Evolution of gene expression regulation at the transcriptional and post-transcriptional levels in the early *Drosophila* embryo

By

EMILY CARTWRIGHT

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

Joanna Chiu

Artyom Kopp

Committee in Charge

2022

ACKNOWLEDGEMENTS

I first want to thank my advisor, Susan Lott, for her unwavering support throughout my time as a graduate student. I am so grateful that she accepted me into her lab and has given me so much scientific and professional guidance over the years. I am especially thankful for having her inspire me to push myself outside of my comfort zone when it comes to public speaking. Anytime there was an opportunity to present my work or gain experience speaking in front of an audience, I not only had her encouragement, but she also would dedicate time to making sure I felt prepared for any presentation. I cannot thank her enough for all I have learned and for being so supportive.

Thank you to my committee members, Artyom Kopp and Joanna Chiu for feedback during all stages of my graduate career. Their questions and suggestions have helped me immensely as I figured out what direction to take my analysis.

Thank you to all the people who have been a part of the Lott lab during my time there. I could not ask to work with a better group of people. Thank you especially to Anna Feitzinger, Charlie Omura, and Gizem Kalay for making time in the lab so enjoyable and for always offering advice and feedback. Thank you to Joel Atallah for always being willing to help with genomic analysis. And thank you to the wonderful undergrads that I have had the opportunity to work with – Amanda Crofton, Nikita Shah, and Ruilin Yu.

Thank you to Chris Brooks, who always believes in me and supports me, even at my worst -- especially as I was trying to finish up my PhD. I promise to return the favor of making you (vegetarian) dinners when you write up your dissertation soon. I also want to thank my mom, Lisa, and dad, David, and my siblings, Max and Lily, for their

support and for going out of their way to make sure we always have time to get together, even when I felt the weight of getting work done. I would also not have made it through nearly as happily without my dog/current officemate, Leeloo, who has made it her mission to make sure we get up and go for walks throughout the day while I work from home. She made working at home so much more bearable during the pandemic. And thank you to Caitlin Scott – I'm so glad that running in Davis led us to be friends. Getting outside for adventures with you on the weekends is always a highlight and helped me along in stressful times.

ABSTRACT

The precise regulation of gene expression is especially important during the earliest stages of development, when critical processes lay the foundation for the rest of development. At this time, embryogenesis is dependent on the transcriptional products of two genomes, from the mother and the zygote. Maternally deposited transcripts are supplied to the oocyte during oogenesis and carry out all initial developmental processes from the beginning of embryogenesis until the zygotic genome is activated. Over time, new transcripts are produced from the zygote as maternally deposited transcripts are degraded. The coordinated handoff of developmental control from the maternal to the zygotic genome is tightly regulated and highly conserved. In my dissertation, I investigate two aspects of gene regulation during early development and ask how they compare and evolve across species of *Drosophila*. In Chapter 1, I looked at how gene regulation evolves for differentially maternally deposited and zygotically transcribed genes by using hybrid crosses between three closely related species, *D. simulans, D. sechellia,* and *D. mauritania*. Surprisingly, the mechanisms of gene regulatory change differed substantially between the maternal and zygotic genome. There were more differences in maternal deposition resulting from changes in *trans* regulation while differences in zygotic transcription resulted from a combination of *cis*, *trans*, and the joint action of *cis* and *trans* changes. This is indicative that the maternal and zygotic genomes are under different sets of regulatory constraints and likely evolve via different mechanisms. Another critical aspect of gene regulation that I address in my dissertation is the trajectory of maternal transcript degradation throughout development, until all maternal transcripts are degraded, and how it compares across species of

Drosophila. We chose four species of *Drosophila*, *D. melanogaster*, *D. persimilis*, *D. virilis*, and *D. yakuba*, which vary in developmental time and geographic origin and represent a range of divergence times. Looking at transcripts that are maternally transcribed and not later transcribed by the zygote (maternal-only), we found that a similar proportion degrades by the end of stage 5 in each of the species examined. This suggests that maternal transcripts are stable for a longer absolute amount of time in species that develop more slowly. We also find that relatively few maternal-only transcripts are common across all species examined while a larger proportion are unique to a specific species, indicating a lack of conservation of maternal-only genes, which is especially surprising given the high degree of conservation of maternal genes overall. Future work will investigate whether these transcripts, particularly those that degrade later in development than have previously been examined, contain motifs that may act as signatures for degradation and look at how those compare across species. Overall, the work in this dissertation gives a better understanding of transcriptional and post-transcriptional gene regulation during early embryogenesis and how it evolves across species of *Drosophila*.

TABLE OF CONTENTS

INTRODUCTION

In all animals, gene regulation during early embryogenesis establishes the trajectory for development, directing all critical processes during this time including setting up the body plan and allowing for cell and tissue differentiation. The beginning of development is characterized by a period of time when only maternally supplied proteins and transcripts are present to drive developmental processes (Tadros and Lipshitz 2009; Vastenhouw et al. 2019). During this time, transcription from the zygotic genome has not yet begun (Ali-Murthy et al. 2013; Kwasnieski et al. 2019) so the supply of maternal factors is critical to start embryogenesis. Over time, maternal transcripts are degraded and the control of development is handed off to the zygotic genome, as it becomes transcriptionally active (Tadros and Lipshitz 2009; Vastenhouw et al. 2019). The coordination between the two genomes, that of the mother and that of the zygote, is highly regulated during this process, which is termed the maternal to zygotic transition (MZT).

Prior work has investigated the MZT extensively in many model systems and, while all animals depend on maternal factors to carry out initial development before the zygotic genome is active, the MZT plays out slightly differently in each system. From mice to fruit flies, the length of time that zygotes are dependent solely on maternal factors differs, both in the absolute amount of time and the proportion of embryonic development this stage represents (Tadros and Lipshitz 2009; Vastenhouw et al. 2019). In mice, the major wave of zygotic transcription is initiated at the two-cell stage (Hamatani et al. 2004), and the MZT lasts for over a day (Vastenhouw et al. 2019). In contrast, widescale zygotic genome activation (ZGA) in flies occurs during nuclear cycle

14 (De Renzis et al. 2007), and only spans a matter of hours (Vastenhouw et al. 2019). Thus, while the MZT is tightly regulated and conserved, it does evolve.

Gene regulation is controlled by two genomes during the MZT, the genome of the mother and the zygote, both of which are under different regulatory constraints. In *Drosophila*, maternal transcripts are produced by polyploid nurse cells during oogenesis and are dumped into the oocyte or transported into the oocyte by microtubule-dependent mechanisms (Mische et al. 2007; Kugler and Lasko 2009). The nurse cells transcribe a large quantity of RNA as we find ~100ng of total RNA in embryos prior to zygotic gene expression. The maternal transcriptome also represents a large proportion of all genes and is produced in a short amount of time (Tadros et al. 2007; De Renzis et al. 2007; Thomsen et al. 2010; Lott et al. 2011; Vastenhouw et al. 2019). Surprisingly little is understood about how transcription in the nurse cells is regulated. In contrast, transcription from the zygotic genome is highly regulated as gene expression often occurs in localized patterns throughout the embryo at distinct development times (Small et al. 1992; Perry et al. 2011). Because of the often precise nature of zygotic transcription, it has served as a model system for studying gene regulation in a complex eukaryotic system. Thus, while maternal and zygotic transcripts are in the embryo at the same or similar times, and coordinate the processes of early development, they originate from vastly different gene regulatory environments.

In addition to transcriptional regulation, regulation at post-transcriptional levels plays an important role during early development. Post-transcriptional regulation appears to have an outsized role in the regulation of maternal transcripts (Tadros et al. 2007; Rouget et al. 2010; Barckmann and Simonelig 2013; Eichhorn et al. 2016),

presumably as no new transcription is available in the stage controlled solely by maternal transcripts. Maternal transcripts are subject to differential splicing (Atallah and Lott 2018), stabilization through cytoplasmic poly(A) polymerases (Benoit et al. 2008), localization within the oocyte (Theurkauf and Hazelrigg 1998), and are targets of specific and targeted degradation (Tadros et al. 2007; Laver et al. 2015b).

Here, I will focus primarily on degradation of maternal transcripts. While maternally deposited transcripts and proteins drive all initial developmental processes, maternal mRNA is degraded as developmental control is handed off to the zygotic genome (Tadros and Lipshitz 2009; Vastenhouw et al. 2019). Several regulators of maternal transcript degradation have been well characterized, including Smaug (Tadros et al. 2007; Benoit et al. 2009), Brain Tumor (Laver et al. 2015b), and miRNAs of the miR-309 cluster in *Drosophila* (Bushati et al. 2008) but these RNA binding proteins and miRNAs only account for the degradation of a specific subset of maternal transcripts (Laver et al. 2015a).

