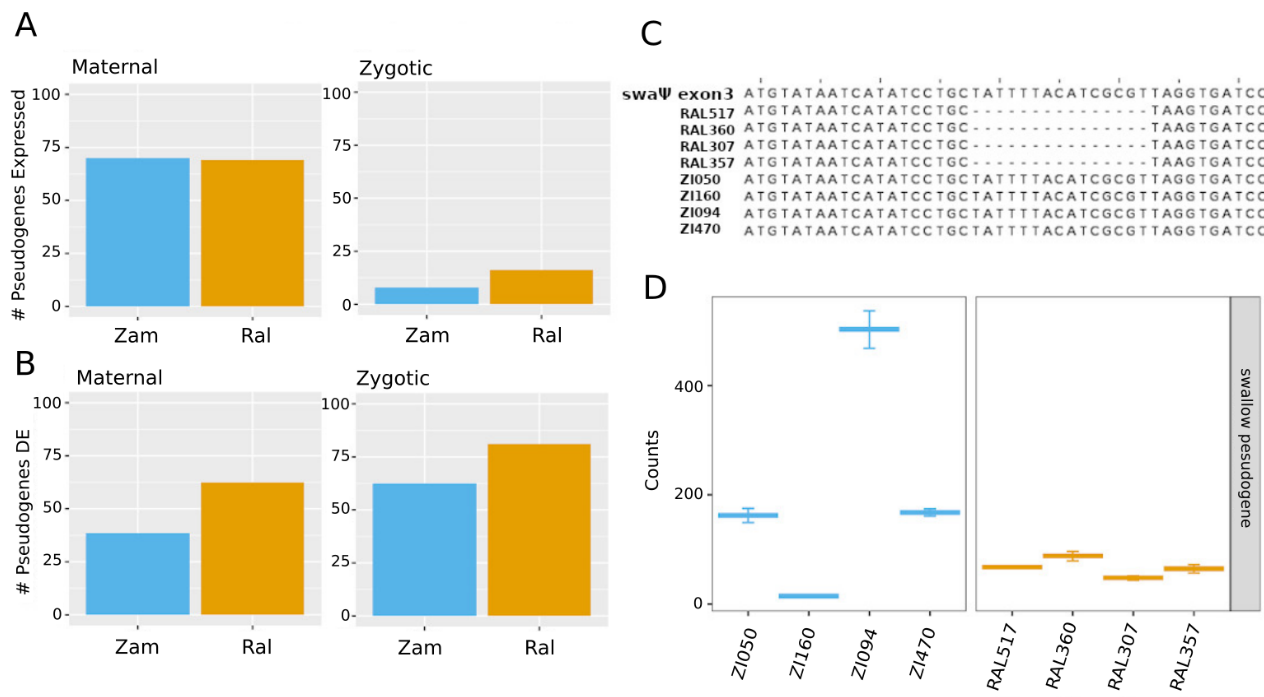


Figure 5. Variation in pseudogene transcript levels. A) At each stage (maternal, zygotic), the number of annotated pseudogenes expressed are similar between populations. The smaller number of pseudogenes expressed at the zygotic stage reflect that this analysis was restricted to zygotic-only genes, which are zygotic genes with no maternal expression. B) Of the pseudogenes expressed at each stage, a larger proportion are differentially expressed in the Raleigh lines. C) One example, the *swallow* pseudogene*,* has a 15bp deletion shared by all of the Raleigh lines at the position shown in the alignment. D) The *swallow* pseudogene is more highly expressed in a number of the Zambia lines, with considerable variation between lines.

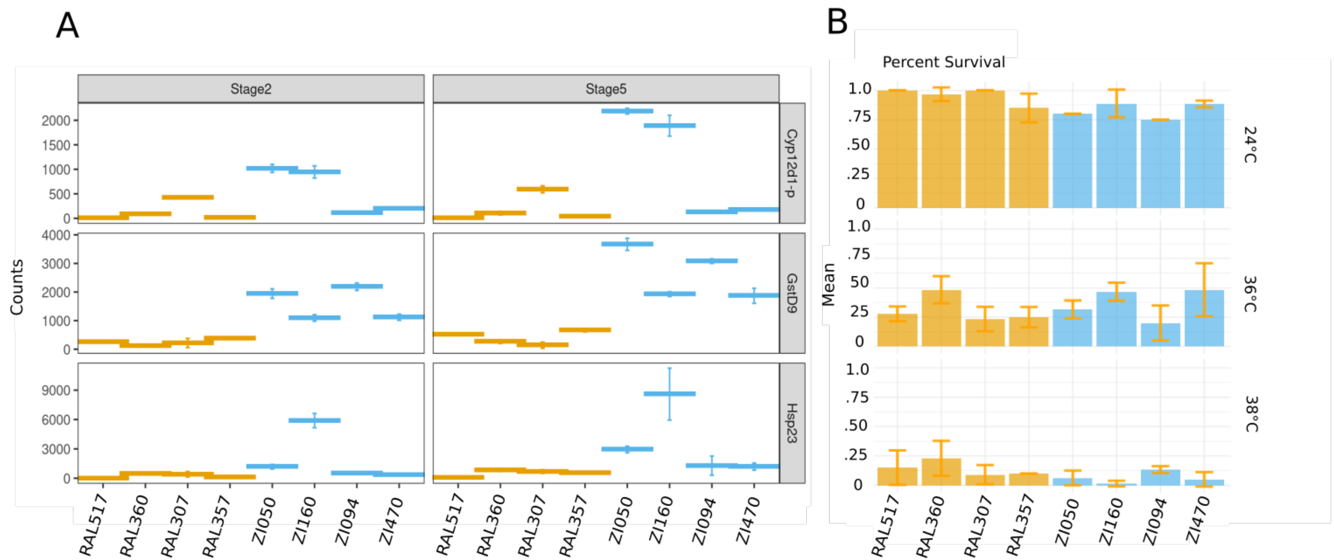
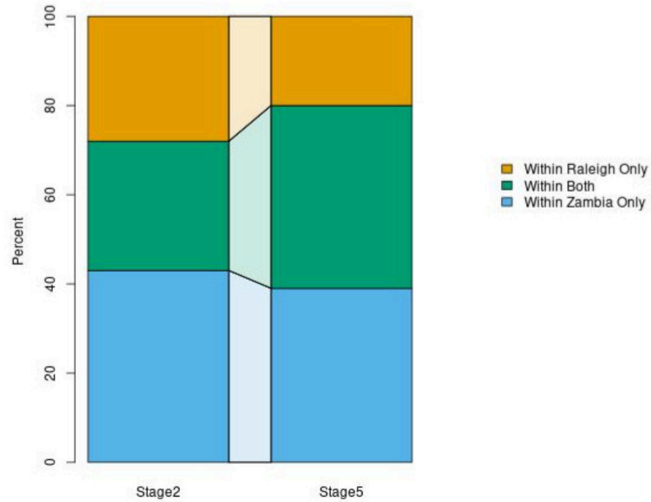


Figure 6. Examples of differentially expressed genes with previous evidence for functional significance. A) Transcript levels for three example genes, shown at both developmental stages labeled across the top, for the Raleigh lines (blue) and the Zambia lines (orange). B) Results of experiments testing survival of embryonic heat shock across lines, showing relative survival at three temperatures. While on average the Raleigh lines have higher survival after heat shock at 24°C and 38°C, they also have higher survival at standard rearing temperatures, results do not correspond well with heat shock transcript levels.



Supplemental Figure S1. Shared expression variation within populations. At both stage 2 and stage 5 there are a set of genes that are differentially expressed within both the Raleigh and Zambia populations (green). This shared set of genes differentially expressed in both populations increasing from stage 2 to stage 5.

Chapter 2

Evolution of mRNA Localization in the Drosophila Oocyte

Anna A. Feitzinger and Susan E. Lott

Department of Evolution and Ecology, University of California, Davis CA 95616

*Corresponding authors: afeitzinger@ucdavis.edu (AAF); selott@ucdavis.edu (SEL)

Abstract

Maternal mRNA localization during oogenesis is a conserved feature of many species. Localized mRNAs in the oocyte serve critical roles in early development such as specifying the body axis and determination of germ cells. Although the genes driving these critical processes are fundamental for proper development, the identity of these genes has changed over the course of evolution in Dipteran flies. For instance, the localized maternal transcript driving anterior fates has evolved several times in Diptera. In order for these changes in the localization of anterior determinants to take place, there must exist variation in both maternal expression and transcript localization. Evidence of changes in maternal deposition on relatively smaller timescales comes from a study characterizing the evolution of maternal transcripts across 12 species of Drosophila, revealing differences in maternal deposition across the genus. Here, we sought to determine if mRNA localization can change on relatively smaller timescales within a genus along the Drosophila phylogeny. We characterize differentially localized mRNAs in 5 species of Drosophila by bisecting stage 14 oocytes into anterior and posterior halves followed by RNA sequencing and differential expression analysis. We find that there is a higher number of significantly differentially enriched mRNAs in the anterior than the posterior in each species. We identify several changes in transcript localization along the phylogeny. We find several genes that have subcellular localization in only one species, or a subset of species, but lack expression in stage 14

oocytes of other species in our analysis. *D. pseudoobscura* was found to have the highest number of cases in this category, with 6 gains in expression and localization. We also find examples of transcripts which are expressed in oocytes of all 5 species but have anterior or posterior enrichment in only a subset. These results shed light on potential mechanisms by which novel mRNA localization schemes arise during oogenesis and exemplify how rapid gains and losses of localization can take place.

Background

mRNA localization in Drosophila oogenesis

The localization of mRNAs is a fundamental biological process that allows for the asymmetric distribution of mRNAs and proteins and is critical for the localized activity of genes in developing embryos and polarized cells [1–5]. Processes dictated by localized mRNAs include embryonic patterning, asymmetric cell fate decisions, epithelial polarity, cell migration, and neuronal morphogenesis [2]. During early *Drosophila melanogaster* development 70% of expressed mRNAs are subcellularly localized [6]. mRNA localization is best characterized for the maternally supplied body axis determinants in the oocyte of *D. melanogaster* [7–9]. During oogenesis the oocyte is considered transcriptionally silent, relying on 15 polyploid support cells, called nurse cells, to

transcribe and deposit large amounts of mRNAs into the oocyte [10,11]. The nurse cells are intracellularly connected to the oocyte by large actin rich cytoplasmic bridges called ring canals. The transport of key developmental transcripts from the site of transcription in the nurse cells to the oocyte relies on motor proteins and the microtubule cytoskeleton [12–15]. Active transport along the microtubules to the oocyte occurs over two days in mid-stage oogenesis (stages 2 to10A) [16]. At a later stage, the nurse cells rapidly dump their contents into the oocytes and undergo programmed cell death in a passive process called nurse cell dumping [17–19]. By stage 14, localized transcripts are anchored to subcellular locations in the oocyte [8].

Relatively little is known about the mechanisms by which mRNAs are localized during oogenesis genome-wide. The current model is that all maternally localized mRNAs are transported by the Egalitarian/Bicaudal-D/Dynein localization machinery from the nurse cells to the oocyte [12,13,20]. The non-canonical RNA binding protein Egalitarian (Egl) in conjunction with the cargo adaptor Bicaudal-D (Bic-D) are the only known adaptors which serve to physically link mRNAs to the minus end directed dynein motor via localization elements [13]. Cis-regulatory elements called localization elements, commonly found in the 3'UTR of transcripts, have been identified for a handful of maternal localized genes including *bicoid (bcd), gurken (grk), fs(1)K10 (K10),* and the *I* Factor retro-transposon [21–26]. Bcd, grk, and *oskar (osk)*, which are required

for body axis patterning, all rely on active microtubule mediated transport from the nurse cells to the oocyte followed by distinctive mechanisms for subcellular localization and anchoring within the oocyte [7,9,27].

The BLE1 cis regulatory sequence is required for the localization of the anterior determinant *bcd* and is one of the best characterized localization sequences [23,26,28]. This sequence, found in the 3'UTR of *bcd*, is required for transport to the oocyte through association with the microtubule cytoskeleton as well its subsequent localization to the anterior of the oocyte. *Osk* mRNA is reliant on cis-regulatory elements in the 3'UTR as well as splicing of the first intron, which creates a short RNA stem-loop for localization to the oocyte [21,24]. However, in mid-stage oogenesis, *osk* is localized to the posterior of the oocyte. This second phase of localization is also reliant on cis-acting elements and on microtubule directed transport, however; unlike its initial localization to the oocyte, posterior localization requires association with the plus-end directed motor Kinesin-1 [29–31].

In several documented cases, the cis-acting localization elements used during oogenesis are also functional during embryogenesis. While asymmetric mRNA localization of anterior-posterior patterning genes *bcd*, *osk* and *grk* utilize localization elements in the oocyte, a number of zygotic transcripts also require cis acting elements

to direct their localization to the apical region of the blastoderm embryo [32]. These include *wingless (wg)* and the pair-rule genes *fushi tarazu (ftz)*, *even-skipped (eve), hairy (h), paired,* and *runt* [33–36]. When *bcd, grk, nos,* and *K10,* all maternal transcripts, are injected into the blastoderm they are efficiently transported and observed to be apically localized [37]. Colcemid, which depolymerizes microtubules, inhibits these localization events [37]. Therefore, localization of mRNAs is an important feature throughout oogensis and early embryogenesis, and both maternal and zygotic transcripts share cis-acting localization elements and machinery.

The localization elements thus far characterized show little sequence conservation, rather, secondary structures have been shown to be important for recognition by Egl [13]. The low primary sequence conservation and the small number of localization elements identified makes computational predictions about which transcripts are transported and localized difficult. However, stem-loops are a shared feature in 3'UTR localization elements [23]. Additionally, NMR spectroscopy of the 44 nucleotide K10 signal sequence has revealed an A'-form RNA helix that is required for recognition by transport machinery and cytoplasmic transport [38].

Some cis-regulatory localization elements have conservation of function across Drosophila species. For example, 3' non coding regions containing Bicoid's BLE1

localization sequence from *D. simulans* and *D.virilis* were able to properly direct Bicoid localization in *D.melangaster* [39]. Additionally, the oskar signal sequence from *D.virilis* is able to direct osk localization to the posterior in *D.melanogaster* oocytes [40]. However, there is novel intermediate localization of *D.virilis* osk in both *D.virilis* oocytes not observed in *D.melanogaster*[40]. The signal sequence of *osk* from *D.virilis* can rescue body patterning in *D. melanogaster* but could not rescue pole cell formation [40]. Therefore, there is evidence of conservation of cis-regulatory

Recent studies have revealed that there are many more localized mRNAs in Drosophila oogenesis and early development than had previously been appreciated [41]. A recent study implementing RIP-seq has found that at least 50 transcripts associate with Egl/Bic-D localization machinery [42]. It has been suggested that Egl may be somewhat promiscuous [13], as over-expression of Egl in embryo was shown to drive a small population of non-localizing transcripts to the apical cytoplasm [43]. These studies have revealed that many transcripts are localized to and within the oocyte, however, much is yet to be learned about the mechanisms driving localization of transcripts genome-wide and how localization can be gained or lost over the course of evolution.

Changes in Maternal mRNA Expression and Localization in Diptera

Evidence that expression and localization of mRNAs during oogenesis changes over evolutionary time comes from a limited number of examples, such as the critical patterning factor, *bicoid*. The maternally supplied *bicoid* transcript is anteriorly localized in the oocyte of *D. melanogaster* and is required for anterior-posterior axis patterning [44–47]. Surprisingly, bicoid is not found outside of higher Diptera and is the result of a duplication of the Hox3 homeodomain transcription factor *zerknüllt (zen) [48]*. It has been hypothesized that *otd,* which is a maternally deposited body axis determinant in the wasp Nasonia and the beetle Tribolium, is the ancestral determinant whose function has been taken over by *bicoid* in cyclorrhaphan flies [49–53]. *Otd* is zygotically expressed, but not maternally supplied in *D. melanogaster,* and it is regulated by maternal *bcd* [54]. Additionally, more recent work has found that the anterior determinant has evolved multiple times in Diptera. Yoon et al. [55] determined the anterior determinants in a number of basal Diptera species including crane flies, moth flies, a common midge, and two mosquito species. This study found that even between mosquito species, different genes drive anterior determination: in culicine mosquitoes (*Culex*, *Aedes*) a previously uncharacterized zinc finger gene, *cucoid,* drove anterior fates, whereas the gene *pangolin* determines the anterior pole in the anopheline mosquito Anopheles. Additionally, *odd-paired* was found to be the anterior determinant

in moth flies. This study revealed that different mechanisms have contributed to the evolution of anterior determinants. In the case of *Drosophila* and *Chironomus*, a common midge, newly evolved genes drive anterior fates without roles outside of axis specification. Conversely, in the other species examined, pre-existing genes evolved anterior determinant function by acquiring maternal expression of alternative isoforms which are localized to the anterior of the oocyte. Still, the regulatory changes that lead to such dramatic changes in maternal expression and localization are unknown.

There have been very few investigations into the natural variation of maternal expression and localization on relatively shorter evolutionary time-scales. One such example is of *orthodentical (otd)* expression and localization in two cyclorrhaphan species; the medfly *Ceratitis capitata* and caribfly *Anastrepha suspensa* [56]. These species represent approximately 124 million years of divergence time [57]. Unlike Drosophila, o*td* is maternally expressed in both these species. However, o*td* is only found to be anteriorly localized in *A. suspensa*, exemplifying how swiftly changes in maternal expression and localization can occur.

Evidence of the rapid changes in maternal expression on even smaller evolutionary timescales comes from experiments done within the Drosophila genus. Single embryo RNA-seq of 14 Drosophila species, spanning 50 million years of

divergence, was performed at a stage before and after zygotic genome activation [58]. Analysis of this data has revealed that hundreds of maternally supplied transcripts are differentially expressed between even the most closely related species in this data set. However, the extent and rate of gains or losses in localization of maternal transcripts in the egg between species is unknown. Leveraging the availability of sequenced *Drosophila* genomes and tools for comparative analysis facilitated a systematic investigation as to how maternal transcript localization may evolve in smaller timescales along a phylogeny.

Given that there have been gains of expression and localization of maternal mRNAs in highly diverged species we hypothesized that changes in localization could arise in two different ways on smaller timescales: 1) a gain in maternal expression of transcripts that already have cis-acting regulatory elements that are able to associate with the pre-existing localization machinery or 2) a change in cis-regulatory sequence that either confers a gain or loss in localization in a gene that has pre-existing maternal expression. To investigate the evolution of maternal transcript localization across Drosophila, we bisected stage 14 oocytes from 5 species, performed RNA sequencing on the anterior and posterior halves of the oocytes, and used differential expression analysis to explore differential transcript localization. We discovered a relatively small number of significant changes in transcript localization across species, with examples of cases where expression and localization appear to have been gained simultaneously

and cases where a previously unlocalized maternal transcript gains localization. Overall, our results point to different ways transcripts may gain or lose localization; a simultaneous gain in maternal expression and localization of transcripts which may be due to gains in expression of transcripts with pre-existing cis-regulatory regions that can associate with localization machinery or changes in cis-regulatory regions that confer a gain of localization in transcripts which have already have maternal expression. We find a small number of transcripts with lineage specific localization, where transcripts are only localized in one of the species in our analysis. These species-specific localized transcripts may be necessary adaptations to species-specific developmental constraints.