Degradation of maternal transcripts is often categorized by when transcripts degrade. Early degradation occurs through machinery supplied by the maternal genome and later degradation is carried out by machinery that is produced from the zygotic genome (Laver et al. 2015a). Smaug acts during the first wave of maternal transcript degradation (Tadros et al. 2007) while Brain Tumor acts during the early and late waves of degradation (Laver et al. 2015b). The miR-309 cluster acts during the late wave of degradation and relies on the activity of Smaug for proper transcription (Benoit et al. 2009; Luo et al. 2016). Smaug, along with cofactors, acts to translationally repress and degrade a large proportion of maternally deposited transcripts (Semotok et al. 2005;

Tadros et al. 2007; Chen et al. 2014). It is estimated that Smaug acts to degrade about two-thirds of the maternal transcripts that are degraded by maternal machinery (Tadros et al. 2007). An additional ~600 transcripts, largely non-overlapping with Smaug associated transcripts, are targeted for degradation by Brain Tumor (Laver et al. 2015b). Pumilio, another RNA binding protein known to act during the later wave of degradation, is thought to associate with maternal transcripts that are degraded (Thomsen et al. 2010; Laver et al. 2015b). Pumilio interacting transcripts largely do not overlap with those associated with Brain Tumor, but it is thought that both may act cooperatively to target maternal transcripts for degradation (Laver et al. 2015b). RNA binding proteins can act through deadenylation of target transcripts where the poly(A) tail is shortened, triggering the degradation of transcripts (Semotok et al. 2005). The miR-309 cluster is thought to target several hundred transcripts for degradation (Bushati et al. 2008), some of which are also predicted to be targets of Brain Tumor (Luo et al. 2016). The mechanism of degradation for all maternal transcripts is not known although there is evidence for additional miRNAs that may bind to maternally deposited transcripts (Thomsen et al. 2010; Luo et al. 2016).

The early and late waves of maternal transcript degradation that have been well-characterized in *D. melanogaster* represent but a small portion of the time maternal transcripts are present (Tadros et al. 2007; De Renzis et al. 2007; Thomsen et al. 2010; Lott et al. 2014). These studies have focused on portions of the first 3-4 hours of development time, up until gastrulation, with most work on the period of time just prior to and just following the activation of the zygotic genome. However, maternal transcripts are likely present up to 5-6 hours of development time in this species. Thus, while the

targets of specific degradation machinery at the MZT have been extensively studied in *D. melanogaster*, how and when the group of late-decaying transcripts is removed is not well understood.

Prior work has found that maternal and zygotic transcriptomes, measured by transcript levels in the early embryo, are highly conserved both within and between species (Atallah and Lott 2018; Feitzinger et al. 2021). The maternal transcriptome appears to be under a high degree of constraint (Atallah and Lott 2018) and more of the maternal genome may be under purifying selection (Feitzinger et al. 2021). However, despite this high degree of conservation, changes in maternal and zygotic transcripts and their abundances have been identified within and between species (Atallah and Lott 2018; Feitzinger et al. 2021). The regulatory mechanisms by which these differences in maternal and zygotic genomes evolve is not well understood. My first chapter of this dissertation addresses the changes in regulation that may underlie the changes in maternal and zygotic transcriptomes in the early embryo. Given the particular importance of post-transcriptional regulation at the maternal stage, it is striking that little is known about how these post-transcriptional mechanisms and their targets evolve. My second chapter focuses on the evolution of maternal transcript degradation, how this process varies across species, and extends this examination of degradation across the entire portion of development where maternal transcripts are present.

The first chapter of my dissertation focuses on how gene regulation can evolve for both maternal and zygotic transcription. Changes in *cis* regulation, such as changes in regulatory sequences like enhancers, and changes in *trans* regulation, such as changes to transcription factors that can bind to regulatory sequences to affect

transcription, are two ways in which gene expression can be altered. While changes in *cis* are likely to only affect the gene where the regulatory change has occurred, changes in *trans* can be pleiotropic and can affect the expression of many genes. Considerable prior research has focused on whether gene expression evolves primarily through changes in *cis*, *trans*, or a combination of *cis* and *trans* regulatory elements (Wittkopp et al. 2004; Graze et al. 2009; McManus et al. 2010; Coolon et al. 2014; Mack et al. 2016; Glaser-Schmitt et al. 2018). Some studies within or between species find more changes in *cis* regulatory elements underlying differences in transcript abundance (Graze et al. 2009; Mack et al. 2016) while others find more changes in *trans* regulatory elements (McManus et al. 2010; Glaser-Schmitt et al. 2018), likely due to the underlying study system or the lines or species compared. In this chapter, I compare regulatory changes underlying differences in maternal transcript deposition and zygotic gene transcription across closely related species of *Drosophila*. By using hybrid crosses between three sister species (*D. mauritiana*, *D. simulans*, and *D. sechellia*), I investigated the regulatory changes underlying evolved differences in gene expression between these species, at two developmental timepoints. The first timepoint sampled was prior to any zygotic gene expression which allowed for studying only the changes in maternal deposition between species. The second timepoint sampled was after ZGA to examine regulatory changes underlying differences in zygotic transcription between species. From the resulting transcriptomic data, species-specific transcript levels were compared between hybrid and parental embryos. This allowed me to investigate regulation in the context of the vastly different regulatory environments of the two developmental stages and ask how gene regulation can evolve between closely related species.

Through this study we found that maternal genes are more likely to change in *trans* regulation while in comparison, zygotic gene regulation for primarily zygotic genes evolves through a combination of *cis*, *trans*, and the dual action of *cis* and *trans* regulatory changes. Through this work and others (Omura and Lott 2020), we have shown that the maternal transcriptome is likely largely regulated at the level of chromatin state. Here, I show that maternal genes were found to be enriched with binding motifs associated with *trans* factors that interact with topologically associated domains and are thought to act as insulators. In contrast, the regulatory changes found for primarily zygotic genes, changes in *cis*, *trans*, and a combination of the two, are reflective of what is known about zygotic gene regulation. The findings in Chapter 1 indicate that maternal and zygotic transcripts are produced in different regulatory environments and can evolve through different mechanisms.

Chapter 2 of my dissertation addresses when maternal transcripts are degraded by tracking degradation across seven developmental stages from a time when only maternal transcripts are present, through ZGA, to a time when we expect all maternal transcripts to be degraded. While a lot of prior work has been done in *D. melanogaster*, the work in this chapter extends studying maternal transcript degradation to other species of *Drosophila*. Studies show that both maternal and zygotic transcripts change quantitatively and qualitatively across species of *Drosophila* (Atallah and Lott 2018). If species evolve maternal deposition of transcripts, they will likely also need to evolve signals for them to degrade during the MZT. The work described in Chapter 2 addresses how maternal degradation evolves across species, focusing on maternal transcripts that likely do not have zygotic expression until after the maternal component

is entirely degraded. We refer to this set of transcripts as maternal-only. The species examined differ in developmental time, divergence time, and have different geographic and climate origins. In this chapter, we ask how maternal transcript degradation compares across species to better understand how this aspect of regulation is conserved or changes over time. We find that among maternal-only genes, many are not shared among all four species and in fact, many are uniquely maternal-only in an individual species. Additionally, maternal-only genes have approximately the same proportion of genes that degrade by stage 5 across all species. Future analyses of these data will examine all maternal transcripts in these species, including ones that are zygotically expressed before the maternal transcripts are completely degraded.

The work in this dissertation addresses how gene regulation can change at the levels of maternal transcript deposition, zygotic transcription, and maternal transcript degradation across species of *Drosophila*. It gives a better understanding of the constraints placed on two distinct developmental stages during early embryogenesis and reveals insights into the gene regulatory environments at each stage. It also explores how developmental timing affects the trajectory of mRNA decay in early embryos of different species in order to better understand how this critical post-transcriptional process affects the evolution of the maternal transcriptome.