Results

Reproducibility of Anterior and Posterior RNA sequencing data

To determine RNA localization changes within the *Drosophila* genus, we chose to characterize anteriorly and posteriorly enriched transcripts from five *Drosophila* species representing different evolutionary distances, with known variation in maternally deposited transcripts. Stage 14 oocytes from *Drosophila melanogaster*, *Drosophila*

simulans, *Drosophila pseudoobscura*, *Drosophila willistoni* and *Drosophila virilis* were dissected from ovaries followed by bisection into anterior and posterior halves. We pooled five oocyte halves per replicate and collected five biological replicates of each pole (anterior or posterior) per species. The RNA from these anterior and posterior samples were then sequenced. Sequences were aligned to their respective genomes via HISAT2 [59] and counts were generated using FeatureCounts [60] (see Methods). For all downstream analysis we only considered genes that have one to one orthologs in all five species (see Methods).

The data produced were highly reproducible, transcript abundances for replicates are highly correlated with Spearman correlation coefficients greater than 0.95 for all biological replicates in our dataset. Correlations between anterior and posterior halves were similarly correlated (all $r > .95$), indicating that most transcripts were not significantly polarized across the anterior-posterior axis. Principal component analysis (PCA) of all samples results in clustering of species along the first principal component, which explains 34.5% of the variation in the count data, however the second principal (22.4% variance) component does not correspond to the identity of the embryo pole (anterior or posterior) (Figure 1). These data together highlight that the mRNAs present in the anterior and posterior halves in stage 14 oocytes are highly correlated, and suggests that differences in RNA localization along the anterior-posterior axis represent a very small proportion of transcripts.

Among the most well documented localized maternal RNAs in *Drosophila* are the anterior and posterior determinants such as *bcd* and *osk*. To validate that we properly separated anterior and posterior halves, we first confirmed that known localized maternal mRNAs were enriched in the anterior and posterior by using differential expression analysis between poles. For each species, *bcd* and *osk* were found to be significantly enriched ($p < 0.05$) in the anterior and posterior poles, respectively (Figure 2). Additional mRNAs that are known to associate with Egl-Bcd machinery in *D. melanogaster* [42] were found to be enriched in *D. melanogaster* samples including *grk* and *nos*. The common set of mRNAs differentially enriched across all five species at this significance threshold (p value < 0.05) were *bcd*, *osk* and *bitesized (btsz)*. Log fold change of *bcd* ranged between 2.5 and 4.8 in the anterior direction among species. Log fold change values for *osk* in the posterior were much lower, ranging from 0.26 to 0.55 among species. This result is consistent with a previous study identified RNAs that coprecipitated with Egl and found that *bcd* had a much higher fold enrichment than *osk*, suggesting less efficient localization of *osk* [20]. Out of the known localized transcripts (*bcd, osk, grk* and *nos*), *bcd* had the lowest p-value and highest fold change in all species.

Proportion of localized transcripts

To examine the proportion of genes enriched in the anterior and posterior halves of each species, we restricted our analysis to genes with a log fold change of greater than 0.5 and a DE significance of $p < 0.05$. For all species, more significantly differentially expressed genes were found in the anterior as compared to the posterior (Figure 3). Of the five species examined, we find that *D. simulans* has the highest proportion of anteriorly localized RNAs and *D. virilis* has the highest proportion of posteriorly localized RNAs.

We then asked, among all the differentially enriched transcripts in each species, what proportion have a shared bias in enrichment across all species? To answer this, we first compiled a list of genes that were differentially enriched in at least one species (n=76). We then determined for each of these genes if it met our expression threshold and log fold cut off in the other species. 17 out of the 76 genes had at least two species that met both the expression and log fold cutoffs. Of these, 13 of the 17 were biased in the same direction in the other species in which the log fold change cut off was met. The remaining 56 genes were only found to be significantly enriched in a single species at these cutoffs; however, we cannot discount that they share similar localization in other species, but were missed beause our cut-offs were too stringent.

Co-evolution of gains in Expression and Localization

Next we sought to determine if a change in maternal deposition can lead to differences in the mRNA localization landscape of the oocyte. To identify cases where a gain in expression and a gain in localization co-occur, we used a stringent cutoff of $p < 0.01$ for anterior versus posterior differential expression analysis for each species. We then did pairwise comparisons between each species asking which genes were differentially expressed in one (with log fold change > 0.5) and below an expression cutoff (counts in both anterior and poster < 20) in the other species in the comparison. Remarkably, we find 13 instances in which we observe a gain in both maternal deposition and localization, including genes that are lineage specific and genes that have gains in more than one species. Ten of these changes are in genes that have anterior localization in only a subset of species, with three transcripts having posterior localization in a subset.

Interestingly, *D. pseudobscura* has the most instances of apparently simultaneous changes of both expression and localization, with a total of seven, two of these being lineage specific. Four of these genes are anteriorly localized in *D. pseudoobscura* with the remaining two being localized to the posterior. One gene, *Ppk11*, an epithelial sodium channel, is the most striking example, with little to no expression in any of the species but high expression and significant anterior localization

in *D. pseudoobscura*. Another striking example of anterior localization in *D. pseudoobscura* is the transcript, *Ccdc85,* which has much higher levels of maternal deposition than any of the other species and is absent in the *D. wilistoni* oocyte (Figure 4).

Consistent with our finding that *D. simulans* has the highest proportion of localized transcripts in the anterior, all changes in both expression and localization in *D. simulans* are in anteriorly enriched transcripts. Out of the three genes which have unique anterior localization in *D.simulans*, one occurs only *D. simulans* and two share maternal expression and anterior bias in either *D.melanogaster* or *D.willistoni*. We find that one transcript, *fire exit*, is only maternally expressed in *D. melanogaster* and *D. simulans* and anterior bias in both, suggesting a gain in expression and localization along the lineage leading to these two species.

Gains in localization of ancestral maternally deposited transcripts

To evaluate cases where localization was gained for a maternal gene that likely was already expressed, we limited our next analysis to genes that are maternally deposited in all species in our data set. Genes with greater than 20 counts in either the anterior or posterior were considered maternally deposited. We then did all possible pairwise comparisons between species pairs to determine which out of the genes

maternally deposited in all species have differential localized in only one of the two species in the comparison. This generated a list of 30 genes from all of the comparisons. However, seven of these genes showed similar bias of localization in all species, with only a subset being significant, therefore, we discounted them from further analysis. The remaining 23 genes had significantly differential localization between species. Sixteen of these genes are localized to the anterior in at least one species, with lack of enrichment, or localization changes in one or more of the other species. The remaining six transcripts have posterior localization in one or more of the species, with one or more of the remaining species lacking significant posterior localization.

We found a number of genes that have appeared to gain localization to the posterior of the *D.virilis* oocyte as well as losses in localization in *D.virilis* of transcripts that are anteriorly localized in other species. Three out of the six posterior localization changes are transcripts that are maternally deposited in all species but have significant posterior localization in *D. virilis*. One of these three transcripts, *vig,* shares non-significant bias to the posterior in *D. pseudoobscura*, with another of these transcripts, *TyrRS-m,* having a non-significant bias to the anterior in *D.pseudoobscura*. The remaining three posterior quantitative differences have significant posterior enrichment in *D. wilistoni*, with only one of these genes sharing non-significant posterior bias with *D.virilis.* In contrast to possible gains in posterior localization in *D. virilis*, we find two transcripts, *mutator 2 (mu2)* and *Xeroderma pigmentosum (Xpc)*, both involved in DNA

break repair, that have anterior bias in all species except for *D. virilis*, suggesting a possible loss of anterior localization in this species.

We find branch specific gains and losses in localization of ancestrally expressed maternal genes. Out of the anterior localization differences, we find two transcripts that have strong significant anterior localization in only *D. simulans* and *D. melanogaster* (Figure 5A). Both *flower (fwe)* and *hu li tai shao (hts)* have strong and significant anterior localization in *D. melanogaster* and *D. simulans* which is absent from the remaining species, suggesting a possible gain in localization along the branch leading to *D. melanogaster and D. simulans*. Conversely, *ripped pocket (rpk)* is anteriorly enriched in D*. pseudoobscura*, *D. wilistoni* and *D. virilis* without enrichment to either pole in *D. melanogaster* or *D. simulans* (Figure 5B). *D. wilistoni* has seven transcripts that are uniquely significantly enriched to the anterior of its oocyte. Only two of these transcripts share similar non-significant bias to the anterior with another species.

Notably, we also found ten cases where transcripts were significantly enriched in at least one species but have non-significant bias to the other pole in another species. Nine of these cases are transcripts with significant anterior localization in at least one species but have greater abundances in the posterior than anterior of *D. pseudoobscura*.

Discussion

While previous studies comparing species in different genera have shown that maternal deposition and localization has changed over larger timescales, here we present evidence of changes in maternal transcript localization within a single genus. We chose to examine the anterior and posterior localization of transcripts in stage 14 oocytes of five species of Drosophila to represent different evolutionary divergence times ranging from ~2-50 million years. In this study, we have identified several changes in maternal transcript localization across these timescales in the Drosophila phylogeny. We present two categories of changes in maternal localization between species. In the first category, we find instances in which genes are maternally deposited and localized in only a subset of species. In the second category, we find genes that are maternally deposited in all the species in our data set, but only show enrichment to either the anterior or posterior pole in a subset of species. These two categories of changes represent two potentially distinct mechanisms by which new instances of maternal mRNA localization occur.

Our results show that in each species there is a higher number of significantly enriched transcripts in the anterior than to the posterior of the oocyte, with *D. simulans* having the highest number of significantly anteriorly enriched transcripts and *D.virilis*

having the least. Previous work has shown that *D. simulans* has the highest proportion of its genome maternally deposited whereas *D. virilis* has the lowest proportion of its genome maternally deposited [58]. The Egl Bic-D machinery, directly linking mRNAs to the minus end directed motor dynein, is hypothesized to be somewhat promiscuous, as embryonic over expression of Egl drives a small population of non-localizing transcripts to the apical cytoplasm [13,43]. We therefore speculate that having a more diverse pool of maternally deposited transcripts may increase the chance of transcripts being localized. However, further experiments are required to further investigate this hypothesis.

We find 13 cases of simultaneous changes in maternal deposition and localization. These data suggest that gain or loss in maternal deposition may coincide with a gain or loss in localization. One possible way a simultaneous gain in deposition and localization may be achieved is if a transcript that gains maternal deposition already has a localization signal. However, another possibility is that maternal deposition was gained followed by changes in cis-regulatory sequences that then lead to a gain in localization. In this second scenario, we have not sampled the intermediate state where maternal transcription is gained yet the transcript is not localized. For these observed gains of transcription and localization, *D. pseudoobscura* was exceptional, with the largest number (four) of transcripts observed to be both maternally deposited and localized uniquely in this species. This finding is consistent with other previous

observations about the unique nature of maternal transcription in the *D. pseudoobscura* group species, in that they have an excess of lineage-specific maternally deposited transcripts [58], and evolved an exceptional number of changes to regulation of the maternal transcriptome, possibly mediated by transposable element expansion in these species [61,62].

We also find 23 cases in which maternal deposition occurs in all five species, but significant enrichment to either the anterior or posterior of the oocyte is observed in only a subset of the species. Interestingly we find two transcripts, *fwe* and *hts*, which have potentially gained localization along the branch leading to *D.simulans* and *D. melanogaster*. Another transcript, *rpk*, has potentially lost localization along the branch leading to *D. simulans* and *D. melanogster*, as it only has bias to the anterior in *D. pseudoobscura, D. wilistoni* and *D.virilis. D. wilistoni* was found to have the highest number of differences in localization of these genes maternally deposited across all species, with seven transcripts having subcellular localization which is absent in other species.

We identify more cases of transcripts that likely gain localization after maternal localization (23) than cases where maternal deposition and localization occur apparently simultaneously (13), according to our criteria. However, these are very small numbers of

changes, and thus we lack the power to determine if localization gain, following already present maternal deposition, occurs with significantly higher frequency. The small numbers do suggest that changes in localization along the anterior-posterior axis are relatively infrequent, likely reflecting either strong conservation of localization or mechanistic barriers in the evolution of localization.

One common theme of previous work on the evolution of new anterior determinants in Diptera is that both maternal expression and localization within the oocyte are needed, which occurs This has been shown to have occurred through different mechanisms in Diptera. One example is the evolution of *bicoid*, which is the result of a gene duplication of the *hox3* gene *zen*. It has been hypothesized that bicoid evolved the ability to regulate transcription factors required for anterior patterning and that localization to the anterior of the oocyte must have also evolved [63]. More recent studies elucidating the anterior determinants of more basal Diptera identified a previously unknown mechanism involving alternative maternally expressed isoforms of old genes. Yoon et al [55] hypothesize that alternative transcription facilitated evolution of anterior determinants by supplying a UTR sequence for isoform-specific localization signals [55]. In each of these cases a gain of localization through changes in cis-acting regulatory sequences are required to facilitate a major shift in the evolution of the highly critical developmental process of axis patterning. In previous studies have shown that there is conservation of localization machinery between different species of Drosophila.

For example, 3' non-coding regions containing *bicoid* localization signals from *D. simulans* and *D.virilis* were able to confer anterior localization of a hybrid *bicoid* transgene in *D.melanogaster* [39]. We cannot say mechanistically how the transcripts previously uncharacterized to have anterior or posterior localization in our findings are recruited to either pole. However, we hypothesize that for genes that we are maternally deposited in all five Drosphila species, but are only localized in a subset, the changes are due to pre-existing sequences in the UTRs that facilitate association with localization machinery. Additional experiments must be done to further validate these localization changes across species and to determine what regulatory changes have occurred that have led to these differences.

Conclusion

We have found that variation in maternal deposition and localization has evolved along the Drosophila phylogeny. While it was known that changes in maternal gene expression and localization have occurred between highly diverged species, we find that differences in anterior and posterior localization of transcripts can occur over smaller evolutionary timescales. Remarkably, we find instances of apparent gains in both expression and localization along the Drosophila phylogeny, where expression and localization are only found in a subset of species. These may be due to gains in

maternal expression of transcripts that have cis-regulatory elements which can associate with localization machinery. Additionally, we find instances of transcripts which are maternally expressed in all species but are only localized in a subset. These cases may be the outcome of changes in cis-regulatory regions in the transcript that lead to a gain or loss of localization. We also find a small number of species-specific localized transcripts, where localization of transcripts is only found in a single species. These species-specific localized transcripts may be necessary adaptations to species-specific developmental constraints. However, the small overall number of localization differences suggests a strong conservation of localization and potential mechanistic barriers to the evolution of novel localization.

Methods

RNA-seq sample preparation and sequencing

Drosophila ovaries were dissected from adult female flies using forceps. Individual oocytes were then sorted, and stage 14 oocytes were chosen for further processing. Individual stage 14 oocytes were transferred to glass slides. The glass slide was then placed on dry ice to allow the oocyte to freeze. Oocytes were then bisected using a cryosectioning blade by hand. Anterior and posterior halves were then placed in 1.5ml microcentrifuge tubes containing 50µl of TriZol (Agilent). Five halves were pooled

per sample. Samples were then frozen at -80°C until further RNA processing. RNA was extracted as in the manufacturer's protocol, with the exception of extracting in an excess of reagent (1mL was used) compared to expected mRNA concentration.

Prior to sequencing, RNA quality and integrity were assessed on an Agilent 2100 bioanalyzer using RNA 6000 Nano Chips (Agilent Technologies, USA). RNA samples were prepared for sequencing using the NEBNext Ultra™ II RNA Library Prep Kit for Illumina Sequencing and sequenced on Novaseq6000 PE150.

Data Processing and analysis

Reads were trimmed and adapters removed using Cutadapt [64], and gently (PHRED Q < 20) trimmed for quality. Mapping was done using genome for each species (from the 12 species project, downloaded from Flybase) and associated annotation file using HISAT2 version 2.1.0 [59] using default parameters. Gene level counts were generated using featureCounts [60] of the subRead [65] package in R (R version 3.4.1) [66]. Analysis was restricted to genes which have one to one orthologs as determined by OrthoDB [67].

To determine anterior and posterior enrichment differential expression analysis was done using the EdgeR [68] package in R. Using EdgeR and ran differential expression analysis between anterior and posterior samples for each species. For principal component analysis (PCA) was also performed in R using the prcomp() function. The results of the differential expression analysis between anterior and posterior halves of each species were presented using volcano plots from the EnhancedVolcano packaged in R.

We restricted the outputs from the differential expression analysis to genes with $p<0.01$ and with a log fold change of greater than 0.5. When determining which transcripts likely gained expression and localization simultaneously in a lineage, we did pairwise comparisons between each species determining which genes were differentially expressed in one species and below an expression cut off 20 counts (counts in both anterior and poster < 20) in the other species in the comparison.

To determine which transcripts were present in all species but localized in only a subset, we only considered genes which had greater than 20 counts in either the anterior or posterior in all species. We then did all possible pairwise comparisons between species pairs to determine if, out the genes expressed in all species (counts >

20), which genes have significant differential localization in only one of the two species in the comparison.

References

1. Medioni C, Mowry K, Besse F. Principles and roles of mRNA localization in animal development. Development [Internet]. 2012;139:3263–76. Available from: http://dev.biologists.org/cgi/doi/10.1242/dev.078626

2. Holt CE, Bullock SL. Europe PMC Funders Group Subcellular mRNA Localization in Animal Cells and Why It Matters Mechanisms of mRNA Localization : Illuminating a Multi-Step Process. 2013;326:1212–6.

3. Gonsalvez GB, Long RM. Spatial regulation of translation through RNA localization. 2012;11.

4. Kumano G. Polarizing animal cells via mRNA localization in oogenesis and early development. Dev Growth Differ. 2012;54:1–18.

5. Vazquez-Pianzola P, Suter B. Conservation of the RNA transport machineries and their coupling to translation control across eukaryotes. Comp Funct Genomics. 2012;2012.

6. Lécuyer E, Yoshida H, Parthasarathy N, Alm C, Babak T, Cerovina T, et al. Global Analysis of mRNA Localization Reveals a Prominent Role in Organizing Cellular Architecture and Function. Cell. 2007;131:174–87.

7. Kugler JM, Lasko P. Localization, anchoring and translational control of oskar, gurken, bicoid and nanos mRNA during drosophila oogenesis. Fly (Austin). 2009;3:15–28.

8. Lasko P. mRNA localization and translational control in Drosophila oogenesis. Cold Spring Harb Perspect Biol. 2012;4:1–15.

9. Chang CW, Nashchekin D, Wheatley L, Irion U, Dahlgaard K, Montague TG, et al. Anterior-posterior axis specification in drosophila oocytes: Identification of novel bicoid and oskar mRNA localization factors. Genetics. 2011;188:883–96.

10. Navarro-Costa P, McCarthy A, Prudêncio P, Greer C, Guilgur LG, Becker JD, et al. Early programming of the oocyte epigenome temporally controls late prophase i transcription and chromatin remodelling. Nat Commun. 2016;7.

11. Mische S, Li M, Serr M, Hays TS. Direct Observation of Regulated Ribonucleoprotein Transport Across the Nurse Cell / Oocyte Boundary □. 2007;18:2254–63.

12. Clark A, Meignin C, Davis I. A Dynein-dependent shortcut rapidly delivers axis determination transcripts into the Drosophila oocyte. Development [Internet]. 2007;134:1955–65. Available from: http://dev.biologists.org/cgi/doi/10.1242/dev.02832

13. Dienstbier M, Boehl F, Li X, Bullock SL. Egalitarian is a selective RNA-binding protein linking mRNA localization signals to the dynein motor. Genes Dev. 2009;23:1546–58.

14. Theurkauf WE, Hazelrigg TI. In vivo analyses of cytoplasmic transport and cytoskeletal organization during Drosophila oogenesis : characterization of a multi-step anterior localization pathway. 1998;3666:3655–66.

15. Cha BJ, Koppetsch BS, Theurkauf WE. In vivo analysis of drosophila bicoid mRNA localization reveals a novel microtubule-dependent axis specification pathway. Cell. 2001;

16. Steinhauer J, Kalderon D. Microtubule polarity and axis formation in the Drosophila oocyte. Dev Dyn. 2006;235:1455–68.

17. Buszczak M, Cooley L. Eggs to die for: Cell death during Drosophila oogenesis. Cell Death Differ. 2000;7:1071–4.

18. Pritchett T, Tanner E. Cracking open cell death in the Drosophila ovary. Apoptosis [Internet]. 2009;14:969–79. Available from: http://www.springerlink.com/index/h55507lk1077745m.pdf

19. Cavaliere V, Taddei C, Gargiulo G. Apoptosis of nurse cells at the late stages of oogenesis of Drosophila melanogaster. Dev Genes Evol. 1998;208:106–12.

20. Sanghavi P, Liu G, Veeranan-Karmegam R, Navarro C, Gonsalvez GB. Multiple roles for egalitarian in polarization of the Drosophila egg chamber. Genetics. 2016;203:415–32.

21. Jambor H, Mueller S, Bullock SL, Ephrussi A. A stem-loop structure directs oskar mRNA to microtubule minus ends. Rna. 2014;20:429–39.

22. Serano TL, Cohen RS. A small predicted stem-leap structure mediates oocyte localization of Drosophila K10 mRNA. Development. 1995;121:3809–18.

23. Snee MJ, Arn EA, Bullock SL, Macdonald PM. Recognition of the bcd mRNA Localization Signal in Drosophila Embryos and Ovaries . Mol Cell Biol. 2005;25:1501–10.

24. Ryu YH, Kenny A, Gim Y, Snee M, Macdonald PM. Multiple cis-acting signals, some weak by necessity, collectively direct robust transport of oskar mRNA to the oocyte. J Cell Sci. 2017;130:3060–71.

25. Thio GL, Ray RP, Barcelo G, Schüpbach T. Localization of gurken RNA in Drosophila oogenesis requires elements in the 5’ and 3’ regions of the transcript. Dev Biol. 2000;221:435–46.

26. Macdonald PM, Kerr K, Smith JL, Leask A. RNA regulatory element BLE1 directs the early steps of bicoid mRNA localization. 1993;1243:1233–43.

27. Irion U, St Johnston D. bicoid RNA localization requires specific binding of an endosomal sorting complex. Nature. 2007;445:554–8.

28. Macdonald PM, Struhl G. Cis- acting sequences responsible for anterior localization of bicoid mRNA in Drosophila embryos. Nature [Internet]. 1988;336:595–8. Available from: http://www.nature.com/doifinder/10.1038/336595a0

29. Zimyanin VL, Belaya K, Pecreaux J, Gilchrist MJ, Clark A, Davis I, et al. In Vivo Imaging of oskar mRNA Transport Reveals the Mechanism of Posterior Localization. Cell [Internet]. Elsevier Inc.; 2008;134:843–53. Available from: http://dx.doi.org/10.1016/j.cell.2008.06.053

30. Huynh JR, Munro TP, Smith-Litière K, Lepesant JA, St Johnston D. The Drosophila hnRNPA/B homolog, Hrp48, is specifically required for a distinct step in osk mRNA localization. Dev Cell. 2004;6:625–35.

31. Micklem DR, Adams J, Grünert S, St Johnston D. Distinct roles of two conserved Staufen domains in oskar mRNA localization and translation. EMBO J. 2000;19:1366–77.

32. Davis I, Ish-Horowicz D. Apical localization of pair-rule transcripts requires 3′ sequences and limits protein diffusion in the Drosophila blastoderm embryo. Cell. 1991;67:927–40.

33. Bullock SL. Differential cytoplasmic mRNA localisation adjusts pair-rule transcription factor activity to cytoarchitecture in dipteran evolution. Development [Internet]. 2004;131:4251–61. Available from: http://dev.biologists.org/cgi/doi/10.1242/dev.01289

34. Bullock SL, Zicha D, Ish-Horowicz D. The Drosophila hairy RNA localization signal modulates the kinetics of cytoplasmic mRNA transport. EMBO J. 2003;22:2484–94.

35. Santos G, Simmonds AJ, Krause HM. A stem-loop structure in the wingless transcript defines a consensus motif for apical RNA transport. 2008;143:133–43.

36. Lall S, Francis-Lang H, Flament A, Norvell A, Schüpbach T, Ish-Horowicz D. Squid hnRNP protein promotes apical cytoplasmic transport and localization of Drosophila pair-rule transcripts. Cell. 1999;98:171–80.

37. Bullock SL, Ish-Horowicz D. Conserved signals and machinery for RNA transport in Drosophila oogenesis and embryogenesis. Nature. 2001;414:611–6.

38. Bullock SL, Ringel I, Ish-horowicz D, Lukavsky PJ. A ′ -form RNA helices are required for cytoplasmic mRNA transport in Drosophila. Nat Publ Gr [Internet]. Nature Publishing Group; 2010;17:703–9. Available from: http://dx.doi.org/10.1038/nsmb.1813

39. Luk SKS, Kilpatrick M, Kerr K, Macdonald PM. Components acting in localization of bicoid mRNA are conserved among Drosophila species. Genetics. 1994;137:521–30.

40. Webster PJ, Suen J, Macdonald PM. Drosophila-Virilis Oskar Transgenes Direct Body Patterning but Not Pole Cell-Formation or Maintenance of Messenger-Rna Localization in Drosophila-Melanogaster. Development. 1994;120:2027–37.

41. Jambor H, Surendranath V, Kalinka AT, Mejstrik P, Saalfeld S, Tomancak P. Systematic imaging reveals features and changing localization of mRNAs in Drosophila development. Elife. 2015;4:1–22.

42. Vazquez-Pianzola P, Schaller B, Colombo M, Beuchle D, Neuenschwander S, Marcil A, et al. The mRNA transportome of the BicD/Egl transport machinery. RNA Biol [Internet]. Taylor & Francis; 2017;14:73–89. Available from: http://dx.doi.org/10.1080/15476286.2016.1251542

43. Bullock SL, Nicol A, Gross SP, Zicha D. Guidance of Bidirectional Motor Complexes by mRNA Cargoes through Control of Dynein Number and Activity. Curr Biol. 2006;16:1447–52.

44. Driever W, Nüsslein-Volhard C. The bicoid protein determines position in the Drosophila embryo in a concentration-dependent manner. Cell. 1988;

45. Berleth T, Burri M, Thoma G, Bopp D, Richstein S, Frigerio G, et al. The role of localization of bicoid RNA in organizing the anterior pattern of the Drosophila embryo. EMBO J. 1988;7:1749–56.

46. Little SC, Wieschaus EF. Shifting Patterns: Merging Molecules, Morphogens, Motility, and Methodology. Dev Cell [Internet]. Elsevier Inc.; 2011;21:2–4. Available from: http://dx.doi.org/10.1016/j.devcel.2011.06.003

47. Porcher A, Dostatni N. The Bicoid Morphogen System. Curr Biol. 2010;20:249–54.

48. Stauber M, Jäckle H, Schmidt-Ott U, Gehring WJ. The anterior determinant bicoid of Drosophila is a derived Hox class 3 gene. Dev Biol. 1999;96:3786–9.

49. Lynch JA, Brent AE, Leaf DS, Pultz MA, Desplan C. Localized maternal orthodenticle patterns anterior and posterior in the long germ wasp Nasonia. Nature. 2006;439:728–32.

50. Kotkamp K, Klingler M, Schoppmeier M. Apparent role of Tribolium orthodenticle in anteroposterior blastoderm patterning largely reflects novel functions in dorsoventral axis formation and cell survival. Development. 2010;137:1853–62.

51. Schinko JB, Kreuzer N, Offen N, Posnien N, Wimmer EA, Bucher G. Divergent functions of orthodenticle, empty spiracles and buttonhead in early head patterning of the beetle Tribolium castaneum (Coleoptera). Dev Biol. 2008;317:600–13.

52. Wolff C, Schröder R, Schulz C, Tautz D, Klingler M. Regulation of the Tribolium homologues of caudal and hunchback in Drosophila : evidence for maternal gradient systems in a short germ embryo. 1998;3654:3645–54.

53. Lynch J, Desplan C. Evolution of development: Beyond bicoid. Curr Biol. 2003;13:557–9.

54. Finkelstein R, Perrimon N. The orthodenticle gene is regulated by bicoid and torso and specifies Drosophila head development. Nature. 1990;346:485–8.

55. Yoon Y, Klomp J, Martin-Martin I, Criscione F, Calvo E, Ribeiro J, et al. Embryo polarity in moth flies and mosquitoes relies on distinct old genes with localized transcript isoforms. Elife. 2019;8:1–30.

56. Schetelig MF, Schmid BGM, Zimowska G, Wimmer EA. Plasticity in mRNA expression and localization of orthodenticle within higher Diptera. Evol Dev. 2008;10:700–4.

57. Krosch MN, Schutze MK, Armstrong KF, Graham GC, Yeates DK, Clarke AR. A molecular phylogeny for the Tribe Dacini (Diptera: Tephritidae): Systematic and biogeographic implications. Mol Phylogenet Evol [Internet]. Elsevier Inc.; 2012;64:513–23. Available from: http://dx.doi.org/10.1016/j.ympev.2012.05.006

58. Atallah J, Lott SE. Conservation and evolution of maternally deposited and zygotic transcribed mRNAs in the early Drosophila embryo. PLOS Genet [Internet]. 2018;14:1–27. Available from: http://dx.plos.org/10.1371/journal.pgen.1007838

59. Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. Nat Protoc [Internet]. Nature Publishing Group; 2016;11:1650–67. Available from: http://dx.doi.org/10.1038/nprot.2016-095

60. Liao Y, Smyth GK, Shi W. FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2014;30:923–30.

61. Omura CS, Lott SE. The conserved regulatory basis of mRNA contributions to the early Drosophila embryo differs between the maternal and zygotic genomes. PLoS Genet [Internet]. 2020;16:1–30. Available from: http://dx.doi.org/10.1371/journal.pgen.1008645

62. Omura CS, Lott SE. Evolution of maternal and early zygotic transcript regulation across Drosophila. bioRxiv [Internet]. 2021;2021.10.28.466359. Available from: http://biorxiv.org/content/early/2021/10/29/2021.10.28.466359.abstract

63. Liu Q, Onal P, Datta RR, Rogers JM, Schmidt-Ott U, Bulyk ML, et al. Ancient mechanisms for the evolution of the bicoid homeodomain's function in fly development. Elife. 2018;7:1–28.

64. Microscopy CZ. RNA extraction from frozen sections Some helpful tips before starting : Microtome cutting and Cresyl Violet staining. Manual.

65. Shi W, Liao Y. Subread / Rsubread Users Guide. 2018;0–55.

66. R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

67. Zdobnov EM, Tegenfeldt F, Kuznetsov D, Waterhouse RM, Simao FA, Ioannidis P, et al. OrthoDB v9.1: Cataloging evolutionary and functional annotations for animal, fungal, plant, archaeal, bacterial and viral orthologs. Nucleic Acids Res. 2017;45:D744–9.

68. Zhang ZH, Jhaveri DJ, Marshall VM, Bauer DC, Edson J, Narayanan RK, et al. A comparative study of techniques for differential expression analysis on RNA-seq data. PLoS One. 2014;

Figures

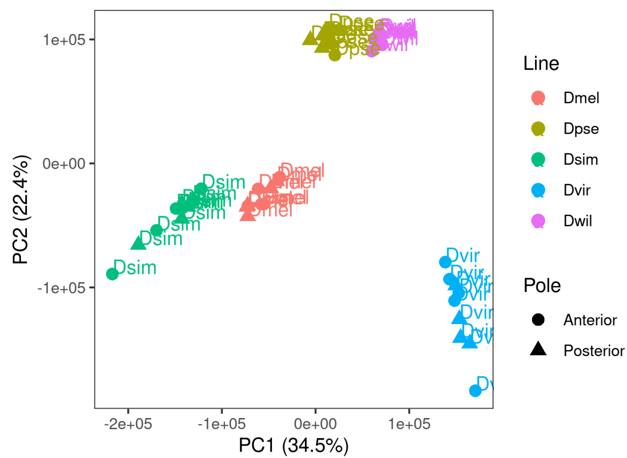


Figure 1. Transcriptomes from anterior posterior halves cluster by species. Principal component analysis shows that all replicates for anterior and posterior halves cluster by species (PC1, which explains a large proportion of the variance at 34.5%). The identity of the pole, anterior of posterior, does not explain a large proportion of the variance in the data set, as halves do not cluster with each other.

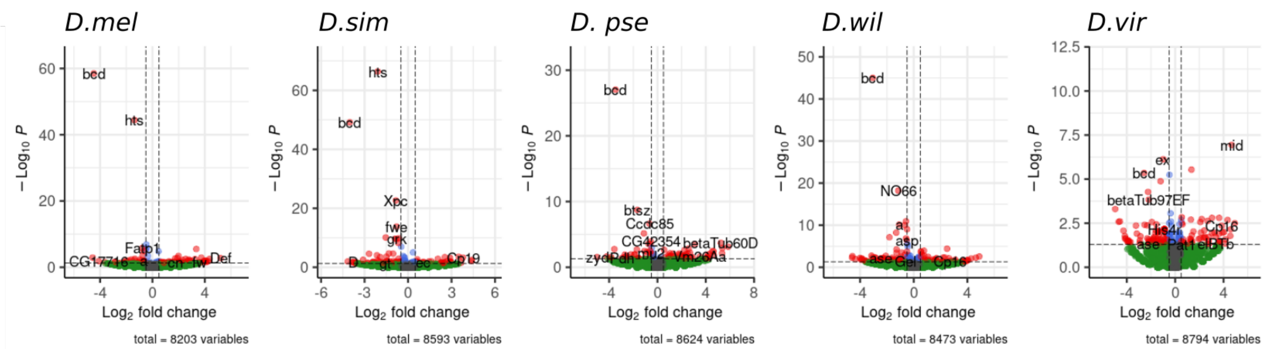


Figure 2. Differential expression between anterior and posterior halves of Stage 14 oocytes

Volcano plots for differential expression analysis of anterior versus posterior halves for each species. Horizontal dotted line represents a p value cut off, with all points above the horizontal line having a p value less than 0.05. Vertical lines represent a log fold change cut off of 0.5, with genes enriched to the anterior to the left of the dotted line and genes enriched to the posterior to the right of the dotted line.

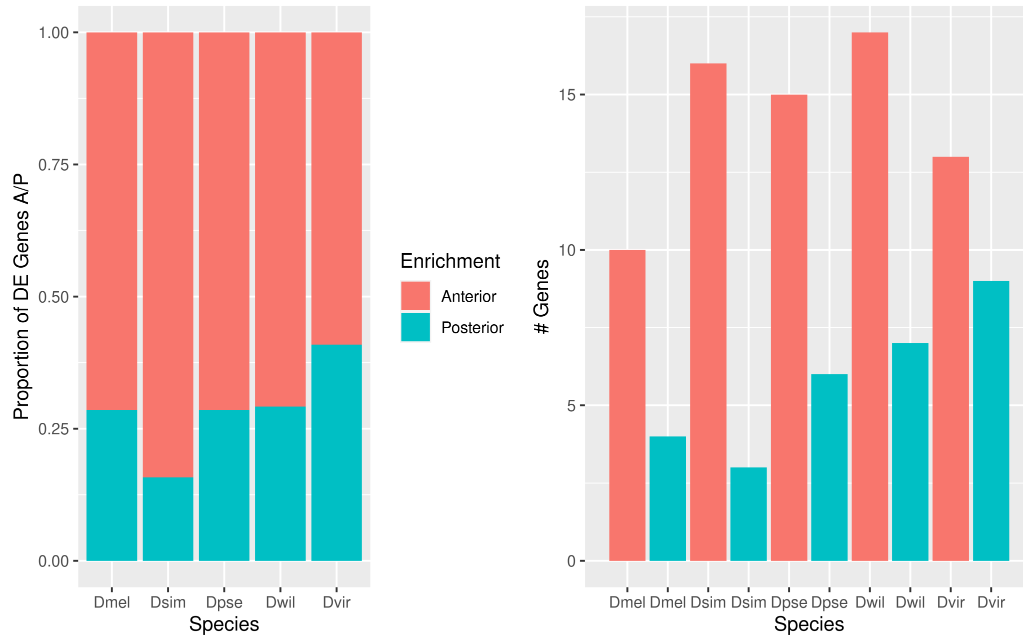


Figure 3. Higher Proportion of Significantly Enriched Transcripts in the Anterior than to Posterior Of The Oocyte

A) For all species we find a higher proportion of transcripts with a p value < .01 and log fold value > .5 in the anterior of the oocyte than posterior of the oocyte. B) Number of significantly enriched genes in the anterior and posterior of each species.