REFERENCES

Ali-Murthy Z., S. E. Lott, M. B. Eisen, and T. B. Kornberg, 2013 An essential role for zygotic expression in the pre-cellular Drosophila embryo. PLoS genetics 9: e1003428. https://doi.org/10.1371/journal.pgen.1003428

Atallah J., and S. E. Lott, 2018 Evolution of maternal and zygotic mRNA complements in the early Drosophila embryo, (K. A. Dyer, Ed.). PLoS genetics 14: e1007838. https://doi.org/10.1371/journal.pgen.1007838

Barckmann B., and M. Simonelig, 2013 Control of maternal mRNA stability in germ cells and early embryos. Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms 1829: 714–724. https://doi.org/10.1016/j.bbagrm.2012.12.011

Benoit P., C. Papin, J. E. Kwak, M. Wickens, and M. Simonelig, 2008 PAP- and GLD-2-type poly(A) polymerases are required sequentially in cytoplasmic polyadenylation and oogenesis in Drosophila. Development 135: 1969–1979. https://doi.org/10.1242/dev.021444

Benoit B., C. H. He, F. Zhang, S. M. Votruba, W. Tadros, *et al.*, 2009 An essential role for the RNA-binding protein Smaug during the Drosophila maternal-to-zygotic transition. Development 136: 923–932. https://doi.org/10.1242/dev.031815

Bushati N., A. Stark, J. Brennecke, and S. M. Cohen, 2008 Temporal reciprocity of miRNAs and their targets during the maternal-to-zygotic transition in Drosophila. Current Biology 18: 501–506. https://doi.org/10.1016/j.cub.2008.02.081

Chen L., J. G. Dumelie, X. Li, M. H. Cheng, Z. Yang, et al., 2014 Global regulation of mRNA translation and stability in the early Drosophilaembryo by the Smaug RNA-binding protein. Genome Biol 15: R4. https://doi.org/10.1186/gb-2014-15-1-r4

Coolon J. D., C. J. McManus, K. R. Stevenson, B. R. Graveley, and P. J. Wittkopp, 2014 Tempo and mode of regulatory evolution in Drosophila. Genome Research 24: 797–808. https://doi.org/10.1101/gr.163014.113

De Renzis S. D., O. Elemento, S. Tavazoie, and E. F. Wieschaus, 2007 Unmasking Activation of the Zygotic Genome Using Chromosomal Deletions in the Drosophila Embryo. PLOS Biology 5: e117. https://doi.org/10.1371/journal.pbio.0050117

Eichhorn S. W., A. O. Subtelny, I. Kronja, and J. K. Elife, 2016 mRNA poly (A)-tail changes specified by deadenylation broadly reshape translation in Drosophila oocytes and early embryos. cdn.elifesciences.org. https://doi.org/10.7554/elife.16955.001

Feitzinger A. A., A. Le, A. Thompson, M. Haseeb, M. K. Murugesan, et al., 2021 Natural variation in the maternal and zygotic mRNA complements of the early embryo in Drosophila melanogaster. Biorxiv 2021.11.01.466760. https://doi.org/10.1101/2021.11.01.466760

Glaser-Schmitt A., A. Zečić, and J. Parsch, 2018 Gene Regulatory Variation in Drosophila melanogaster Renal Tissue. Genetics genetics.301073.2018-46. https://doi.org/10.1534/genetics.118.301073

Graze R. M., L. M. McIntyre, B. J. Main, M. L. Wayne, and S. V. Nuzhdin, 2009 Regulatory Divergence in Drosophila melanogaster and D. simulans, a Genomewide Analysis of Allele-Specific Expression. Genetics 183: 547–561. https://doi.org/10.1534/genetics.109.105957

Hamatani T., M. G. Carter, A. A. Sharov, and M. S. H. Ko, 2004 Dynamics of Global Gene Expression Changes during Mouse Preimplantation Development. Dev Cell 6: 117–131. https://doi.org/10.1016/s1534-5807(03)00373-3

Kugler J.-M., and P. Lasko, 2009 Localization, anchoring and translational control of oskar, gurken, bicoid and nanos mRNA during Drosophila oogenesis. Fly 3: 15–28. https://doi.org/10.4161/fly.3.1.7751

Kwasnieski J. C., T. L. Orr-Weaver, and D. P. Bartel, 2019 Early genome activation in Drosophila is extensive with an initial tendency for aborted transcripts and retained introns. Genome Res 29: 1188–1197. https://doi.org/10.1101/gr.242164.118

Laver J. D., A. J. Marsolais, C. A. Smibert, and H. D. Lipshitz, 2015a *Regulation and Function of Maternal Gene Products During the Maternal-to-Zygotic Transition in Drosophila*. Elsevier Inc.

Laver J. D., X. Li, D. Ray, K. B. Cook, N. A. Hahn, et al., 2015b Brain tumor is a sequence-specific RNA-binding protein that directs maternal mRNA clearance during the Drosophila maternal-to-zygotic transition. Genome biology 16: 94. https://doi.org/10.1186/s13059-015-0659-4

Lott S. E., J. E. Villalta, G. P. Schroth, S. Luo, L. A. Tonkin, et al., 2011 Noncanonical Compensation of Zygotic X Transcription in Early Drosophila melanogaster Development Revealed through Single-Embryo RNA-Seq. PLOS Biology 9: e1000590. https://doi.org/10.1371/journal.pbio.1000590

Lott S. E., J. E. Villalta, Q. Zhou, D. Bachtrog, and M. B. Eisen, 2014 Sex-Specific Embryonic Gene Expression in Species with Newly Evolved Sex Chromosomes, (H. S. Malik, Ed.). PLoS Genetics 10: e1004159. https://doi.org/10.1371/journal.pgen.1004159

Luo H., X. Li, J. M. Claycomb, and H. D. Lipshitz, 2016 The Smaug RNA-Binding Protein Is Essential for microRNA Synthesis During the Drosophila Maternal-to-zygotic Transition. G3: Genes|Genomes|Genetics 6: 3541–3551. https://doi.org/10.1534/g3.116.034199

Mack K. L., P. Campbell, and M. W. Nachman, 2016 Gene regulation and speciation in house mice. Genome Research gr.195743.115. https://doi.org/10.1101/gr.195743.115

McManus C. J., J. D. Coolon, M. O. Duff, J. Eipper-Mains, B. R. Graveley, et al., 2010 Regulatory divergence in Drosophila revealed by mRNA-seq. Genome Research 20: 816–825. https://doi.org/10.1101/gr.102491.109

Mische S., M. Li, M. Serr, and T. S. Hays, 2007 Direct Observation of Regulated Ribonucleoprotein Transport Across the Nurse Cell/Oocyte Boundary, (S. Wente, Ed.). Molecular Biology of the Cell 18: 2254–2263. https://doi.org/10.1091/mbc.e06-10-0959

Omura C. S., and S. E. Lott, 2020 The conserved regulatory basis of mRNA contributions to the early Drosophila embryo differs between the maternal and zygotic genomes. Plos Genet 16: e1008645. https://doi.org/10.1371/journal.pgen.1008645

Perry M. W., A. N. Boettiger, and M. Levine, 2011 Multiple enhancers ensure precision of gap gene-expression patterns in the Drosophila embryo. Proceedings of the National Academy of Sciences 108: 13570–13575. https://doi.org/10.1073/pnas.1109873108

Rouget C., C. Papin, A. Boureux, A.-C. Meunier, B. Franco, et al., 2010 Maternal mRNA deadenylation and decay by the piRNA pathway in the early Drosophila embryo. Nature 467: 1128-U144. https://doi.org/10.1038/nature09465

Semotok J. L., R. L. Cooperstock, B. D. Pinder, H. K. Vari, H. D. Lipshitz, et al., 2005 Smaug Recruits the CCR4/POP2/NOT Deadenylase Complex to Trigger Maternal Transcript Localization in the Early Drosophila Embryo. Curr Biol 15: 284–294. https://doi.org/10.1016/j.cub.2005.01.048

Small S., A. Blair, and M. Levine, 1992 Regulation of Even-Skipped Stripe-2 in the Drosophila Embryo. Embo Journal 11: 4047–4057. https://doi.org/10.1002/j.1460-2075.1992.tb05498.x

Tadros W., A. L. Goldman, T. Babak, F. Menzies, L. Vardy, et al., 2007 SMAUG Is a Major Regulator of Maternal mRNA Destabilization in Drosophila and Its Translation Is Activated by the PAN GU Kinase. Developmental Cell 12: 143–155. https://doi.org/10.1016/j.devcel.2006.10.005

Tadros W., and H. D. Lipshitz, 2009 The maternal-to-zygotic transition: a play in two acts. Development 136: 3033–3042. https://doi.org/10.1242/dev.033183

Theurkauf W. E., and T. I. Hazelrigg, 1998 In vivo analyses of cytoplasmic transport and cytoskeletal organization during Drosophila oogenesis: characterization of a multi-step anterior localization pathway. Dev Camb Engl 125: 3655–66.