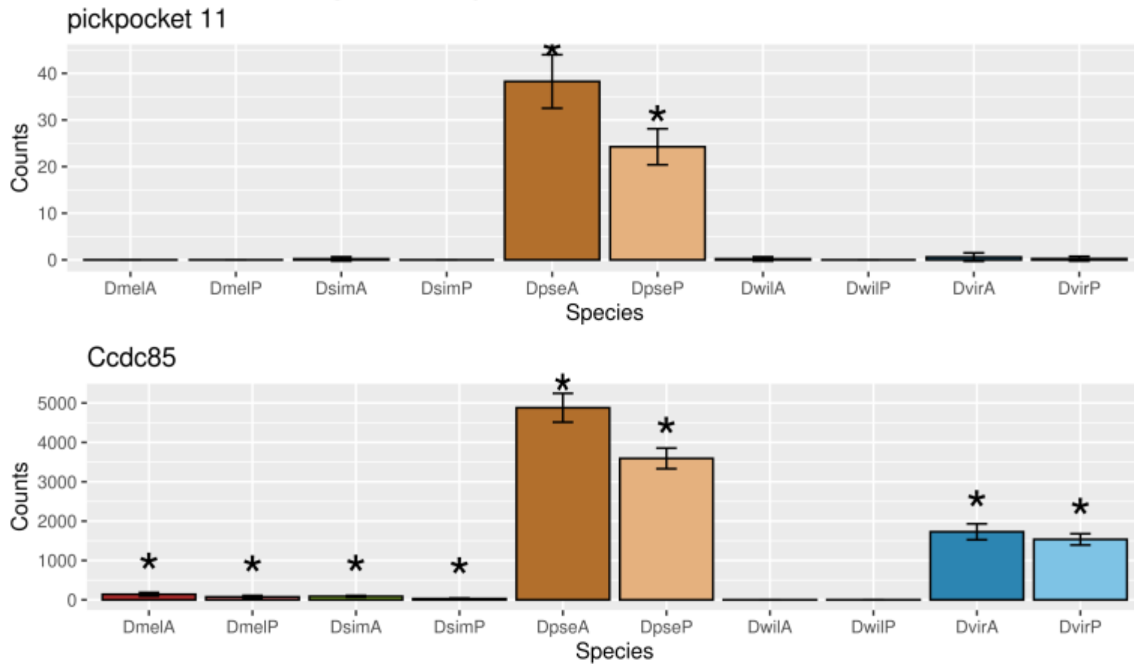


Figure 4. Changes in Both Expression and Localization
Examples of changes in both expression and localization. A) Pickpocket 11 (*Ppk11*) is only expressed in *D.pseudoobscura* oocytes, where it is localized to the anterior. B) Ccdc85 is expressed at high levels in the oocytes of both *D.pseudoobscura* and *D.virilis* however, it is not expressed in the D.wilistoni oocyte.

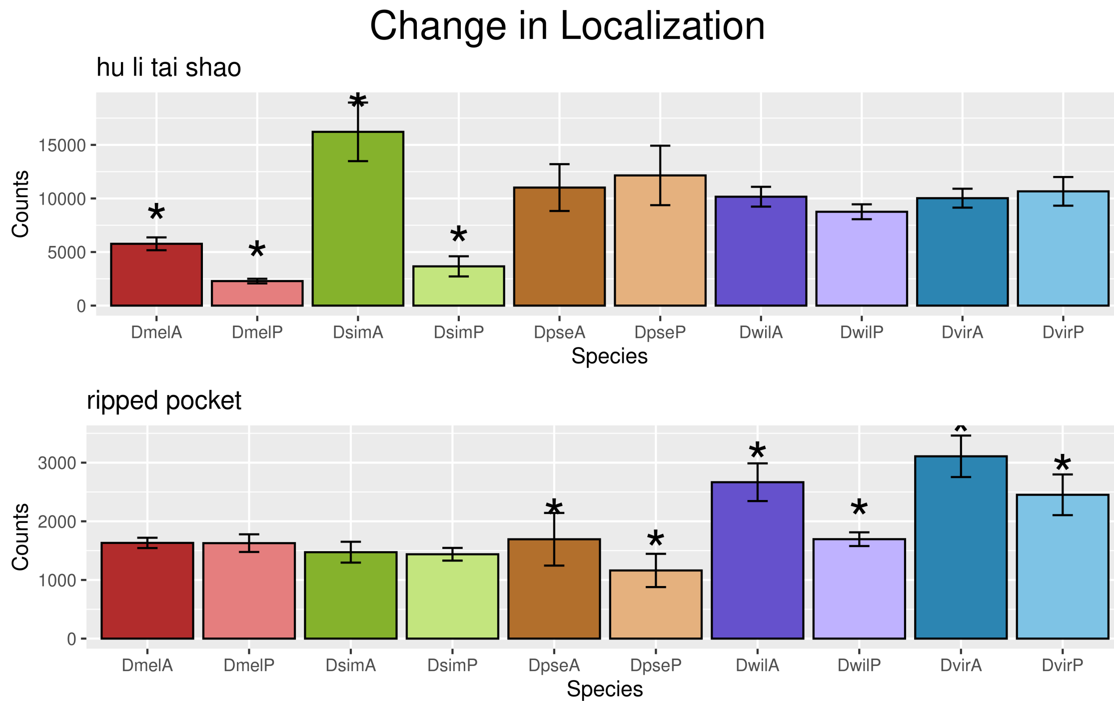
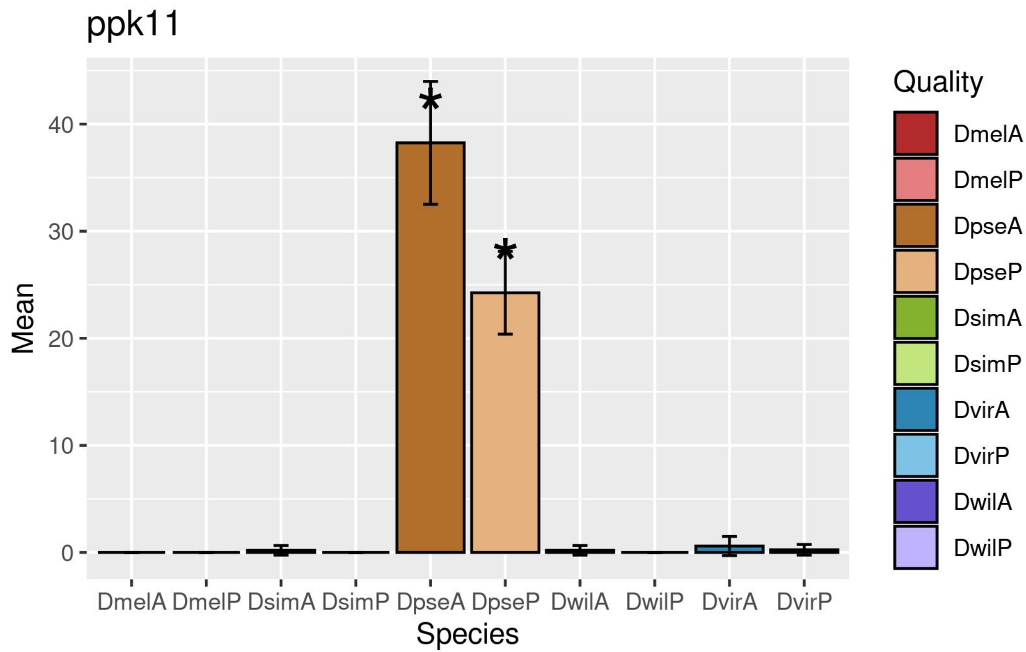
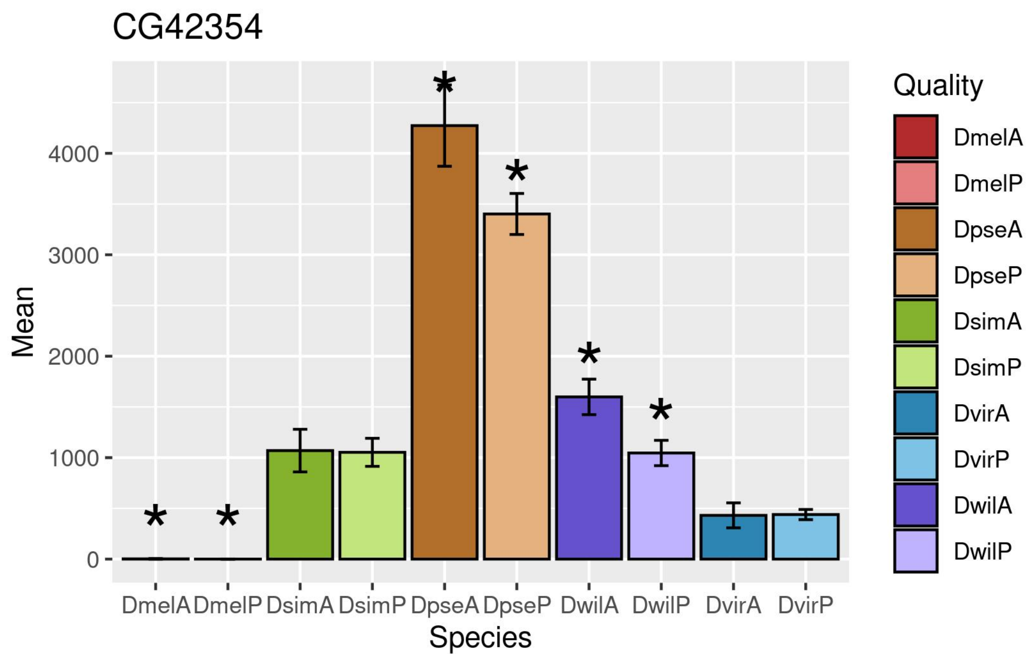


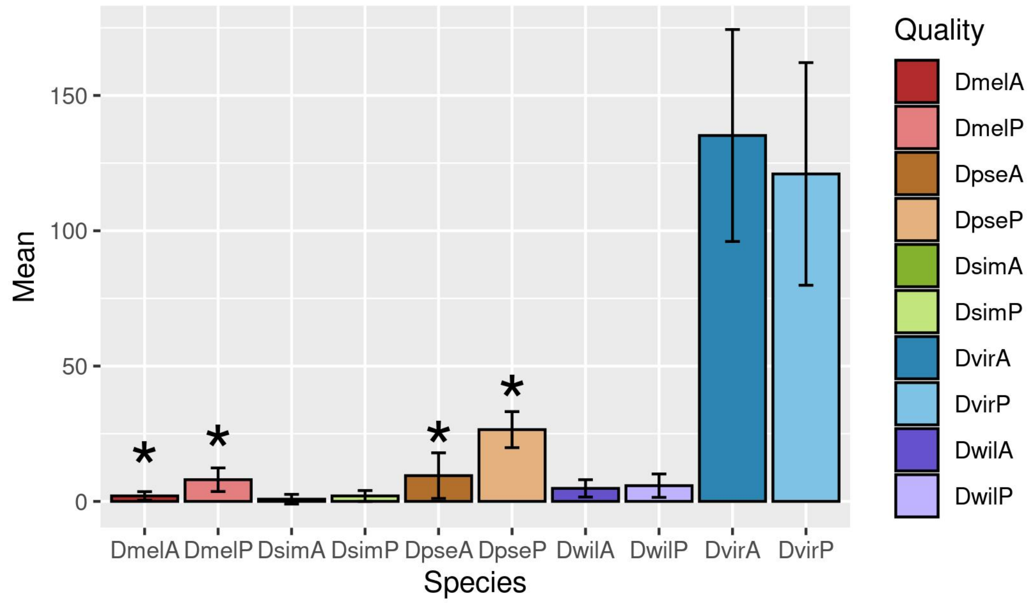
Figure 5. Changes in Localization
Examples of changes in localization. A) Hu li tai (*hts*) and B) ripped pocket (*rpk*) are expressed in the oocytes of all five species. *Hts* is enriched in the anterior of *D.melanogaster* and *D.simulans* oocytes, but lacks anterior enrichment in the remaining species. *rpk* is maternally deposited in all species, but is only enriched in the anterior of *D.pseudoobscura*, *D.wilistoni* and *D.virilis*.

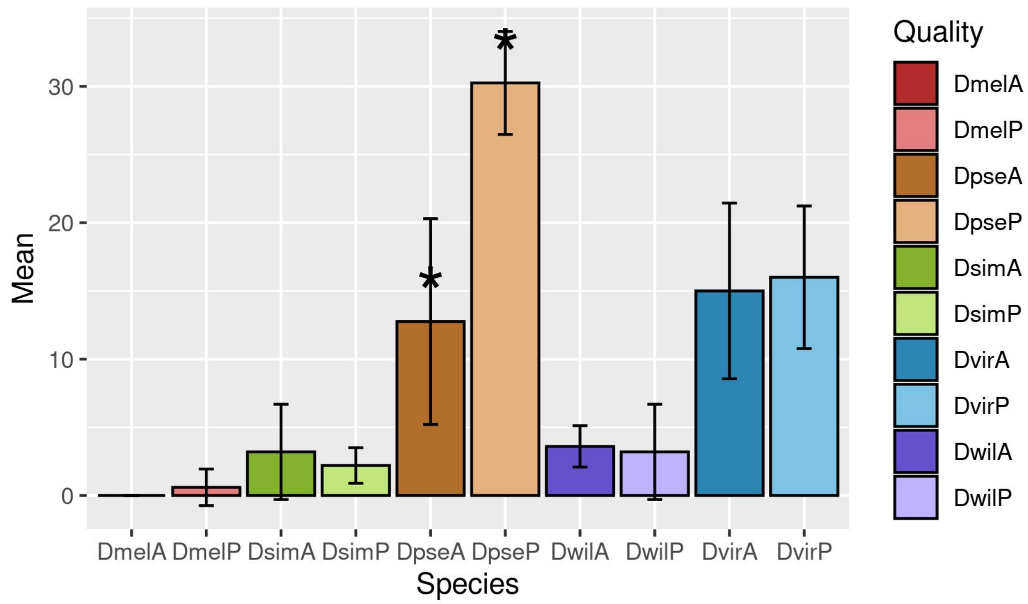
Supplemental Material

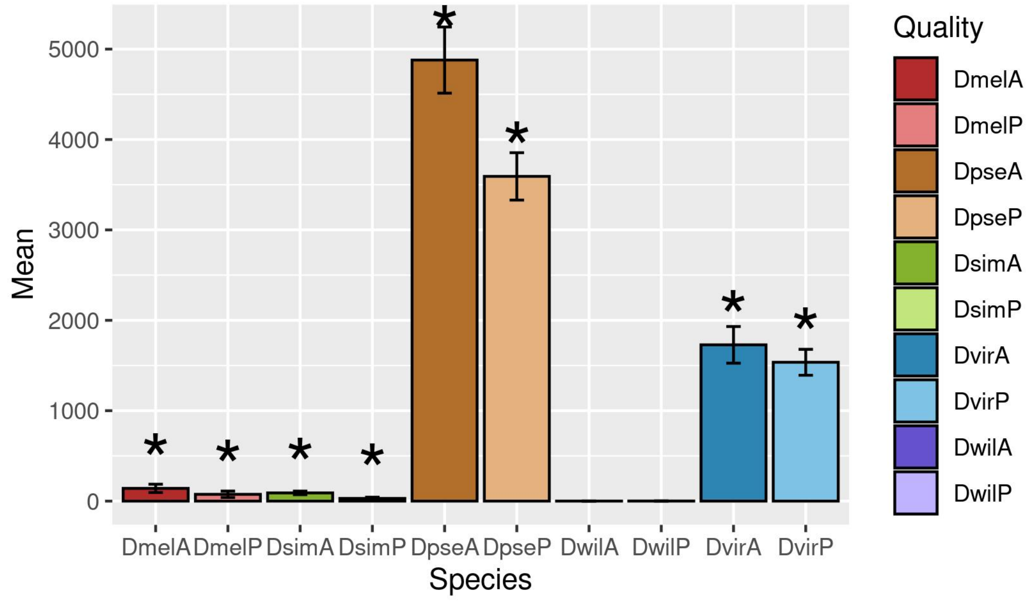
Supplemental Figures 1. Changes in Expression and Localization.

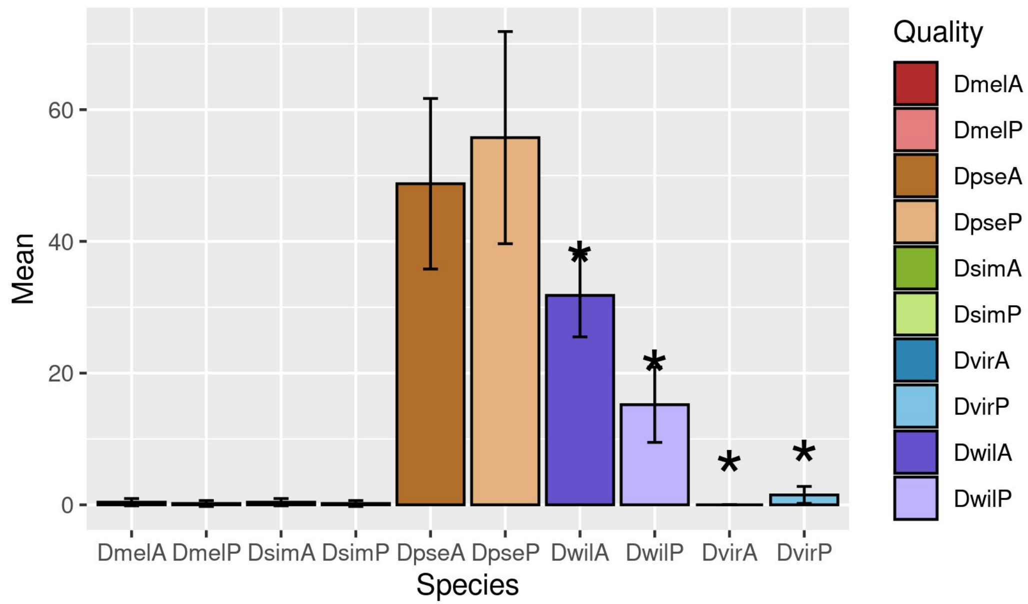


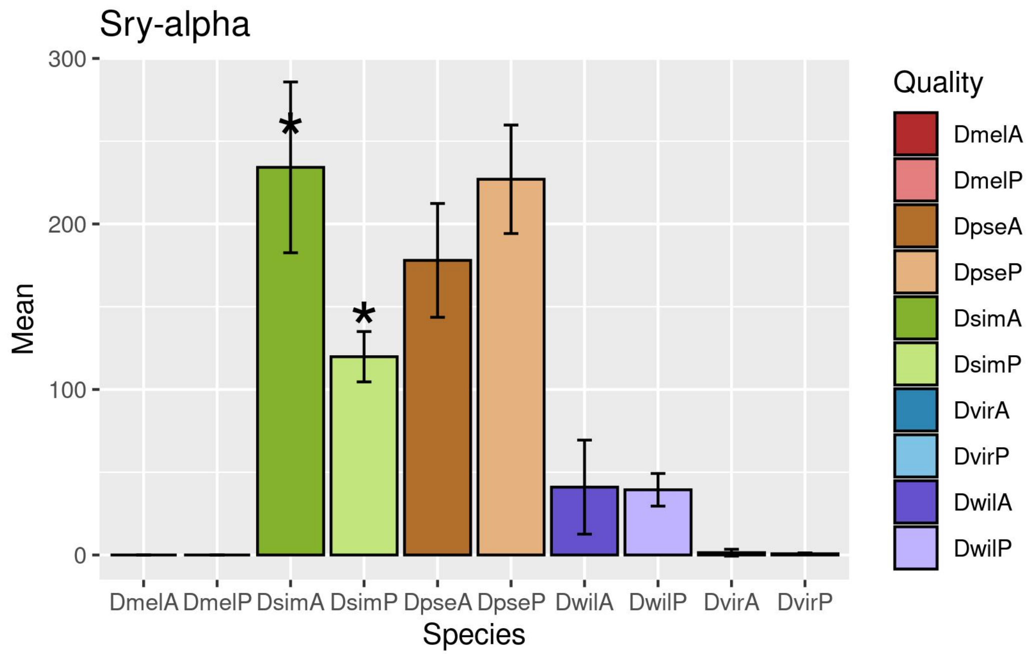


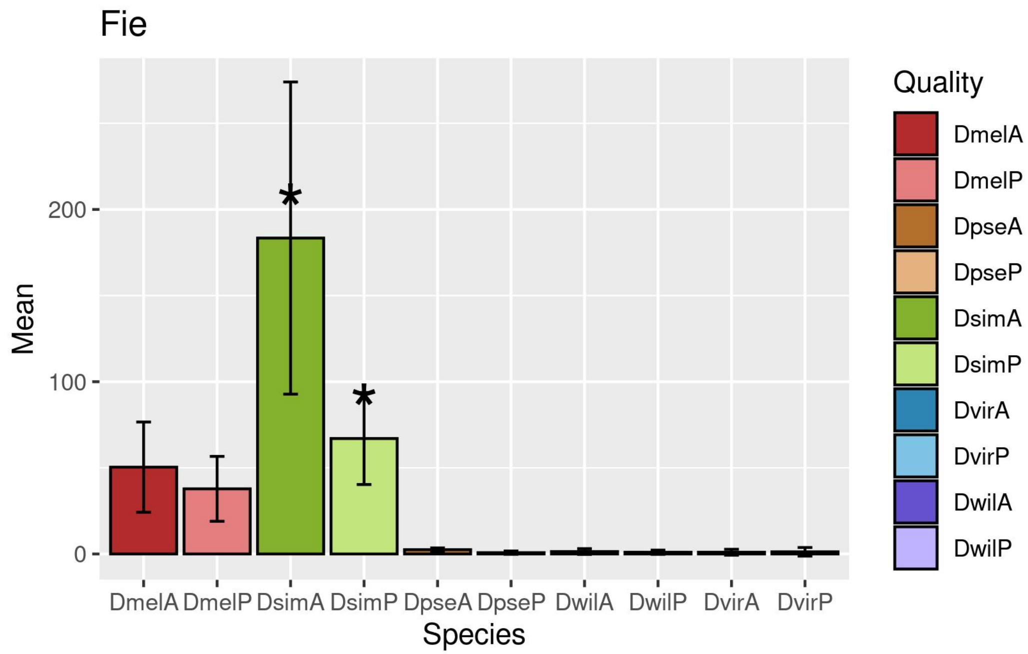

SP1173
Mean
150
100
50
0
Quality
DmelA
DmelP
DpseA
DpseP
DsimA
DsimP
DvirA
DvirP
DwilA
DwilP
DmelA DmelP DsimA DsimP DpseA DpseP DwilA DwilP DvirA DvirP
Species

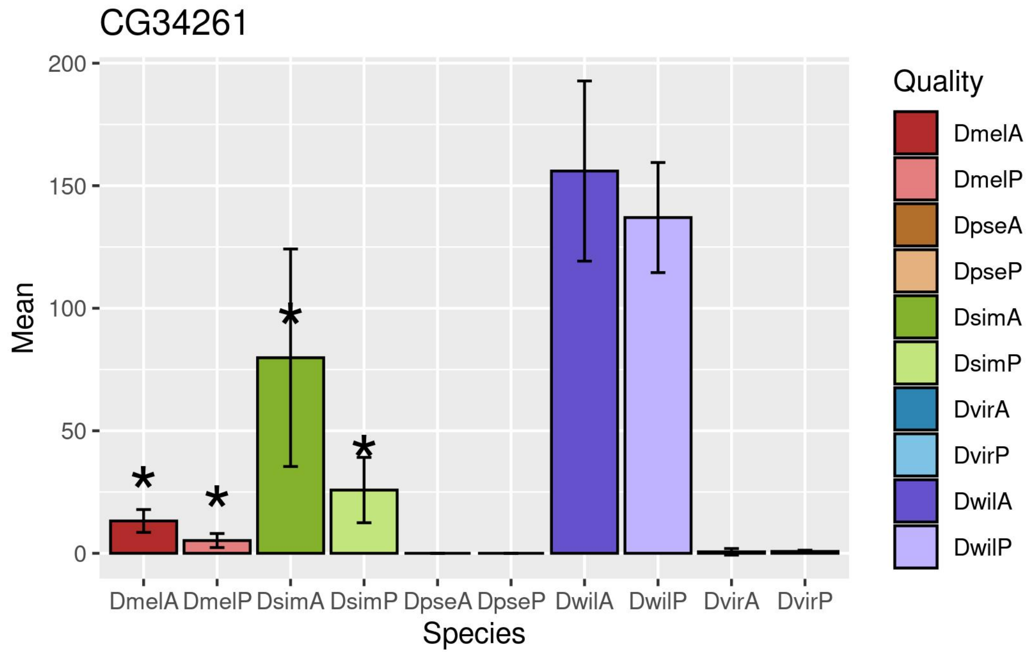

Scgbeta
Mean
30
20
10
0
Quality
DmelA
DmelP
DpseA
DpseP
DsimA
DsimP
DvirA
DvirP
DwilA
DwilP
DmelA DmelP DsimA DsimP DpseA DpseP DwilA DwilP DvirA DvirP
Species

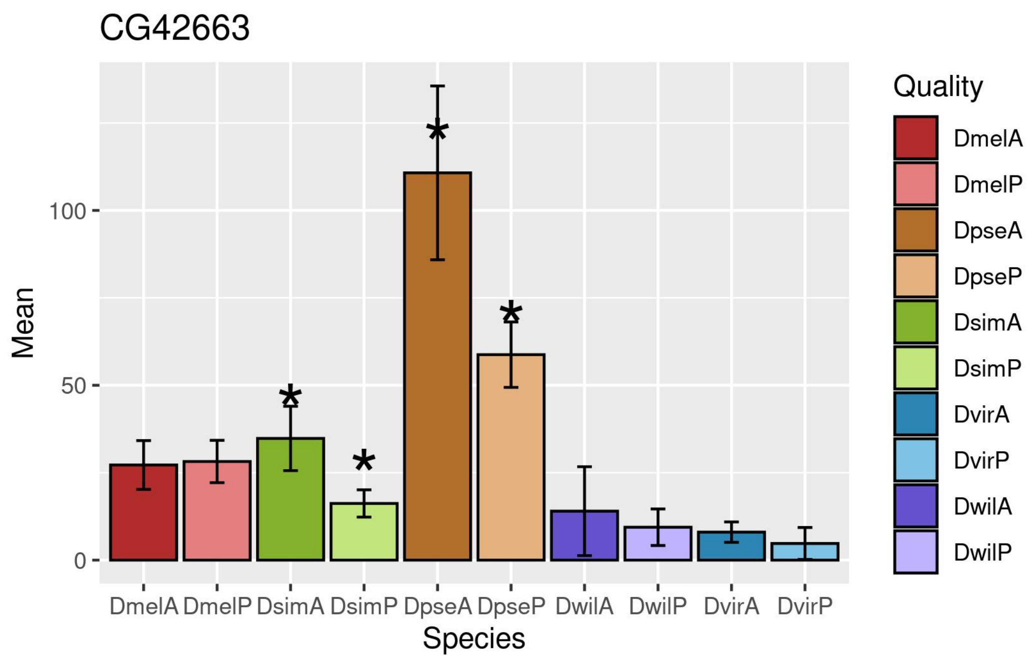

Ccdc85
Mean
5000
4000
3000
2000
1000
0
DmelA
DmelP
DsimA
DsimP
DpseA
DpseP
DwilA
DwilP
DvirA
DvirP
Species
Quality
DmelA
DmelP
DpseA
DpseP
DsimA
DsimP
DvirA
DvirP
DwilA
DwilP

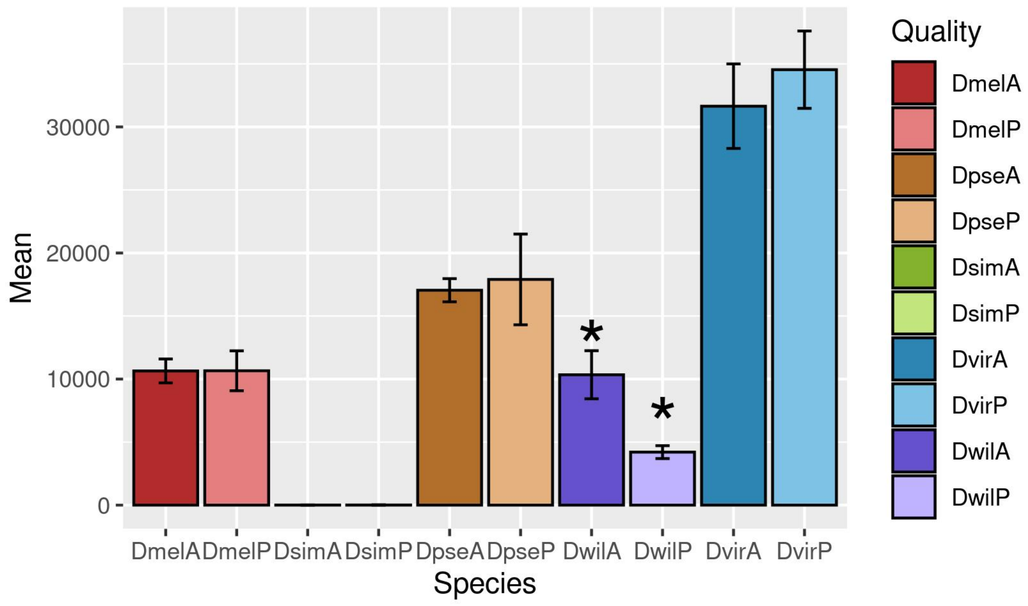

SPE
Mean
60
40
20
0
DmelA
DmelP
DsimA
DsimP
DpseA
DpseP
DwilA
DwilP
DvirA
DvirP
Species
Quality
DmelA
DmelP
DpseA
DpseP
DsimA
DsimP
DvirA
DvirP
DwilA
DwilP

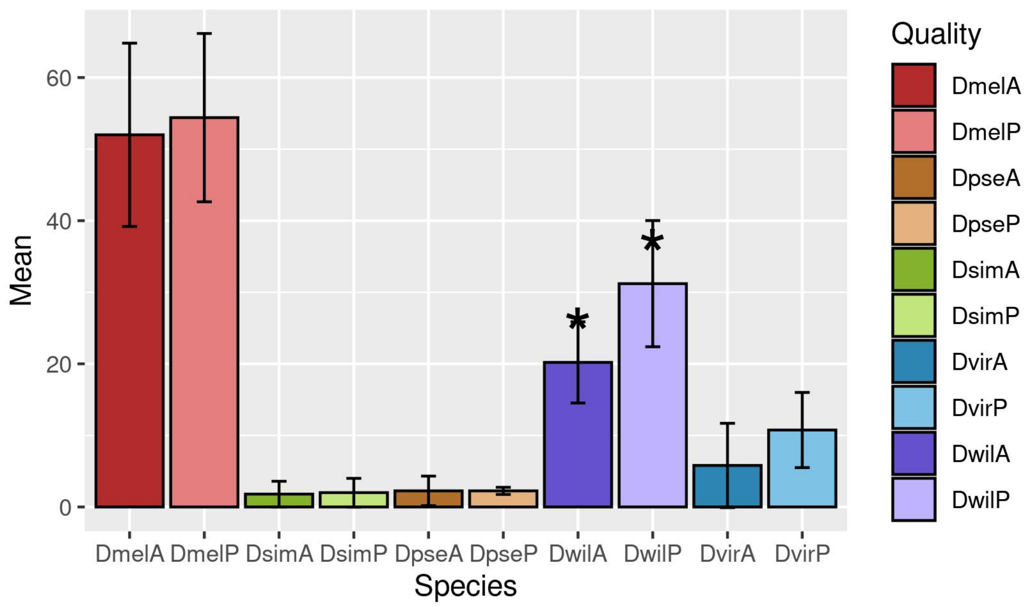

Sry-alpha
Mean
300
200
100
0
*
*
DmelA DmelP DsimA DsimP DpseA DpseP DwilA DwilP DvirA DvirP
Species
Quality
DmelA
DmelP
DpseA
DpseP
DsimA
DsimP
DvirA
DvirP
DwilA
DwilP

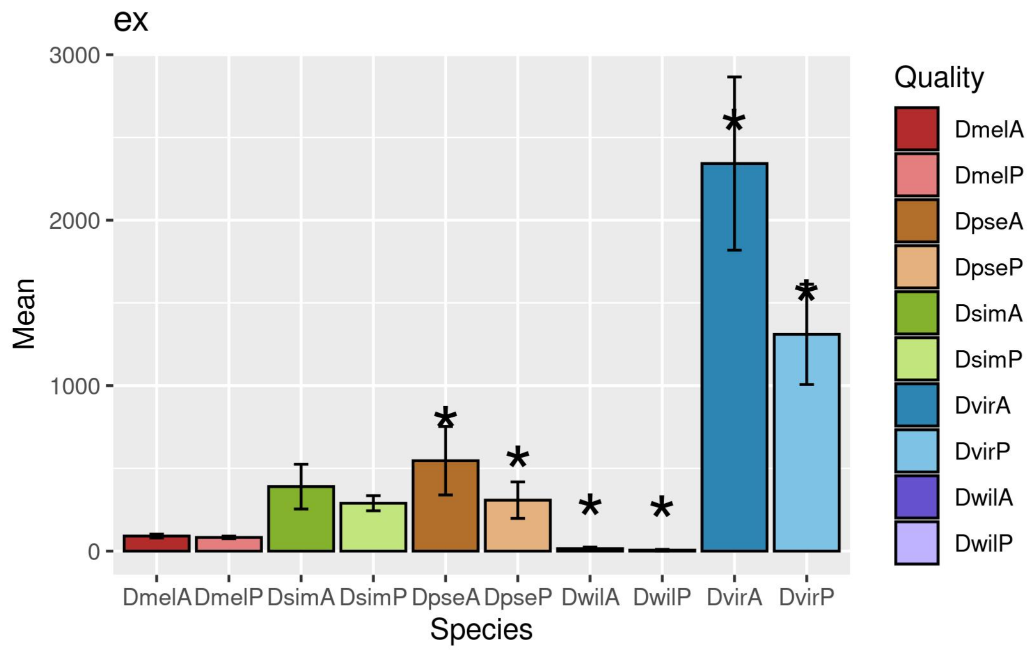

Fie
Mean
200
100
0
*
*
DmelA DmelP DsimA DsimP DpseA DpseP DwilA DwilP DvirA DvirP
Species
Quality
DmelA
DmelP
DpseA
DpseP
DsimA
DsimP
DvirA
DvirP
DwilA
DwilP


CG34261
Mean
Species
Quality
DmelA
DmelP
DpseA
DpseP
DsimA
DsimP
DvirA
DvirP
DwilA
DwilP
0
50
100
150
200

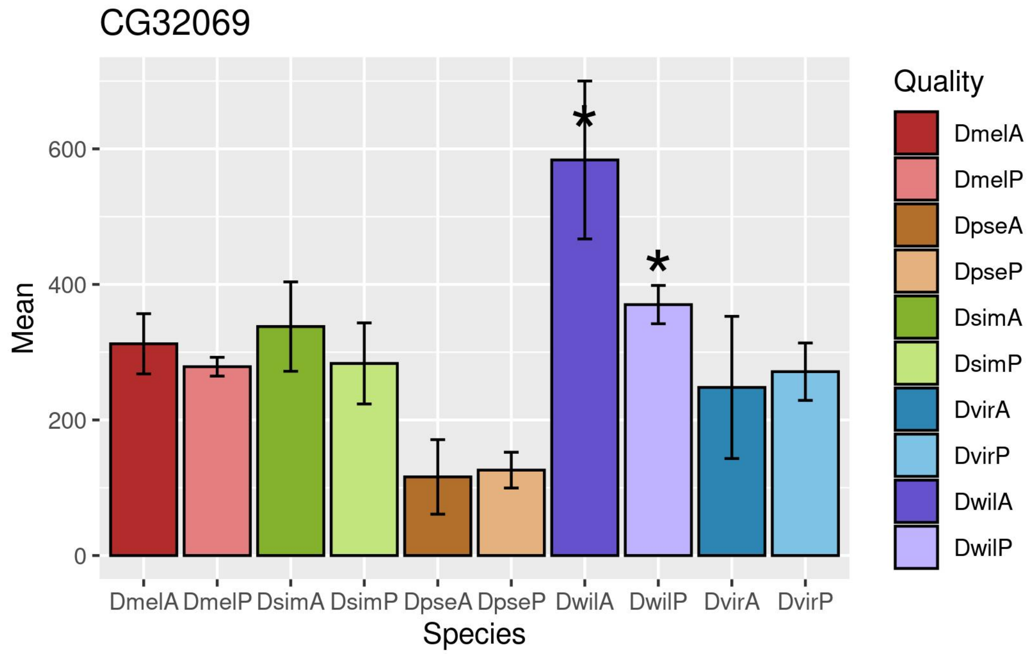

CG42663
Mean
Species
Quality
DmelA
DmelP
DpseA
DpseP
DsimA
DsimP
DvirA
DvirP
DwilA
DwilP
0
50
100