Thomsen S., S. Anders, S. C. Janga, W. Huber, and C. R. Alonso, 2010 Genome-wide analysis of mRNA decay patterns during early Drosophila development. Genome biology 11: R93. https://doi.org/10.1186/gb-2010-11-9-r93

Vastenhouw N. L., W. X. Cao, and H. D. Lipshitz, 2019 The maternal-to-zygotic transition revisited. Development 146: dev161471-20. https://doi.org/10.1242/dev.161471

Wittkopp P. J., B. K. Haerum, and A. G. Clark, 2004 Evolutionary changes in cis and trans gene regulation. Nature 430: 85–88. https://doi.org/10.1038/nature02698

CHAPTER 1:

Evolved differences in *cis* and *trans* regulation between the maternal and zygotic mRNA complements in the *Drosophila* embryo

Emily L. Cartwright and Susan E. Lott

The work in this chapter is published as:

Cartwright, E. L. & Lott, S. E. Evolved Differences in cis and trans Regulation Between the Maternal and Zygotic mRNA Complements in the Drosophila Embryo. *Genetics* **216**, genetics.303626.2020 (2020).

Supplemental material is available on figshare:
https://doi.org/10.25386/genetics.12935120

ABSTRACT

How gene expression can evolve depends on the mechanisms driving gene expression. Gene expression is controlled in different ways in different developmental stages; here we ask whether different developmental stages show different patterns of regulatory evolution. To explore the mode of regulatory evolution, we used the early stages of embryonic development controlled by two different genomes, that of the mother and that of the zygote. During embryogenesis in all animals, initial developmental processes are driven entirely by maternally provided gene products deposited into the oocyte. The zygotic genome is activated later, when developmental control is handed off from maternal gene products to the zygote during the maternal-to-zygotic transition. Using hybrid crosses between sister species of *Drosophila* (*D.* simulans, *D. sechellia*, and *D. mauritiana*) and transcriptomics, we find that the regulation of maternal transcript deposition and zygotic transcription evolve through different mechanisms. We find that patterns of transcript level inheritance in hybrids, relative to parental species, differ between maternal and zygotic transcripts, and maternal transcript levels are more likely to be conserved. Changes in transcript levels occur predominantly through differences in *trans* regulation for maternal genes, while changes in zygotic transcription occur through a combination of both *cis* and *trans* regulatory changes. Differences in the underlying regulatory landscape in the mother and the zygote are likely the primary determinants for how maternal and zygotic transcripts evolve.

INTRODUCTION

Since the proposal that regulation of gene expression may be one of the primary drivers of morphological evolution was introduced (King and Wilson 1975; Carroll 1995), much research has been directed toward elucidating mechanisms underlying the evolution of gene expression. The mechanistic basis of regulatory control determines the substrate for evolution of gene expression. At the transcriptional level, gene expression is controlled by *cis*-regulatory elements (regulatory DNA sequences, such as enhancers or promoters) and by *trans* acting factors (such as transcription factors or miRNAs that bind the regulatory DNA of many genes). Due to their modular structure (Britten and Davidson 1969; Davidson and Peter 2015), changes in *cis*-regulatory elements can lead to the evolution of altered gene expression and thus new traits (Prud'homme *et al.* 2006; Chan *et al.* 2010; Kvon *et al.* 2016) without the pleiotropic consequences for other critical traits controlled by the same gene. On the other hand, transcription factors that bind to enhancers often have higher pleiotropy and thus can affect the expression of many genes. This poses a fundamental mechanistic question as to how gene regulatory evolution occurs: whether changes are more likely to occur in *cis*-regulatory elements or *trans*-acting factors.

In order to determine the relative contributions of *cis* and *trans* elements to the evolution of gene expression genome-wide, previous studies implemented the use of genetic hybrids and methods of detecting allele-specific expression (Wittkopp *et al.* 2004; Landry *et al.* 2005; Graze *et al.* 2009; McManus *et al.* 2010; Coolon *et al.* 2014; León-Novelo *et al.* 2014). Several studies point to differences in *cis* regulation as the primary mechanism of change in transcript abundance within or between species

(Graze *et al.* 2009; Mack *et al.* 2016), while other studies indicate that *trans* changes are more widespread (McManus *et al.* 2010; Glaser-Schmitt *et al.* 2018). Previous research has proposed that the difference in the abundance of *cis* versus *trans* changes affecting gene expression can be explained by the divergence times of the strains or species being compared (Coolon *et al.* 2014), or the particular tissue type examined (Buchberger *et al.* 2019). Given this body of previous work, it is surprising that, to our knowledge, no previous study has compared *cis* and *trans* contributions to gene expression evolution across developmental stages in a model organism.

In this study, we ask whether contributions of *cis* and *trans* changes to gene regulatory evolution differ across developmental stages. We chose stages during embryogenesis that are close together in developmental time, yet are likely to broadly differ in the mechanistic basis of regulation. The general regulatory architecture of a particular stage likely affects how regulation at this stage can evolve. The early stages of development utilized here are under regulatory control of entirely different genomes: that of the mother, and that of the zygote. The earliest developmental processes in embryogenesis are regulated by maternally provided RNA and protein, which lay the foundation for the rest of development (Tadros and Lipshitz 2009; Vastenhouw *et al.* 2019). These maternally derived gene products carry out all initial developmental events because at the time of fertilization, the zygotic genome is transcriptionally silent. Because the zygotic genome is not yet transcriptionally active, post-transcriptional mechanisms also play an important role in regulating the amount of maternal gene products present (Tadros *et al.* 2007; Rouget *et al.* 2010; Barckmann and Simonelig 2013). As the zygotic genome is activated, control of developmental processes is

handed off from the maternally deposited factors to those derived from the zygotic genome in a process known as the maternal-to-zygotic transition (MZT). The MZT is a highly conserved and regulated process during early development that occurs in all animals and in some species of flowering plants (Baroux *et al.* 2008; Tadros and Lipshitz 2009; Vastenhouw *et al.* 2019).

Regulation of early zygotic gene expression has been extensively studied over many years (Mannervik 2014; Schulz and Harrison 2019), while regulatory control of maternal genes is not well understood. Studies of maternal transcripts have largely focused on their transport into the oocyte (Mische *et al.* 2007; Kugler and Lasko 2009), their localization and movement within the oocyte and embryo (Theurkauf and Hazelrigg 1998; Kugler and Lasko 2009), activation of their translation (Salles *et al.* 1994), and their degradation (Tadros *et al.* 2007; Bushati *et al.* 2008; Laver *et al.* 2015). Zygotic gene expression is precisely regulated, with classic examples such as the *even-skipped* gene having multiple enhancer elements along with multiple transcription factors responsible for producing complex expression patterns in developmental time and embryonic space (Small *et al.* 1992; Perry *et al.* 2011; Mannervik 2014). In contrast, maternal transcripts are produced by support cells called nurse cells during oogenesis, which are polyploid and highly transcriptionally active (Kugler and Lasko 2009; Lasko 2012), rapidly producing large amounts of transcripts. Roughly 50-75% of the genome is maternally deposited (Tadros *et al.* 2007; De Renzis *et al.* 2007; Thomsen *et al.* 2010; Lott *et al.* 2011; Vastenhouw *et al.* 2019) and there is considerable post-transcriptional regulation of maternal factors (Tadros *et al.* 2007; Bushati *et al.* 2008; Laver *et al.* 2015). For these reasons, regulation at the transcript level may not need to be as

precise. While the mechanisms behind maternal transcription are not well understood, the regulatory environments driving the production of the maternal and zygotic transcriptomes are quite different.

To study the regulatory basis of differences in transcript levels at the maternal and zygotic stages of early development between species, we focused on three closely related species of *Drosophila* (*D. simulans*, *D. sechellia*, and *D. mauritiana*). Despite having a relatively close divergence time of 250,000 years (McDermott and Kliman 2008), these sister species have differences in the pools of transcripts present in the developing embryo both at a stage where only maternal transcripts are present, and at a stage after zygotic genome activation (ZGA; Atallah and Lott 2018). By comparing hybrids and parental lines of the species *D. simulans*, *D. sechellia*, and *D. mauritiana*, we asked whether, at each of these two developmental stages, changes in gene expression between species occurred due to changes in *cis*, in *trans*, or in a combination of the two.

We found that patterns of gene regulatory changes between species are distinct across developmental stages for maternally deposited transcripts and for genes with primarily zygotic expression (see Methods) in early embryogenesis. Differences in maternal transcripts occur much more frequently due to *trans* as opposed to *cis* regulatory changes, while differences in zygotic gene transcription occur through a mix of *cis*, *trans*, and the combined action of *cis* and *trans* regulatory changes. The complex pattern of changes found in our study at the zygotic stage speaks to what has been known about regulation at this stage for some time, that both *cis* and *trans* elements are necessary for the intricate control of gene expression in time and space at this stage of

embryogenesis. The large proportion of *trans* regulatory signal found at the maternal stage may reveal fundamental properties of the regulatory architecture during oogenesis. *Trans* regulators can affect a large number of genes at the same time, as might be necessary to maximize mRNA production to load sufficient numbers of transcripts into the egg. We also identified motifs associated with *trans* regulation at the level of chromatin at the maternal stage, which lends evidence to an emerging picture of how gene expression might be regulated genome-wide during oogenesis. Overall, we find distinct patterns of gene regulatory changes at the two embryonic timepoints, before and after ZGA, indicating evolved changes in gene regulation differ based on the developmental context.

MATERIALS AND METHODS

Crossing scheme and sample acquisition

Three *Drosophila* species were used for this study: *D. sechellia* (Dsec/wild-type;14021-0248.25) and *D. simulans* (Dsim/w[501]; 14021-0251.011) from the 12 Genomes study (Clark *et al.* 2007) and *D. mauritiana* (Dmau/[w1];14021-0241.60). For interspecific crosses, each vial was set up using 7-12 virgin females from one species and 7-10 males from another. We did not cross *D. sechellia* females and *D. simulans* males, as this combination is known to be incompatible (Lachaise *et al.* 1986). Two types of hybrid crosses were established from which embryos were collected: 1) to determine the regulatory basis of changes in zygotic gene expression, hybrid F1 embryos were collected; and 2) to determine the regulatory basis of changes in maternal gene expression, embryos produced by hybrid F1 mothers were collected. To

investigate the regulatory basis of changes in zygotic gene expression, we collected hybrid F1 embryos at the very end of blastoderm stage, stage 5 (Bownes' stages; Bownes 1975; Campos-Ortega and Hartenstein 2013), a timepoint after the zygotic genome is activated. We define late stage 5 by morphology; it is the point when cellularization is complete, but gastrulation has not yet begun. In order to determine the regulatory basis of changes in maternal gene expression, similar crosses were established with hybrid females from the F1 generations of the initial crosses and males that were the same species as the maternal species in the parental cross. We set up crosses in this manner in order to establish consistency amongst crosses, although the male genotype is unlikely to affect our data. The contribution of sperm mRNA to the zygote is debated but known contributions are small (Fischer *et al.* 2012) and likely not detectable via RNA-sequencing (Ali-Murthy *et al.* 2013). This second set of crosses was used to collect stage 2 embryos (Bownes' stages; Bownes 1975; Campos-Ortega and Hartenstein 2013), during which time only maternal gene products are present. At this point in development, the cytoplasm has retracted from the vitelline membrane at the anterior and posterior poles of the embryo but pole cells have not yet migrated to the posterior (Ashburner 1989). As is conventional in *Drosophila* genetics, we denote our crosses by listing the female genotype first and the male genotype second. For example, in a cross between *D. mauritiana* and *D. simulans*, we write the genotype of the resulting stage 5 hybrid embryo as *mau* x *sim*. We describe the hybrid genotype of stage 2 embryos from backcrosses of hybrid F1 females and males of the maternal species in the initial cross in a similar way, e.g. (*mau* x *sim*) x *mau* (also see Figure 1.1 for cross diagram).

All flies were raised in vials on standard cornmeal media at 25°C. Flies were allowed to lay eggs for ~2 hours (for collecting stage 2 embryos) and ~3 hours (for collecting stage 5 embryos) before they were transferred to a new vial so that the eggs could be harvested. Eggs were collected from 4-14 day old females, dechorionated using 50% bleach and moved into halocarbon oil on a microscope slide for staging. Embryos were staged at the appropriate developmental time point under a microscope (Zeiss AxioImager M2), imaged, and promptly collected at stage 2 or at the end of stage 5 (Bownes' stages; Bownes 1975; Campos-Ortega and Hartenstein 2013), of embryonic development.

To collect the samples after staging, the embryos were quickly transferred with a paintbrush to Parafilm (Bemis) and rolled (to remove excess halocarbon oil) into a drop of TRIzol (Ambion). The embryos were ruptured with a needle so that the contents dissolved in the TRIzol and were transferred to a tube to be frozen at -80°C until extraction. RNA was extracted using glycogen as a carrier (as per manufacturer instructions) in a total volume of 1mL TRIzol. Approximately 80-120ng total RNA was extracted from individual embryos, measured using a Qubit 2.0 fluorimeter (Invitrogen). The quality of the RNA was validated on RNA Pico Chips using an Agilent Bioanalyzer.

Genotyping was performed to determine embryo sex for stage 5 samples, as dosage compensation is not complete and transcript levels for genes on the X chromosome may differ for males and females at this time in development (Lott *et al.* 2014). DNA was extracted from each sample along with the RNA as per manufacturer instructions and amplified using a whole genome amplification kit (illustra GenomePhi v2, GE Healthcare). Sex-specific primers (Table S1.1) designed for use with all three

species, two sets for the Y chromosome (to the *ORY* and *kl2* genes) and one control set (to *ftz*), were used to genotype the single embryos after genome amplification. For the stage 5 samples, a total of three male and three female embryos from each cross were used for sequencing. One noted exception is that in the *sim* x *mau* cross, a total of four female and two male embryos were collected, as determined from the transcriptomic data. A total of three stage 2 embryos were collected. We did not perform genotyping for embryo sex on the stage 2 embryos because the zygotic genome is not yet active at this stage in development.

Library preparation and sequencing

The RNA from single embryos was treated with DNase (TurboDNA-free, Life AM1907) using manufacturer instructions and RNA sequencing libraries were constructed with Illumina TruSeq v2 kits following the manufacturer's low sample protocol. The Illumina protocol uses oligo (dT) beads to enrich for polyadenylated transcripts. Because poly(A) tail length is important in many post-transcriptional processes during early development, including translational efficiency, it is important to ensure that the method used for mRNA selection does not produce a biased set of poly(A) tail lengths. Previous datasets report poly(A) length distributions for transcripts during oogenesis and early development (Lim *et al.* 2016; Eichhorn *et al.* 2016). We could not directly compare our data to previous reports, as these studies were done using *D. melanogaster*, which may have a different poly(A) tail length distribution than the species used in our analysis. However, previous studies comparing distributions of poly(A) tail lengths of all genes to poly(A) tail lengths of transcripts recovered through

poly(A) selection in *D. melanogaster* have demonstrated that poly(A) selection with commonly used methods does not bias which transcripts are recovered from the total pool of transcripts present (Eichhorn *et al.* 2016). This includes studies that used the same single embryo approaches utilized here (Crofton *et al.* 2018; Atallah and Lott 2018). Therefore, it seems unlikely that poly(A) selection heavily biases the extracted RNA relative to the RNAs present at these developmental stages. cDNA libraries were quantified using a Qubit 2.0 fluorimeter (dsDNA BR Assay Kits) and the quality of the libraries were assessed on High Sensitivity DNA chips on an Agilent Bioanalyzer. The libraries were pooled (11-12 samples per lane) and sequenced (100bp, paired-end) in four lanes on an Illumina HiSeq2500 sequencer. Sequencing was done at the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley.

Data Processing

Raw reads were processed to remove adapter sequences and gently trimmed (PHRED Q<5; Macmanes 2014) using Cutadapt (version 1.7.1; Martin 2011). TopHat (version 2.0.13; Trapnell *et al.* 2012) was used to align reads to the *D. simulans* (version r2.02) and *D. sechellia* (version r1.3) genome assemblies (from the twelve species project, downloaded from Flybase) and to the *D. mauritiana* MS17 assembly (Nolte *et al.* 2013). Because the *D. mauritiana* line used for sequencing and the line used to construct the genome assembly differed, variant sites from the lab line, called using Genome Analysis Toolkit's (GATK) Haplotypecaller, were incorporated into the MS17 assembly using Pseudoref (http://yangjl.com/pseudoRef/; Xu *et al.* 2020). Additionally, an updated annotation file for the MS17 assembly (Torres-Oliva *et al.* 2016) was used

during alignment and in subsequent processing steps. Annotation files for *D. simulans* and *D. sechellia* were obtained from the same versions of the genome release of each species. Read alignment, mismatches, edit distance, and gap length were all set to three when using TopHat (version 2.0.13; Trapnell *et al.* 2012) to allow for a higher rate of read alignment.