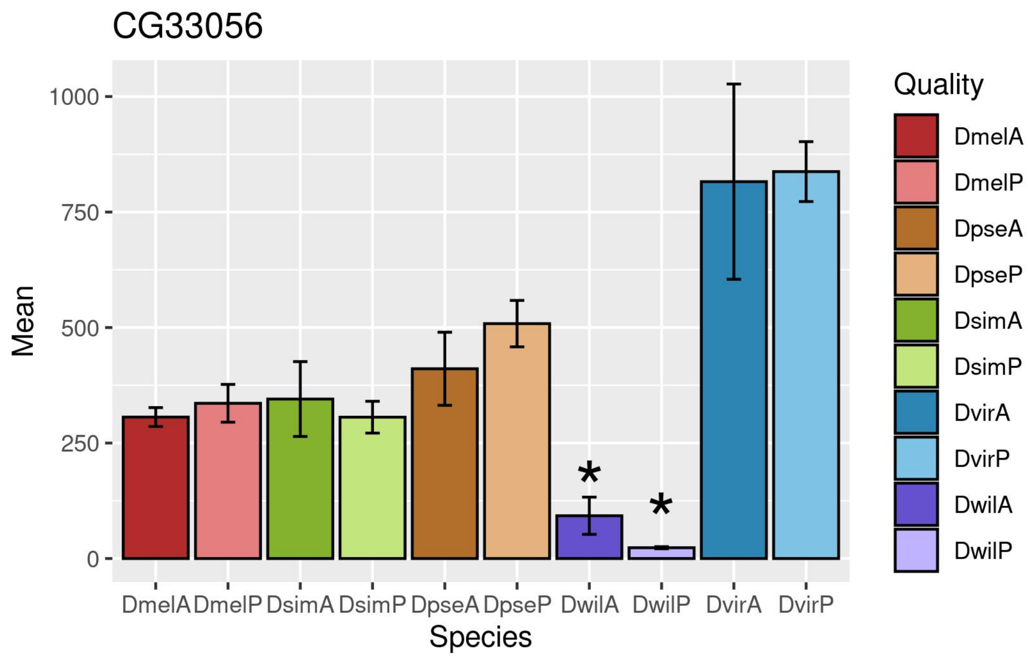

NO66
Mean
30000
20000
10000
0
*
*
DmelA DmelP DsimA DsimP DpseA DpseP DwilA DwilP DvirA DvirP
Species
Quality
DmelA
DmelP
DpseA
DpseP
DsimA
DsimP
DvirA
DvirP
DwilA
DwilP

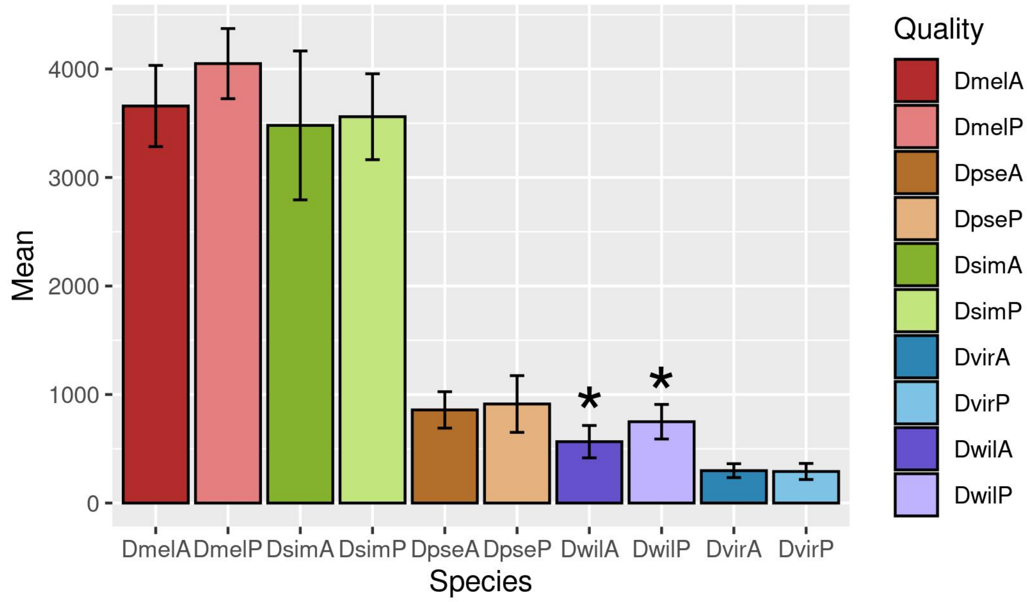

Rpt3R
Mean
60
40
20
0
*
*
DmelA DmelP DsimA DsimP DpseA DpseP DwilA DwilP DvirA DvirP
Species
Quality
DmelA
DmelP
DpseA
DpseP
DsimA
DsimP
DvirA
DvirP
DwilA
DwilP

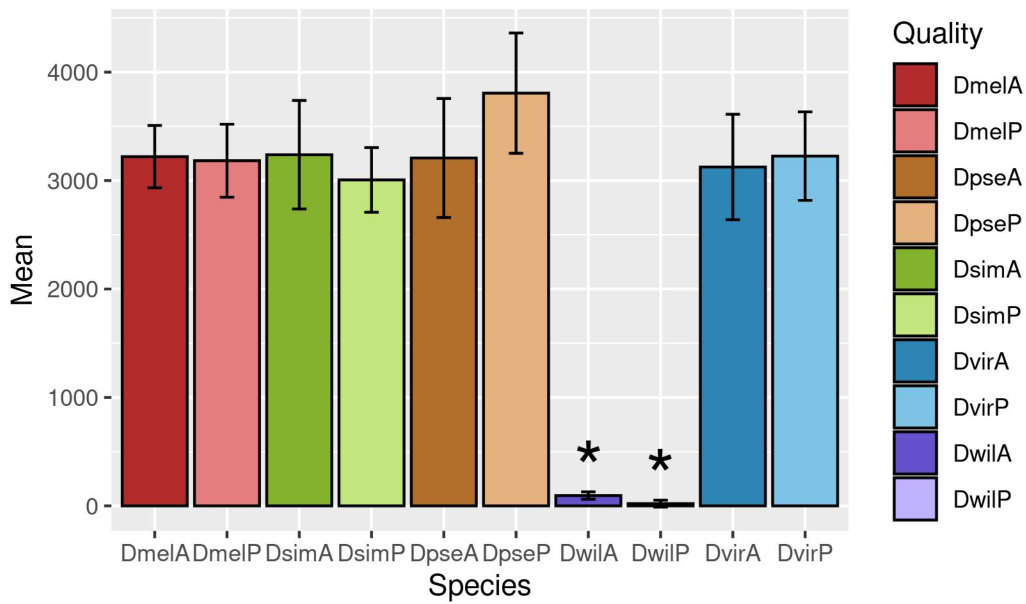

ex
3000
2000
1000
0
Mean
DmelA
DmelP
DsimA
DsimP
DpseA
DpseP
DwilA
DwilP
DvirA
DvirP
Species
Quality
DmelA
DmelP
DpseA
DpseP
DsimA
DsimP
DvirA
DvirP
DwilA
DwilP

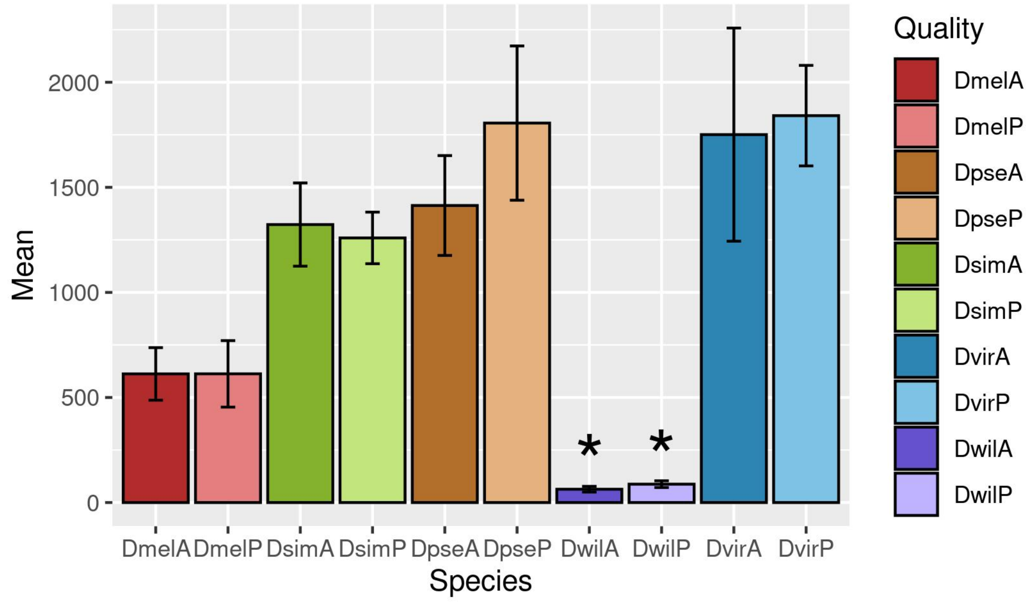
Supplemental Figures 2. Changes in Localization.

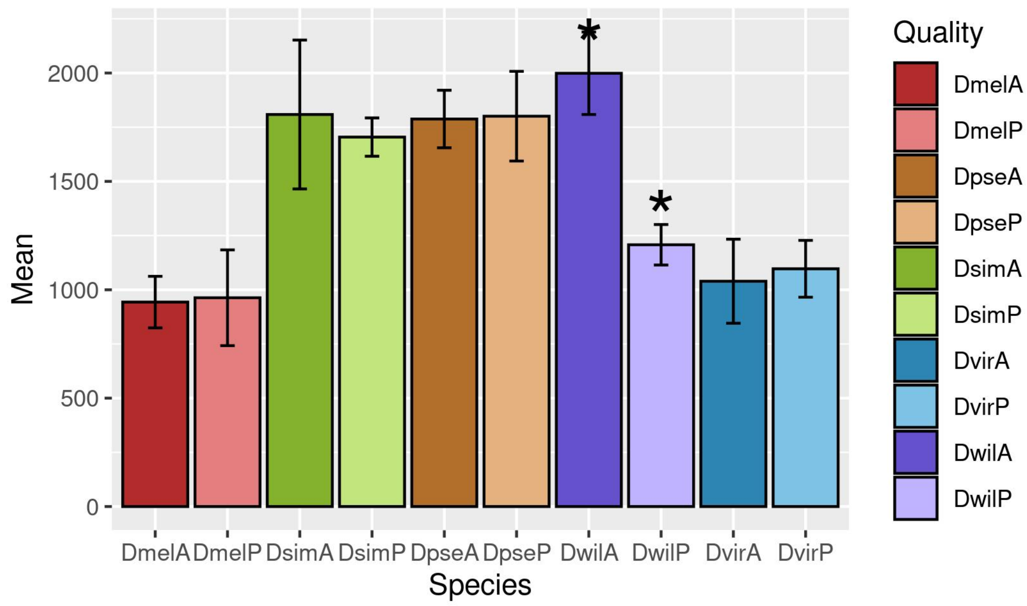


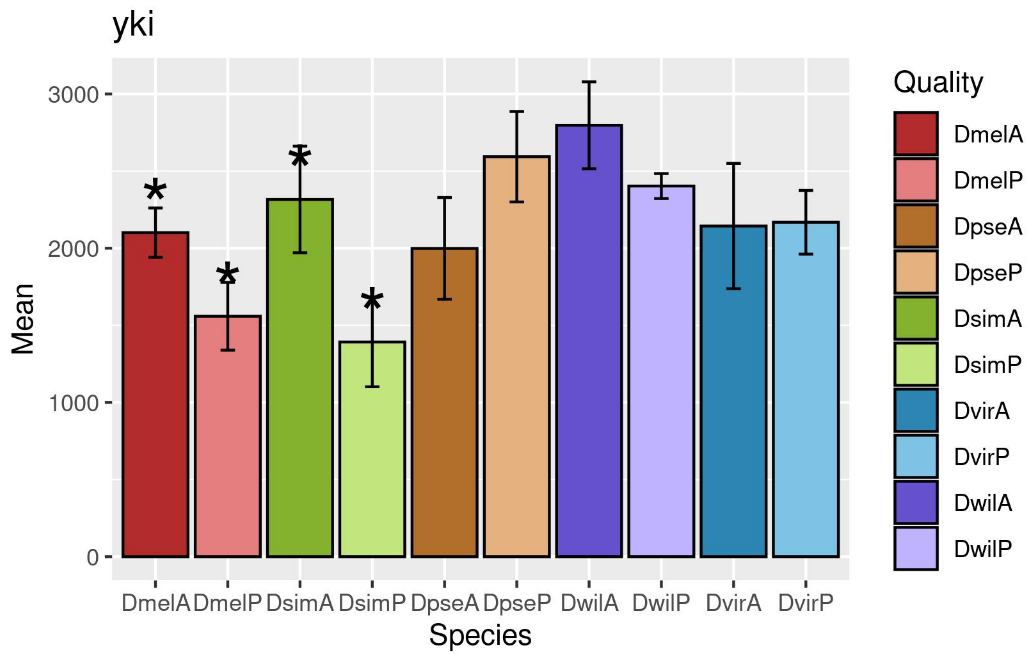


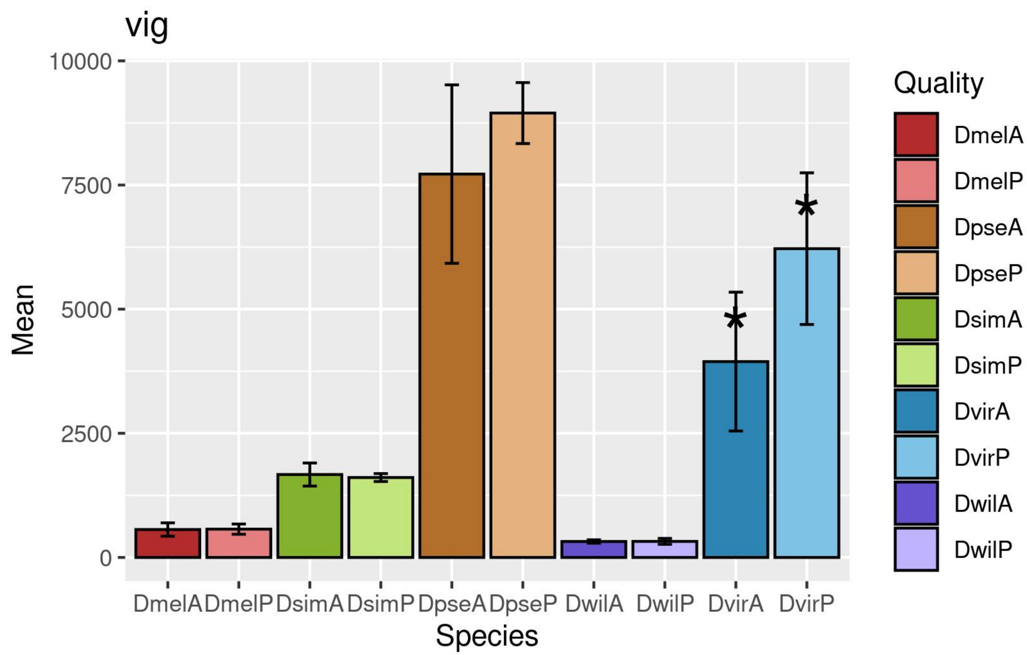

PI4KIIIalpha
Mean
4000
3000
2000
1000
0
DmelA
DmelP
DsimA
DsimP
DpseA
DpseP
DwilA
DwilP
DvirA
DvirP
Species
Quality
DmelA
DmelP
DpseA
DpseP
DsimA
DsimP
DvirA
DvirP
DwilA
DwilP

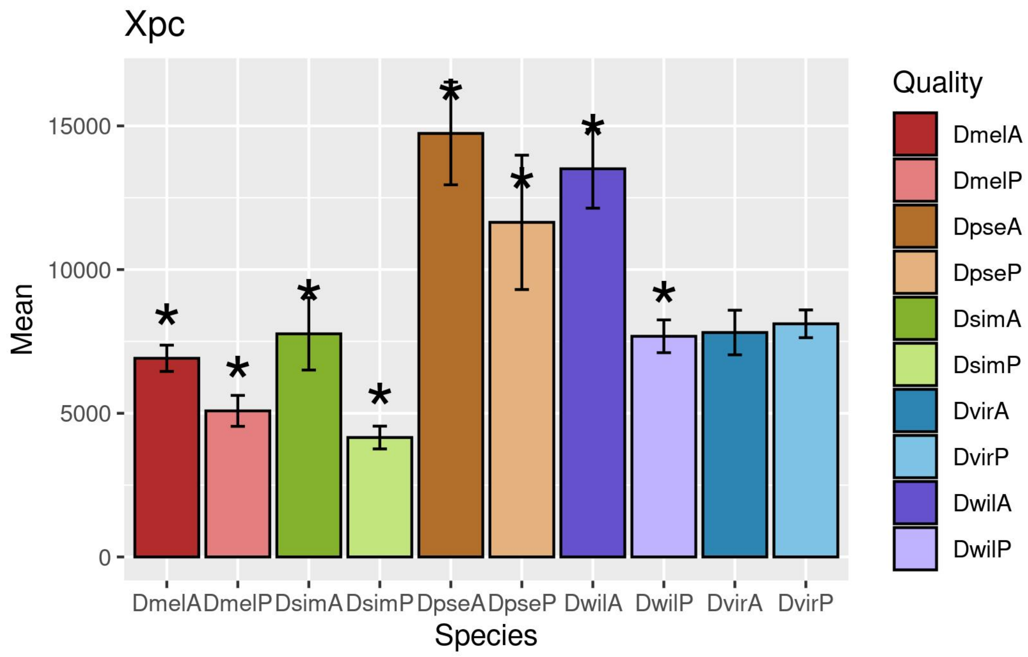

CG9911
Mean
4000
3000
2000
1000
0
DmelA
DmelP
DsimA
DsimP
DpseA
DpseP
DwilA
DwilP
DvirA
DvirP
Species
Quality
DmelA
DmelP
DpseA
DpseP
DsimA
DsimP
DvirA
DvirP
DwilA
DwilP