In order to differentiate reads derived from each parental species, variant sites were called between the genomes of the species used in this analysis. RNA-seq reads from parental species samples (from previous data from Atallah and Lott, 2018, GEO accession GSE112858) were aligned to every other parental genome in each pairwise comparison using TopHat (version 2.0.13; Trapnell *et al.* 2012). The BAM files from the TopHat output were sorted and indexed using Samtools (version 1.2; Li *et al.* 2009). Picard tools (version 2.7.1) and GATK tools (Van der Auwera *et al.* 2013) were then used to identify variant sites by using the following programs: AddorReplaceReadGroups, MarkDuplicates, ReorderSam, SplitNCigarReads, and HaplotypeCaller. Additionally, indels were excluded and sites with single variants selected using the SelectVariants tool. The variants were ordered using a Pysam script (Python version 2.7.10) and read assignments to parental genomes were subsequently organized with custom R scripts using the variant sites that exist between the parental genomes (R version 3.4.1; R Core Team 2017) (Files S1, S2, S3 and S4). This pipeline was also used to update the *D. mauritiana* MS17 assembly (Nolte *et al.* 2013) with variants present in the line we used in the lab (Dmau/[w1];14021-0241.60). Normalization to the upper quantile was performed across all samples in each set of

pairwise species comparisons. This was used to account for differences in the number of reads for each sample as a result of sequencing.

Read counts represent the number of reads mapping to variant sites within a gene. A cutoff of 5 or more reads mapped to any given gene was set to determine if a gene was expressed. Genes with read counts <5 in both species in any pairwise comparison were not considered to be expressed in either species so were removed from the analysis. This cutoff was tested empirically and was set to exclude genes with low count numbers that had a higher frequency of mapping in a biased manner to both parental genomes. Genes analyzed in this analysis were also limited to those with annotated orthologs in both species in any pairwise comparison. An orthologs table from Atallah and Lott, 2018, was updated using the annotations available on FlyBase (v2017) and an updated set of annotations from Torres-Oliva, *et al*. 2016. This revised orthologs table (Table S1.2) was used to compare genes between each species and in each direction of mapping.

Mapping bias due to differing genome quality may occur when using two different reference genomes. In our study, mapping bias can result when a higher proportion of reads from one allele map to the genome of a species used in the cross than reads from the other allele. In order to alleviate mapping bias that may occur when mapping the parental and hybrid samples to each parental reference genome, Poisson Gamma (PG) models (León-Novelo *et al.* 2014) were employed to calculate mapping bias for every set of mappings, in each pairwise comparison of species. We compared the 95% confidence intervals (CIs) from PG models (with fixed bias parameter, $q = 0.5$) in each direction of mapping. We set a slightly conservative standard for classifying allelic

imbalance where genes with CIs below 0.49 or above 0.51 were called differentially expressed. Genes with CIs close to 0.50 did not appear differentially expressed when looking at the count data, so we used a more conservative cutoff. Genes that appeared differentially expressed in one direction of mapping (CIs fell outside of the range of 0.49 - 0.51 when comparing the expression levels of parental alleles in each replicate) but not in the other direction of mapping were removed from the analysis, as this was determined to be a result of mapping bias between the two genomes. We also removed genes that had disparate confidence intervals in the two mapping directions (i.e. one mapping direction yielded a CI that fell above the 0.49 - 0.51 range and the other direction of mapping yielded a CI that fell below the 0.49 - 0.51 range).

We found that between 9.6% and 10.9% of genes expressed at stage 2 (with a count >5) mapped in a biased way to parental genomes when compared to the total number of orthologous genes between any pair of species. Each between-species comparison has a different number of orthologous genes so the proportion of biased genes varies based on the pair of species compared in a cross. In contrast to the maternal stage, we found that between 5.0% and 6.2% of genes expressed at stage 5 mapped in a biased way to parental genomes when compared to the total number of orthologous genes between a pair of species. Overall, when looking at the total proportion of biased genes, not just those that were called “expressed” in our analysis, we found that between 24.8% and 28.3% of genes at stage 2 and between 20.0% and 21.2% of genes at stage 5 mapped in a biased manner when compared to the total number of orthologous genes in any comparison between species in a cross. All the genes that mapped in a biased manner were removed from our analysis. Genes that

were not biased in their mapping and had a read count of >5 reads were retained for analysis.

Genes used for stage 5 analysis

To focus on the gene regulation from the zygotic genome after ZGA, we removed genes with high levels of maternal transcript deposition from our analysis. We limited the pool of genes analyzed to those that are mostly zygotic because roughly half of maternal transcripts are not entirely degraded by stage 5 (although studies are somewhat variable in the percent reported; Tadros *et al.* 2007; De Renzis *et al.* 2007; Thomsen *et al.* 2010; Lott *et al.* 2014) and we wanted to examine only those genes that have a larger contribution to expression from the zygotic genome. For this analysis, we included genes with "zygotic-only" expression (those that are not maternally deposited) and genes that are "mostly zygotic" (those with 8-fold higher expression at stage 5 relative to stage 2, a $\log_2$ difference greater than three). We added a count of 1 to the transcript count for genes at the maternal and zygotic stages when calculating the $\log_2$ difference to avoid errors associated with taking the $\log_2$ of zero. We tested several cutoffs but chose the 8-fold threshold because at this conservative cutoff, most genes with high maternal transcript deposition are removed from the analysis. Additionally, for this analysis we used confidence intervals and averages generated from only female samples for genes on the X chromosome because dosage compensation is not complete at stage 5 (Lott *et al.* 2011).

Correlation analysis and PCA

We performed correlation analysis (Figure 1.2, Table S1.3) between single embryos across replicates, stages, and genotypes in R (R Core Team 2017) using the Spearman option within the *corr* function. Principal component analysis (PCA) was also performed in R using the *prcomp* function (Figure S1.1).

Cis/Trans analysis

To identify evolved regulatory changes between species, we first determined which genes showed differential expression between alleles using the 95% CIs from PG models (León-Novelo *et al.* 2014) that were also used to interpret mapping bias. We compared expression levels of alleles for individual genes within the hybrids and also between parental samples using this model. Genes with CIs that fell outside of the 49% - 51% range were defined as differentially expressed, either within the hybrid or between parental samples, while those falling within the 49% - 51% range were identified as having the same level of expression. Genes were then categorized using *cis*, *trans*, *cis* + *trans*, *cis x trans*, *compensatory*, and *conserved* categories as described in Landry, *et al*. 2005; McManus, *et al*. 2010; and Coolon, *et al*. 2014 (Figures 1.3, S1.2 and S1.3). We assigned the following categories for regulatory change based on the CIs generated from PG models for individual genes (see Figure 1.4 for individual examples):

cis: Genes categorized as having changes in *cis* are those that are differentially expressed (CIs do not overlap 49% - 51%) between the parental species and in the hybrids. (CIs for parental species and hybrids overlap each other for changes

purely in *cis*. To determine this, we used the CIs generated from mapping to the *D. simulans* genome for *D. simulans/D. mauritiana* and *D. simulans/ D. sechellia* comparisons and CIs generated from mapping to the *D. sechellia* genome for *D. sechellia/D. mauritiana* comparisons.)

trans: Genes that are differentially expressed between the parental species (CI does not overlap 49% - 51%) but are not differentially expressed in the hybrid (CI overlaps 49% - 51%).

cis + *trans*: Genes that are differentially expressed in the hybrids and between the parental species (CI does not overlap 0.49% - 0.51%) and the CI is in the same direction for both the parents and the hybrid (i.e. both are greater than 51% but the CIs for the parents and hybrid do not overlap. For this comparison, we used the CIs generated from mapping to the *D. simulans* genome for *D. simulans/D. mauritiana* and *D. simulans/ D. sechellia* comparisons and CIs generated from mapping to the *D. sechellia* genome for *D. sechellia/D. mauritiana* comparisons.)

cis x *trans*: Genes that are differentially expressed in the hybrids and between the parental species (CI does not overlap 49% - 51%) and the CI is in opposite directions for the parents and the hybrid (i.e. one is greater than 51%, the other is less than 49%)

compensatory: Genes that are not differentially expressed between the parental species (CI overlaps 49% - 51%) but are differentially expressed in the hybrids (CI does not overlap 49% - 51%).

conserved: Genes are not differentially expressed between the parental species or within the hybrids (CIs overlap 49% - 51%).

Inheritance Patterns

Previous studies from Gibson, *et al*. 2004 and McManus, *et al*. 2010 identified and outlined ways to classify inheritance patterns of transcript abundance in hybrids in relation to parental samples. We used these methods in our study to compare the averages of total expression levels in the hybrids relative to those of parental samples. Gene expression was considered conserved if the expression level between parental samples and the total expression in the hybrid (sum of the expression of the two species-specific alleles in the hybrid) were within 1.25-fold of one another, a $\log_2$-fold change of 0.32. Overdominant genes were expressed at least 1.25-fold more in the hybrid than in either parent while underdominant genes were expressed at least 1.25-fold lower in the hybrid than in either parent. Genes that were expressed at an intermediate level in the hybrid in comparison to the parental species samples involved in the cross were defined as additive. Dominance was determined when the hybrid had expression within 1.25-fold of one of the parental species such that total transcript levels in the hybrid was more similar to transcript levels in one parental species than in the other parental species.

Candidate transcription factor identification

We took a computational approach to identify potential transcription factors that may change in *trans* regulation between the species in our analysis. We used motif enrichment programs to find potential binding sites in the upstream regions of genes changing in regulation in *D. sechellia* and *D. simulans*. We omitted *D. mauritiana* from this analysis because the *D. mauritiana* genome is not as well annotated as the genomes for *D. simulans* and *D. sechellia*. We used the Differential Enrichment mode in MEME (Bailey and Elkan 1994) as well as the findMotifs.pl script in HOMER (Heinz *et al.* 2010) to identify overrepresented motifs in the regions 500bp upstream of the annotated starting location for genes changing in regulation or with conserved regulation between species in every set of comparisons at stage 2. We utilized a 500bp region as this was empirically determined to give the highest enrichment of signal for motifs in target genes relative to background (see Omura and Lott 2020 for more information). In MEME, we used options to find motifs with any number of repetitions and a motif width of 8-12. We used default options for HOMER and supplied a background FASTA file for enrichment analysis. The background lists supplied were 500bp upstream regions from all annotated genes in the species except for those that were in the target set (either those genes with conserved or changing regulation in any set of comparisons). The 500bp regions were extracted from FASTA files (versions were the same as ones used for mapping) for each species using BEDTools (Quinlan and Hall 2010). Significantly overrepresented motifs in the target lists relative to the background supplied were then compared against databases of known transcription

factor binding sites using Tomtom (MEME suite) and HOMER. All enriched motifs that appeared in both HOMER and MEME analyses are included in Table S1.4. All potential targets of discovered motifs with significant E-values (MEME) or high Match Rank scores in HOMER (>0.8) are also listed in Table S1.4 (see Figure S1.4 for transcript levels of differentially maternally deposited targets in embryos of parental species).

Gene Ontology

Gene ontology (GO) analysis was done with the statistical overrepresentation test in PANTHER (Mi *et al.* 2019) using the default settings. We looked at the GO complete annotations for biological processes and molecular function but did not find any significant terms represented in the cellular component categories. For this analysis, we set a cutoff of Bonferroni adjusted p-value < 0.05. We searched for enrichment of GO categories amongst genes that change in *trans* in each cross, compared to the background of genes that are expressed (having a count >5) in each cross. We used REVIGO (Supek *et al.* 2011) to reduce the number of redundant GO categories and used the small (0.5) level of similarity as a cutoff for redundant GO terms. GO categories shared between two or more crosses at stage 5 are represented in Figure 1.5 and GO categories unique to a cross are shown in Figure S1.5. All enriched categories are listed in Table S1.5.

Data Availability

All sequencing data and processed data files from this study are available at NCBI/GEO at accession number: GSE136646. Supplementary material is available on Figshare.

RESULTS

In order to compare the regulatory basis of evolved changes in gene expression at different stages of early embryogenesis, one stage where all the transcripts are maternally provided and the other after zygotic genome activation (ZGA), we performed a series of crosses between closely related species followed by RNA-seq on resulting embryos (Figure 1.1). We used the sister species *D. simulans*, *D. sechellia*, and *D. mauritiana*, all of which may be crossed reciprocally (with the exception of *D. sechellia* females to *D. simulans* males; Lachaise *et al.* 1986). As transcripts in the two early embryonic stages of interest are produced by different genomes, that of the mother and that of the zygote, we performed crosses to produce a hybrid genome in the appropriate generation (that of the mother or that of the zygote; Figure 1.1). To investigate regulatory changes in zygotic gene expression, the three species were crossed pairwise (with the noted exception), to produce F1 hybrid embryos, which were collected at a stage after zygotic genome activation (end of blastoderm stage or the end of stage 5, Bownes' stages; Bownes 1975; Campos-Ortega and Hartenstein 2013). While the zygotic genome is fully activated at this developmental stage, maternal transcripts are not yet entirely degraded so we limited our analysis to those genes that are expressed at a much higher level after ZGA than before the zygotic genome is activated (see Methods). To discover the regulatory basis of changes in maternal transcript deposition, the F1 females were crossed to males of the same species as the maternal species in the initial cross. Resulting embryos were collected at stage 2 (Bownes' stages; Bownes 1975; Campos-Ortega and Hartenstein 2013), when all the transcripts in the egg are

maternal in origin (Figure 1.1). Three replicate samples were obtained for each cross at stage 2, and since stage 5 features incomplete X chromosomal dosage compensation (Lott *et al.* 2011), 6 replicates were obtained for each cross at late stage 5 (3 female and 3 male embryos with one noted exception, see Methods). mRNA-sequencing libraries were constructed from each embryo sample using poly(A) selection. Libraries were sequenced paired-end, 100bp, on an Illumina HiSeq2500.

Reproducibility of Single Embryo RNA Sequencing Data

Previous studies have shown that single-embryo RNA-seq data is highly reproducible, despite replicate samples representing both biological and technical replicates (Lott *et al.* 2011, 2014; Paris *et al.* 2015; Atallah and Lott 2018). Our current study extends this to include replicates of F1 crosses and subsequent backcrosses between closely related species, which are as reproducible as the within-species replicates. Spearman's rank correlation coefficients are high between replicate samples of the same species or cross at the same developmental stage (Figure 1.2, A,B,D,E, Table S1.3). For example, when comparing the RNA-seq data from stage 2 samples of the (*mau* x *sim*) x *mau* and (*sim* x *mau*) x *sim* hybrid crosses, correlation coefficients range from 0.965 to 0.995 (Table S1.3). Stage 5 hybrids from the *mau* x *sim* cross have equally high correlation coefficients, ranging from 0.980 to 0.996 (Table S1.3). Similarly, correlation coefficients for the RNA-seq data from *D. simulans* stage 5 embryos, when compared with other *D. simulans* stage 5 embryos, range from 0.985 to 0.990. The high correlation coefficients between replicates may be due, in part, to the removal of genes

with differential mapping to either parental genome and those genes with very low transcript abundances (see Methods) from this analysis.

Transcript levels for embryos of the same stage but different genotypes (parental lines and hybrids) are highly similar, as indicated by their Spearman's rank correlation coefficients (Table S1.3), with one notable exception. When we compare RNA-seq profiles from stage 5 hybrids to stage 5 embryos of the paternal species in the cross, we see more divergent patterns of gene expression than when we compare stage 5 hybrids to stage 5 embryos of the maternal species in the cross. This is due to the fact that many maternal transcripts are still present at the zygotic stage, and thus the hybrid zygotic embryo has many remaining maternal transcripts from the maternal species, not the paternal species, in the cross. For example, comparisons between *D. simulans* stage 5 embryos and stage 5 embryos of the *sim* x *mau* cross, where *D. simulans* is the maternal species in the cross, yield high correlation coefficients, ranging from 0.955 to 0.972. In contrast, correlation coefficients are much lower when comparing *sim* x *mau* stage 5 hybrid embryos to stage 5 embryos of the paternal species in the cross, *D. mauritiana*, ranging from 0.863 to 0.887. In this particular comparison, the lower correlation coefficients are likely due to having *D. simulans* as the maternal species in the hybrid cross for the *sim x mau* embryos. Remaining maternal transcripts are from the *D. simulans* alleles and likely explain why these hybrid embryos correlate more highly with *D. simulans* stage 5 embryos.

In contrast to highly correlated samples within a stage, comparing the transcript abundance of different stages yields strikingly lower correlation coefficients (Figure 1.2C and F, Table S1.3), emphasizing the turnover of transcripts between these stages. For

example, when comparing stage 2 hybrids from crosses with *D. mauritiana* and *D. simulans* to stage 5 hybrids from the same cross, correlation coefficients range from 0.483 to 0.573 (Table S1.3). The correlation coefficients are lower when comparing transcript abundances of embryos of different stages than when comparing embryos within a stage, indicating that the pool of transcripts present at the maternal stage is different from that at the zygotic stage of development.

The above finding is reinforced by principal component analysis (PCA; Figure S1.1) for RNA expression profiles of the samples in each set of pairwise comparisons between species. We found that the first principal component corresponds to developmental stage and explains between 80.65% and 81.86% of the variance in the three sets of comparisons. The second principal component of this PCA accounts for between 6.94-8.44% of the variance in the three sets of pairwise comparisons between species and corresponds to genotype. This indicates that there is a more substantial difference between pools of transcripts at the maternal and the zygotic stage of development than there is between the pools of transcripts present in embryos of different genotypes (parental species and hybrids) at the same developmental stage.