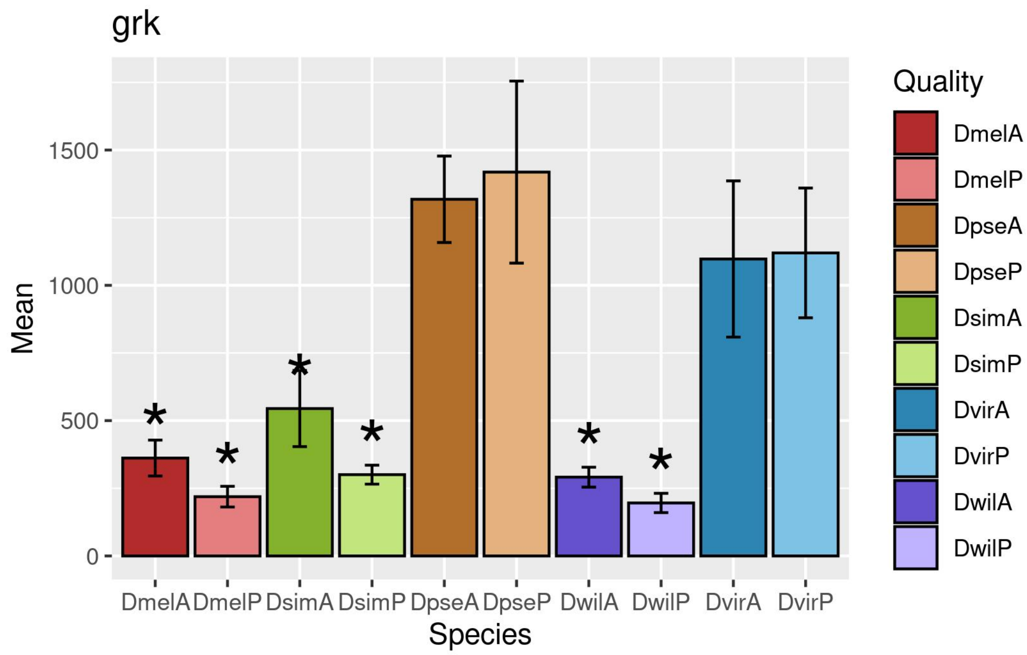

CG15019
Quality
DmelA
DmelP
DpseA
DpseP
DsimA
DsimP
DvirA
DvirP
DwilA
DwilP
Mean
2000
1500
1000
500
0
*
*
DmelA DmelP DsimA DsimP DpseA DpseP DwilA DwilP DvirA DvirP
Species

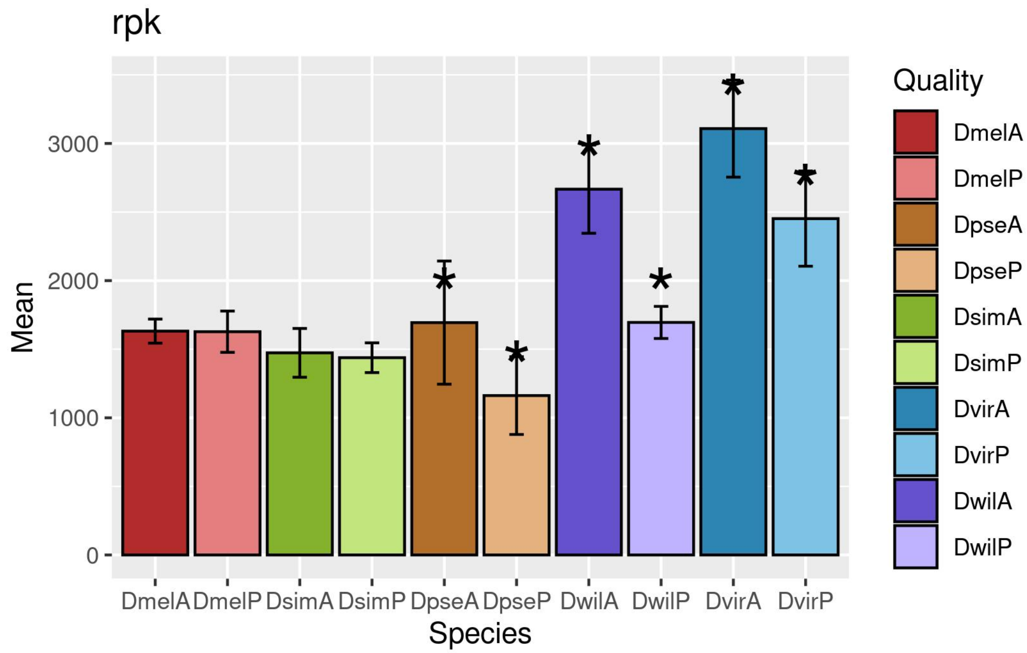

CG6425
Quality
DmelA
DmelP
DpseA
DpseP
DsimA
DsimP
DvirA
DvirP
DwilA
DwilP
Mean
2000
1500
1000
500
0
*
*
DmelA DmelP DsimA DsimP DpseA DpseP DwilA DwilP DvirA DvirP
Species


yki
Quality
DmelA
DmelP
DpseA
DpseP
DsimA
DsimP
DvirA
DvirP
DwilA
DwilP
Mean
3000
2000
1000
0
DmelA DmelP DsimA DsimP DpseA DpseP DwilA DwilP DvirA DvirP
Species

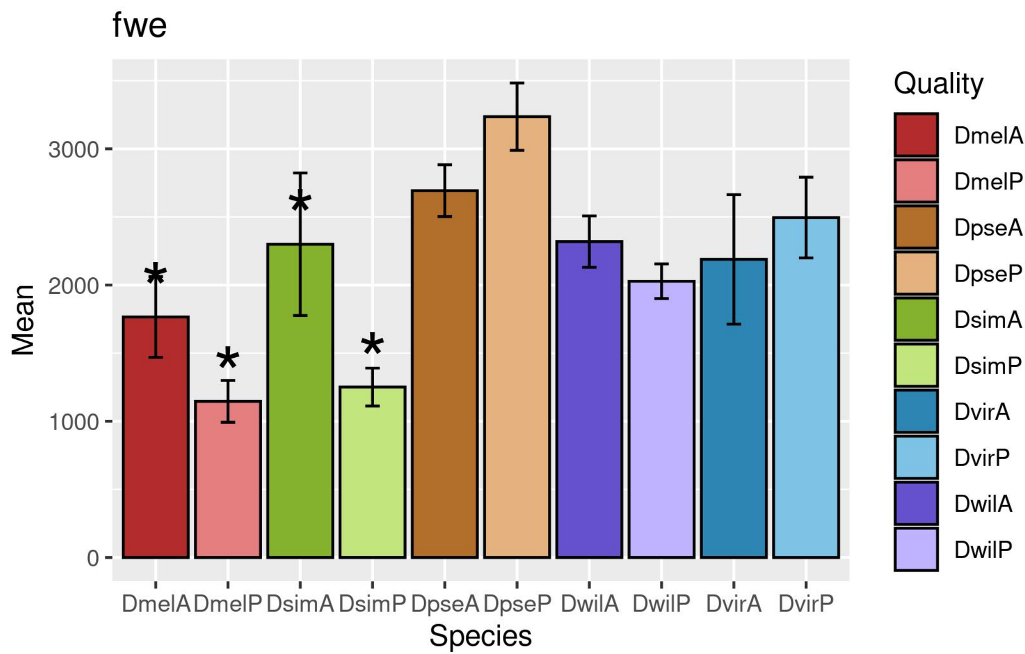

vig
Quality
DmelA
DmelP
DpseA
DpseP
DsimA
DsimP
DvirA
DvirP
DwilA
DwilP
Mean
10000
7500
5000
2500
0
DmelA DmelP DsimA DsimP DpseA DpseP DwilA DwilP DvirA DvirP
Species


Xpc
Quality
DmelA
DmelP
DpseA
DpseP
DsimA
DsimP
DvirA
DvirP
DwilA
DwilP
Mean
15000
10000
5000
0
DmelA DmelP DsimA DsimP DpseA DpseP DwilA DwilP DvirA DvirP
Species

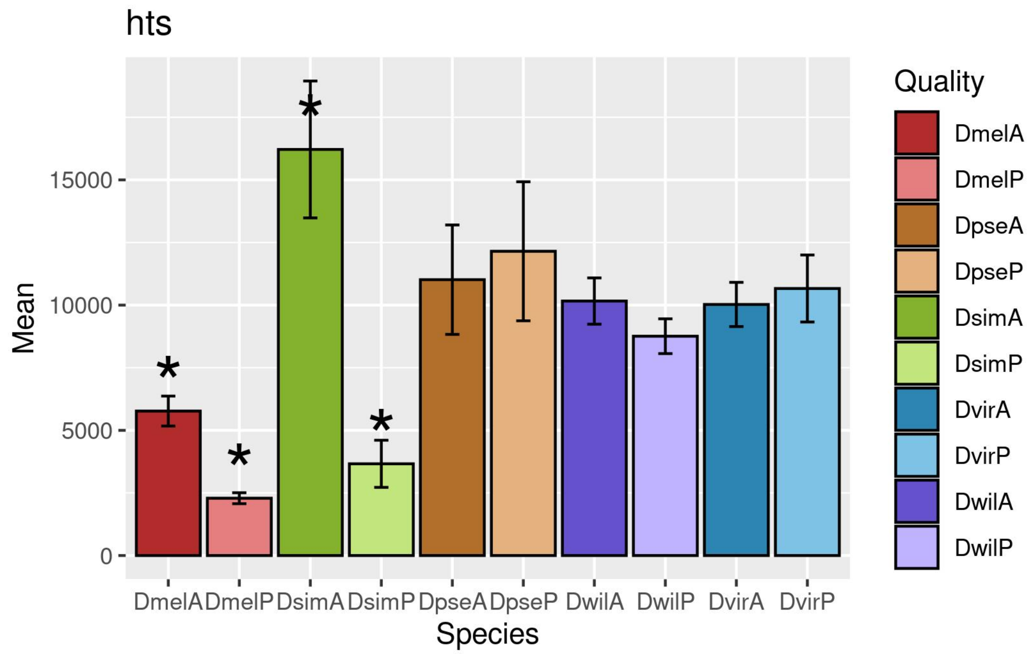

grk
Quality
DmelA
DmelP
DpseA
DpseP
DsimA
DsimP
DvirA
DvirP
DwilA
DwilP
Mean
1500
1000
500
0
DmelA DmelP DsimA DsimP DpseA DpseP DwilA DwilP DvirA DvirP
Species

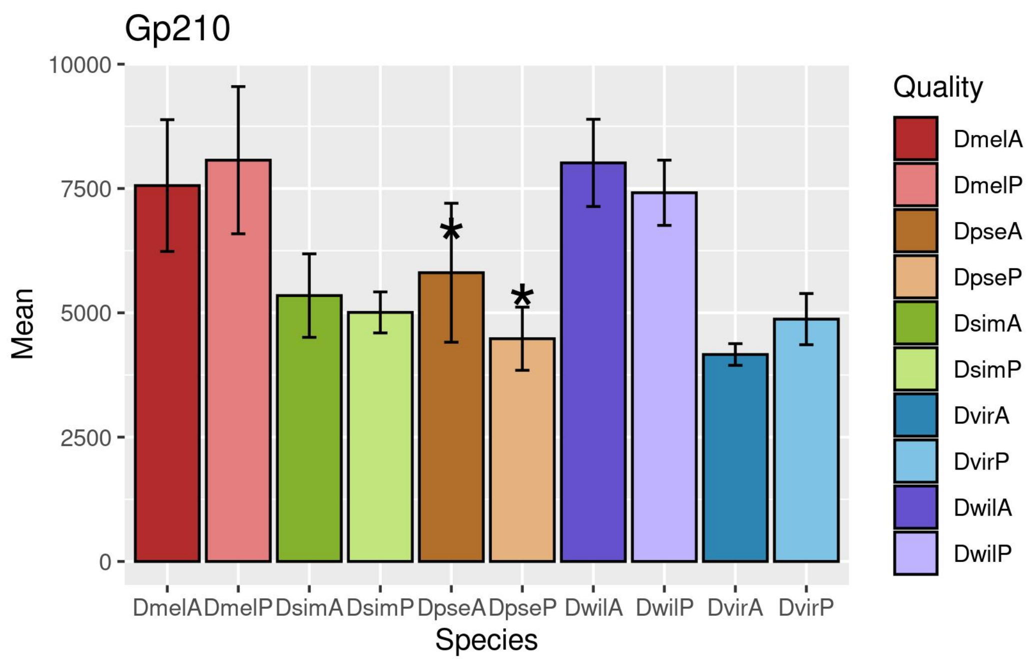


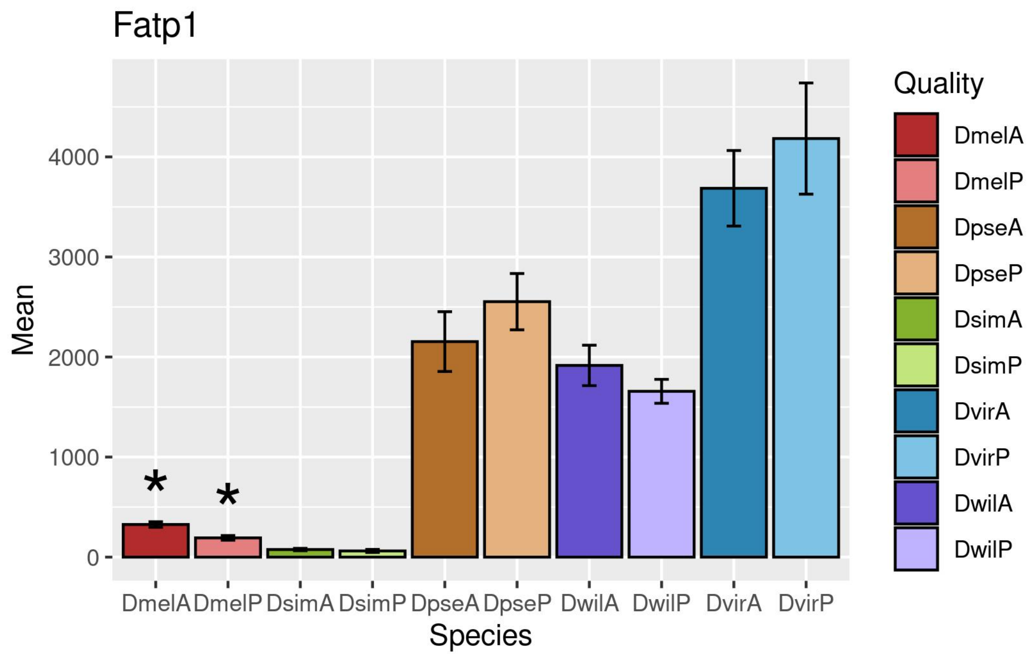
rpk
Quality
DmelA
DmelP
DpseA
DpseP
DsimA
DsimP
DvirA
DvirP
DwilA
DwilP
Mean
3000
2000
1000
0
DmelA DmelP DsimA DsimP DpseA DpseP DwilA DwilP DvirA DvirP
Species

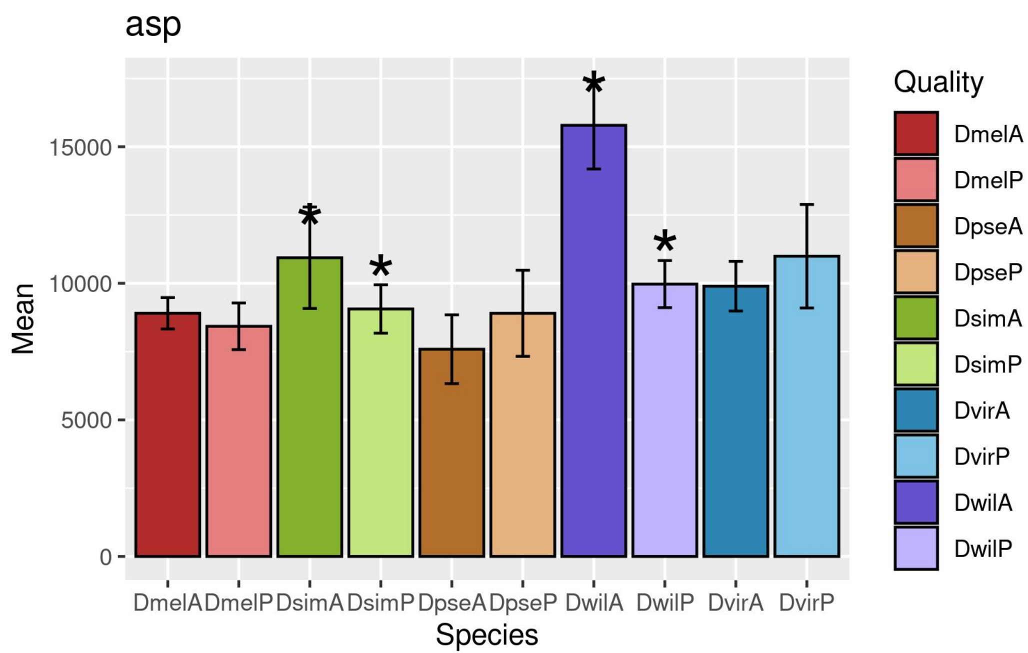


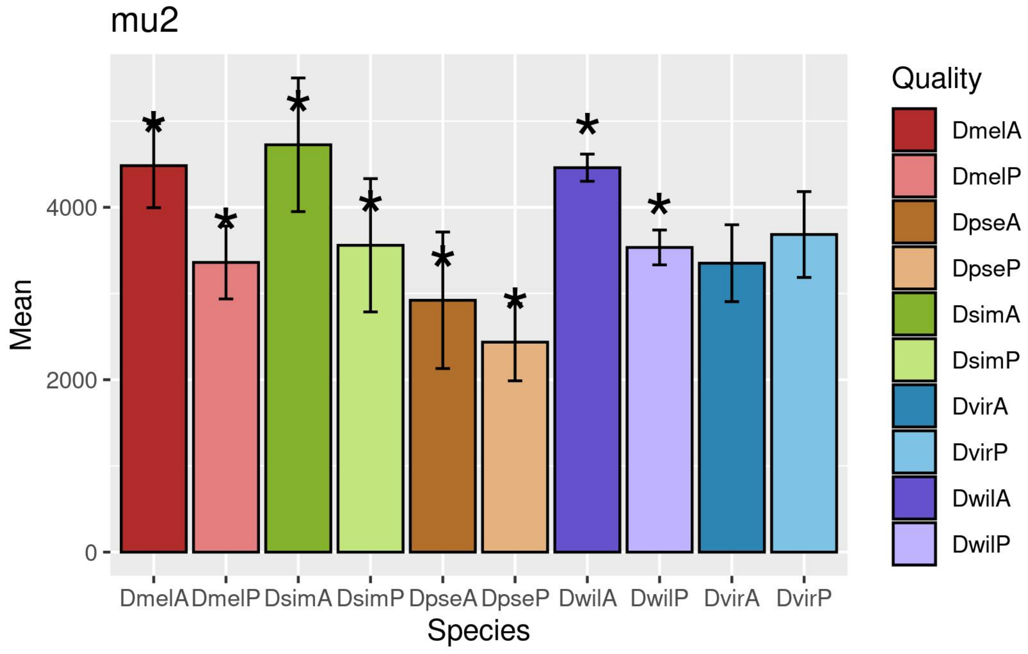
fwe
Quality
DmelA
DmelP
DpseA
DpseP
DsimA
DsimP
DvirA
DvirP
DwilA
DwilP
Mean
3000
2000
1000
0
DmelA DmelP DsimA DsimP DpseA DpseP DwilA DwilP DvirA DvirP
Species

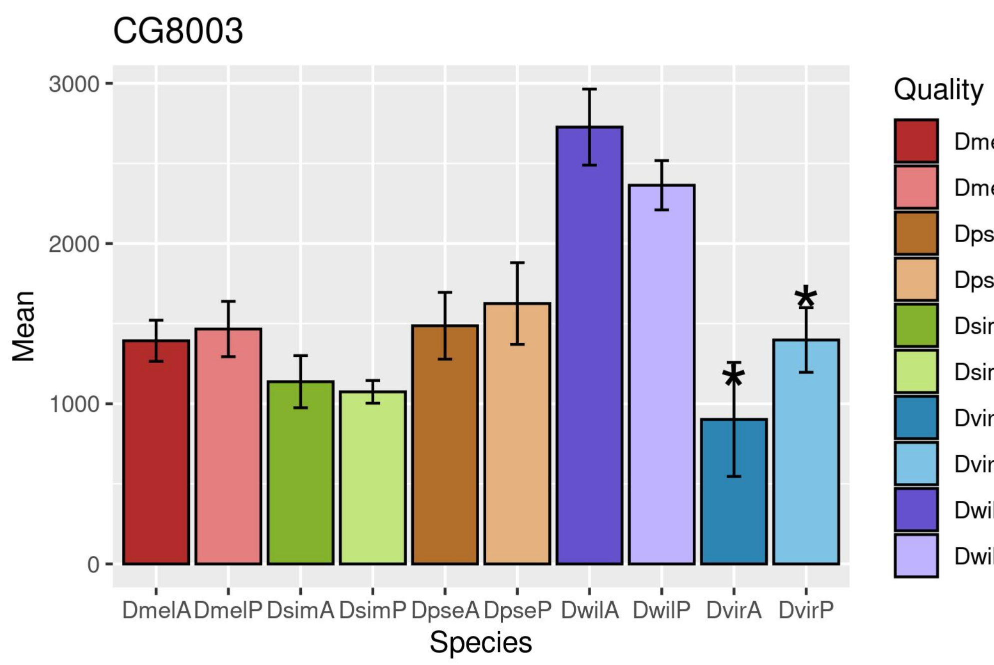

hts
Mean
15000
10000
5000
0
Species
DmelA DmelP DsimA DsimP DpseA DpseP DwilA DwilP DvirA DvirP
Quality
DmelA
DmelP
DpseA
DpseP
DsimA
DsimP
DvirA
DvirP
DwilA
DwilP

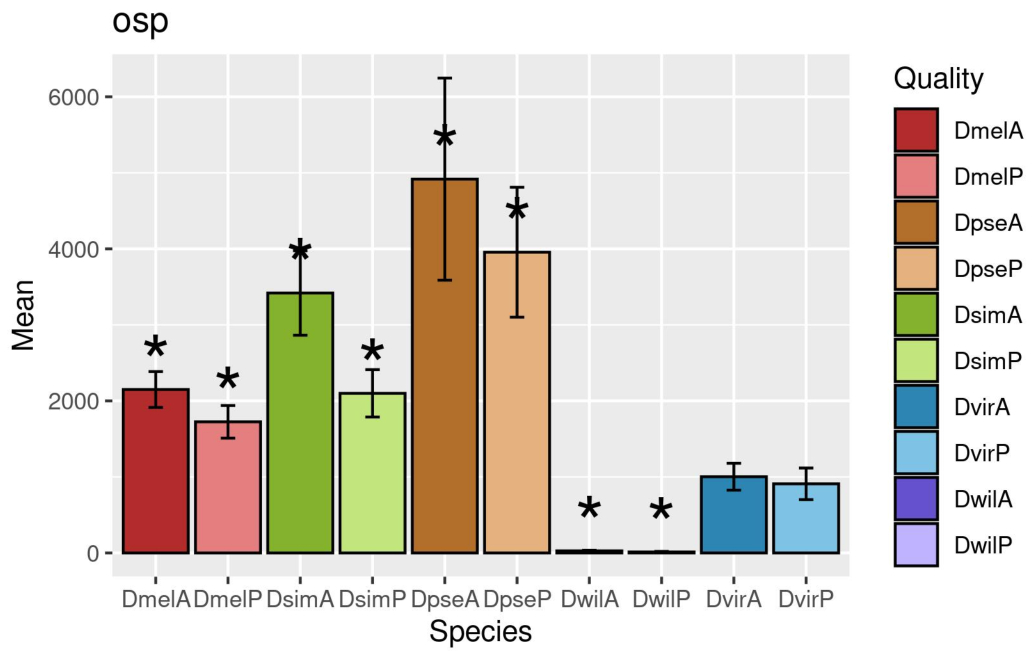

Gp210
Mean
10000
7500
5000
2500
0
Species
DmelA DmelP DsimA DsimP DpseA DpseP DwilA DwilP DvirA DvirP
Quality
DmelA
DmelP
DpseA
DpseP
DsimA
DsimP
DvirA
DvirP
DwilA
DwilP

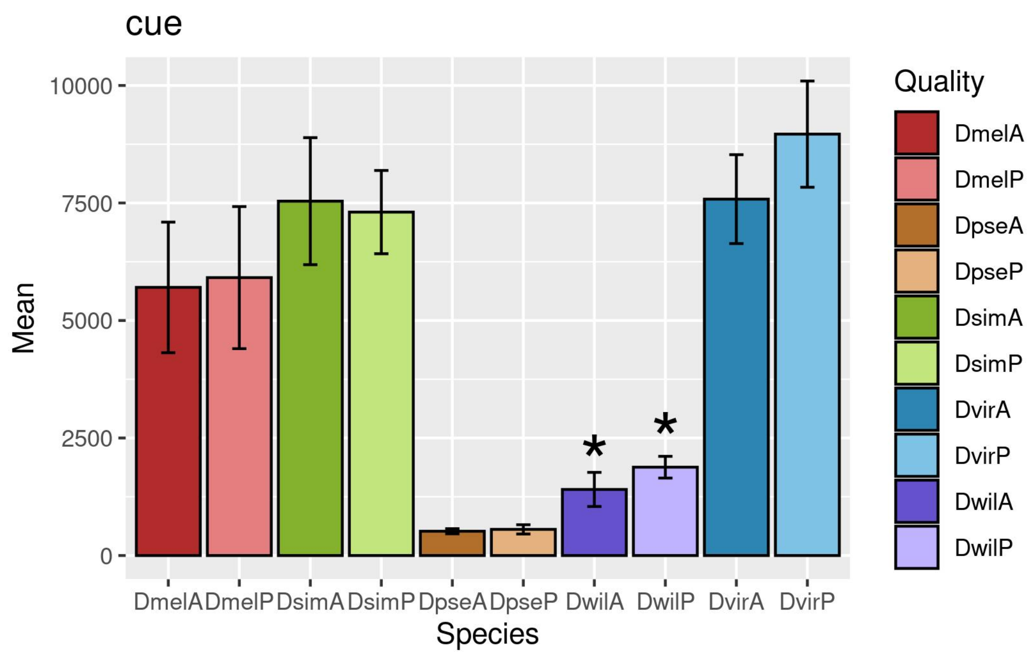

Fatp1
Mean
Species
4000
3000
2000
1000
0
DmelA DmelP DsimA DsimP DpseA DpseP DwilA DwilP DvirA DvirP
Quality
DmelA
DmelP
DpseA
DpseP
DsimA
DsimP
DvirA
DvirP
DwilA
DwilP

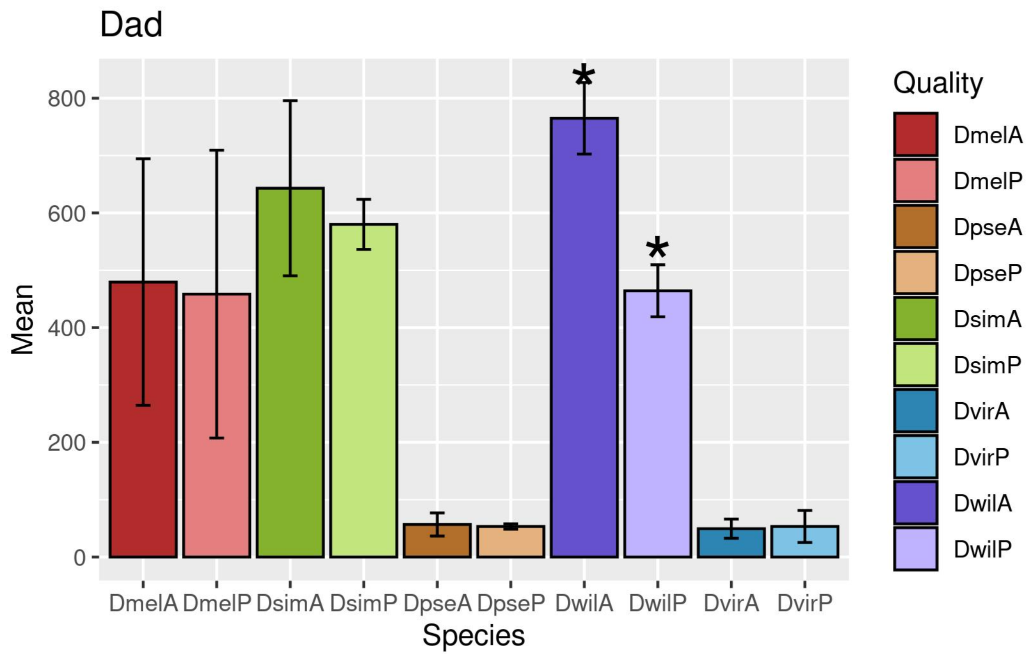

asp
Mean
Species
15000
10000
5000
0
DmelA DmelP DsimA DsimP DpseA DpseP DwilA DwilP DvirA DvirP
Quality
DmelA
DmelP
DpseA
DpseP
DsimA
DsimP
DvirA
DvirP
DwilA
DwilP

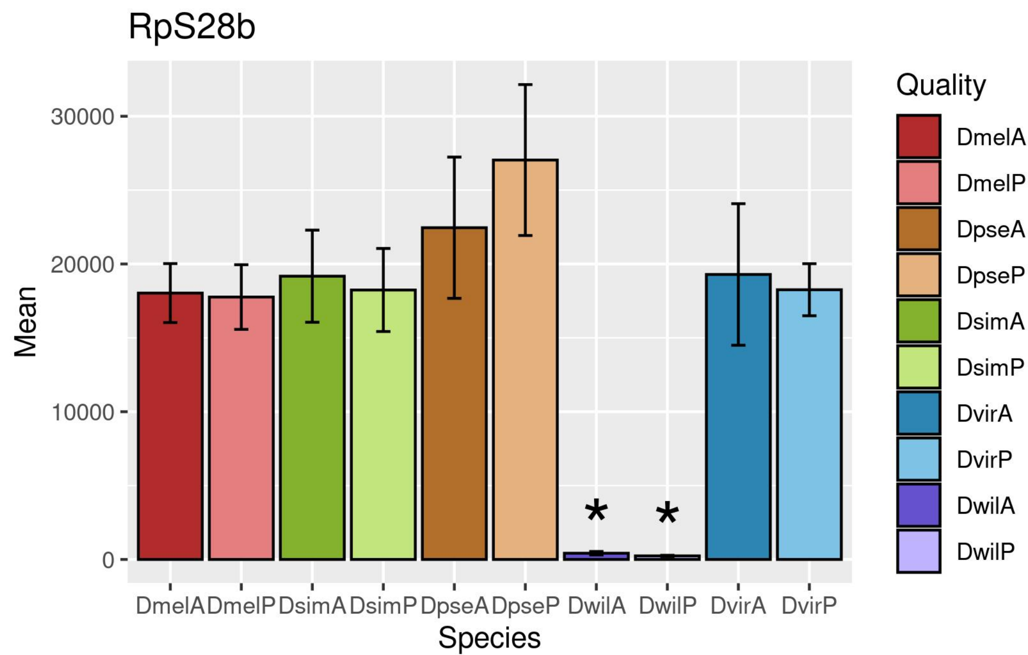

mu2
Mean
4000
2000
0
DmelA DmelP DsimA DsimP DpseA DpseP DwilA DwilP DvirA DvirP
Species
Quality
DmelA
DmelP
DpseA
DpseP
DsimA
DsimP
DvirA
DvirP
DwilA
DwilP

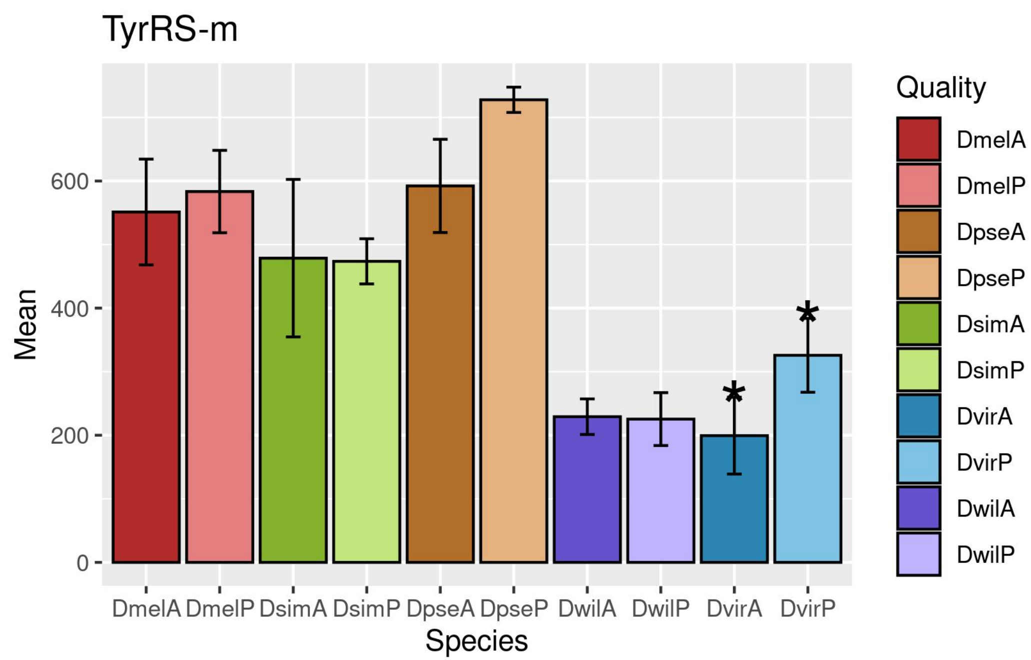

CG8003
Mean
3000
2000
1000
0
DmelA DmelP DsimA DsimP DpseA DpseP DwilA DwilP DvirA DvirP
Species
Quality
DmelA
DmelP
DpseA
DpseP
DsimA
DsimP
DvirA
DvirP
DwilA
DwilP


osp
Mean
6000
4000
2000
0
DmelA DmelP DsimA DsimP DpseA DpseP DwilA DwilP DvirA DvirP
Species
Quality
DmelA
DmelP
DpseA
DpseP
DsimA
DsimP
DvirA
DvirP
DwilA
DwilP


cue
Mean
10000
7500
5000
2500
0
DmelA DmelP DsimA DsimP DpseA DpseP DwilA DwilP DvirA DvirP
Species
Quality
DmelA
DmelP
DpseA
DpseP
DsimA
DsimP
DvirA
DvirP
DwilA
DwilP


Dad
Quality
DmelA
DmelP
DpseA
DpseP
DsimA
DsimP
DvirA
DvirP
DwilA
DwilP
800
600
400
200
0
Mean
DmelA DmelP DsimA DsimP DpseA DpseP DwilA DwilP DvirA DvirP
Species
*
*


RpS28b
Quality
DmelA
DmelP
DpseA
DpseP
DsimA
DsimP
DvirA
DvirP
DwilA
DwilP
30000
20000
10000
0
Mean
DmelA DmelP DsimA DsimP DpseA DpseP DwilA DwilP DvirA DvirP
Species
*
*


TyrRS-m
Mean
600
400
200
0
DmelA DmelP DsimA DsimP DpseA DpseP DwilA DwilP DvirA DvirP
Species
*
*
Quality
DmelA
DmelP
DpseA
DpseP
DsimA
DsimP
DvirA
DvirP
DwilA
DwilP

Conclusion

The earliest events of embryogenesis are crucial for overall organismal fitness. Starting from a single cell, the correct trajectory of events must occur to form an adult organism. In many animals, maternal mRNAs must be properly localized within the oocyte during oogenesis for proper development after fertilization. One crucial role for localized maternal mRNAs is the axis patterning of the embryo. In Drosophila, a high percentage of expressed transcripts are localized at one point throughout oogenesis. Additionally, a large proportion of the genome, ~75%, is maternally deposited. These maternal inputs drive the earliest events of embryogenesis before the zygotic genome is activated to carry on the development of the embryo. The maternal and zyogotic genomes must orchestrate this hand off of developmental control and therefore must be under strong evolutionary constraint. Previous work has shown that changes in the localization of maternal mRNAs critical for early development have occurred between genera in Diptera. Additionally, maternal and zygotic transcriptomes are highly conserved within the Drosophila genus, yet differences in expression do evolve. In examining the maternal and zygotic transcriptomes from different populations of a single species, *Drosophila melanogaster,* I determined the extent of natural variation in critical early stages of development. Additionally, in examining how mRNA localization changes between species of the same genus, I determined how gains and losses of subcellular localization may evolve on relatively small timescales.

In chapter 1, I showed that variation in maternal and zygotic transcriptomes correlates with the underlying genetic diversity of two populations, African and non-African, of *Drosophila melanogaster*. Consistent with what is known about these mRNA complements between species, I found more variation at stage 5, post zygotic activation, within and between populations than at stage 2, where all transcripts are maternally provided. I found more expression variation within populations than fixed expression differences between them. However, I found that expression variation between species is higher than between lines of the same species at both stages. Interestingly, I found an enrichment of differentially expressed genes between populations at stage 2 on the X chromosome, suggesting that the maternal genome may be under unique selective constraints. Additionally, several genes with known selection signatures consistent with out-of-African expansion of *D. melanogaster* are differentially expressed between populations. I also observed a case of a pseudogenized gene which has previous uncharacterized maternal expression in the Zambia population. I identified a population-specific deletion in this gene in the Raleigh population, exemplifying both DNA loss and expression loss of a gene over time within a species.

In chapter 2, I demonstrated that changes in maternal mRNA localization within the oocytes of species of the genus Drosophila. I did this by bisecting oocytes at a time

point in oogenesis after which mRNA localization in oogenesis has occurred; I chose five species of Drosophila with varying divergences times (~5-50 million years). I sequenced transcripts from the anterior and posterior halves of the oocytes to find transcripts with anterior or posterior enrichment. I found cases of apparent gains in both maternal expression and localization. I hypothesize that simultaneous gains in both expression and localization are maternal expression gains of genes which have pre-existing cis-regulatory elements that can associate with transport machinery. I also found transcripts that are maternally deposited in all five species, but are only localized in a subset. I hypothesized that these differences are due to changes in cis-regulatory elements within the transcript, which confer a gain in localization. However, the small overall number of localization differences suggests a strong conservation of localization and potential mechanistic barriers to the evolution of novel localization.