Regulatory changes at the maternal stage of development

Changes in gene expression can occur at many levels of regulation: transcriptional, post-transcriptional, translational, or post-translational. Here, we address whether at the transcriptional level, changes in gene expression between species occurred due to changes in *cis* or in *trans* and whether the pattern of regulatory changes differs based on developmental stage. Changes in *cis* regulation can occur

through changes in the DNA of regulatory regions proximal to the gene that they regulate. These types of regulatory changes have an allele-specific effect on gene expression. In contrast, changes in *trans* regulation typically occur via changes in factors that bind to the DNA, such as transcription factor proteins. Changes in *trans* regulation affect the expression of both alleles.

In order to determine regulatory changes in *cis* and in *trans* that lead to differences in maternal transcript deposition between *D. simulans*, *D. sechellia*, and *D. mauritiana*, we used Poisson Gamma (PG) models (León-Novelo *et al.* 2014). This allowed us to determine mapping bias as well as differential expression between alleles, between parental lines and within hybrid embryos (see Methods). We identified differential maternal transcript deposition between the parental lines as well as between species-specific alleles in the stage 2 embryos produced by hybrid mothers. We then compared the two sets of analyses to determine the proportion of *cis* and *trans* regulatory changes underlying differential maternal transcript deposition between species. We used the logic of Landry, *et al*. 2005 to classify genes as having changed in *cis* or in *trans* regulation, by comparing confidence intervals of the bias parameters generated through the PG models (see Methods).

We found that most regulatory changes underlying differentially maternally deposited transcripts occurred in *trans* between each pair of species examined (Figure 1.3A and C, Figure S1.2), where a change in a transcription factor or other *trans*-acting regulatory factor affects both alleles equally (shown in Figure 1.4A and B). In all pairs of comparisons between species, the proportion of *trans* changes was higher than any other category of changes. Comparisons between *D. simulans* and *D. mauritiana* had

the highest percentage of *trans*-only regulatory changes (between 49.4% and 50.8%) while comparisons between *D. simulans* and *D. sechellia* had a lower percentage of regulatory changes solely in *trans* (32.9%). The second highest proportion (between 15.0% and 26.7%) of regulatory changes between species at the maternal stage occurred only in *cis* regulation. Slightly fewer regulatory changes occurred due to a combination of *cis* and *trans* acting factors (between 13.4% and 15.8% in all comparisons). Most genes that change in *cis* and in *trans* regulation are assigned to the *cis* + *trans* category, which indicates that the allele with higher expression in the parental lines is also preserved as the allele with higher expression in the hybrid (the changes in *cis* and in *trans* affect gene expression in the same direction; Landry *et al.* 2005; McManus *et al.* 2010; Coolon *et al.* 2014). We found a smaller proportion of genes changed in regulation through *cis* x *trans* interactions, where changes in *cis* and in *trans* have opposing effects on gene expression and the allele with the lower level of expression in the hybrid is from the parental line with the higher level of expression (Landry *et al.* 2005; McManus *et al.* 2010; Coolon *et al.* 2014). We also found a percentage of genes with conserved levels of maternal transcript deposition between species, between 16.4% and 25.1% in all crosses. *D. simulans* and *D. sechellia* have the highest percentage of conserved genes while *D. simulans* and *D. mauritiana* have the lowest percentage of conserved genes. We also found a small proportion of genes, between 4.2% and 4.7% in all comparisons, that have evolved compensatory mechanisms of regulation, where the genes are not differentially expressed between the parental samples but are differentially expressed in hybrids. This implies that while

transcript levels are the same between species, regulatory changes have occurred which then become visible in the environment of the hybrid.

Genes with expression differences due to *trans* regulatory changes (see Figure 1.4A and B for examples) include regulators with critical functions in important processes governed by maternal gene products, such as *Cdk1*, a cell-cycle regulator necessary for the rapid cleavage cycles in early development (Farrell and O'Farrell 2014).

Binding sites for chromatin modifiers are enriched in the regulatory regions of maternally deposited genes

As *trans* regulatory changes can affect numerous genetic loci, we asked whether there are *trans* regulatory factors that may affect the differential deposition of a number of maternal transcripts between the species studied. For this, we identified binding sites in the predicted *cis*-regulatory regions of all differentially expressed genes and compared them to identified binding sites in the *cis*-regulatory regions of genes with conserved expression between species at the maternal stage. In the pool of genes with altered expression between species, we included not only those with differences in *trans* regulation, but also those with differences in other regulatory categories (*trans*, *cis*, *cis* + *trans*, *cis* x *trans*, and compensatory). This is because genes with changes in *cis* regulation may have had changes that affect the binding of the same *trans* regulators. For genes with differential expression, and separately for genes with conserved expression between species, we took a computational approach. We used both HOMER and MEME (Bailey and Elkan 1994; Heinz *et al.* 2010), to search for

overrepresented motifs within 500 base pairs upstream of transcription start sites, as compared to the rest of the genome (see Methods). We used the upstream regions of genes in *D. simulans* and *D. sechellia* because the *D. mauritiana* genome is not as well annotated, as compared to the other two species in this study.

Interestingly, we found that the *cis*-regulatory regions from both genes with conserved and genes with differential transcript levels between species are enriched in motifs associated with insulator binding (Table S1.4). Specifically, we found that the Dref/BEAF-32 binding site (BEAF-32 and Dref bind overlapping DNA sequences; Hart *et al.* 1999) is the most significantly enriched (Table S1.4). These factors are annotated as insulators (Matzat and Lei 2014; Ali *et al.* 2016) and known to be associated with topologically associated domains (TADs) (Liang *et al.* 2014; Ramírez *et al.* 2018). The binding site for M1BP also appeared significantly enriched in both sets of genes that change in regulation and in ones that are conserved in regulation across species (Table S1.4). M1BP is involved in transcriptional regulation and RNA polymerase II pausing at the promoter of genes (Li and Gilmour 2013), which may also be associated with regulating chromatin state (Ramírez *et al.* 2018). Our findings are consistent with previous studies that identified the enrichment of binding sites for M1BP, BEAF-32 and Dref in the promoter regions of genes that are maternally deposited (Chen *et al.* 2013; Omura and Lott 2020). These binding sites have also been associated with housekeeping genes (Zabidi *et al.* 2015), and the pool of maternal transcripts is enriched with housekeeping genes (Liu *et al.* 2014). However, maternal genes that are not housekeeping genes are even more highly enriched for these binding sites (Omura and Lott 2020), thus the involvement of insulators or other chromatin regulators in

maternal regulation is unlikely solely due to the inclusion of housekeeping genes among maternal genes. Transcript abundance data from our study indicates that *Dref*, *BEAF-32*, and *M1BP* are differentially maternally deposited in several between-species comparisons (Figure S1.4), although in certain crosses, hybrid reads mapped in a biased way to *Dref*, *BEAF-32* and *M1BP*, and thus they were excluded from our regulatory analysis. As the motifs for these *trans*-acting factors are significantly enriched in the upstream regions of all maternal genes, relative to upstream regions of all annotated genes, they are likely important regulators of transcription during oogenesis, and therefore also likely targets of regulatory evolution between species.

Evolution of regulation for zygotically expressed genes

To determine the regulatory basis of changes in zygotic transcript abundance between species, we compared expression levels in late stage 5 parental species samples to late stage 5 hybrid samples and used PG models to identify *cis* and *trans* regulatory changes, similar to our maternal analysis (see Methods). We limited our analysis at the zygotic stage to those genes that are mostly-zygotic: zygotically expressed but not maternally deposited (zygotic-only) or expressed at the zygotic stage at an 8-fold higher level when compared to the maternal stage (we will refer to these as mostly zygotic genes, see Methods).

While we found that most gene expression changes at the maternal stage of development are due to changes in *trans* regulation between the three sister species, we see strikingly different patterns of regulatory changes after ZGA. At the zygotic stage, differences in gene expression between the three species examined occur

mostly due to regulatory changes in both *cis* and *trans*, either by *cis* + *trans* or *cis* x *trans* interactions (Figure 1.3B and D, Figure S1.3). Changes in both *cis* and *trans* regulatory elements (either *cis* + *trans* or *cis* x *trans* interactions) account for expression differences in 39% to 47% of zygotic genes at stage 5 in our between-species comparisons. We also see a higher proportion of these interactions occurring in a *cis* + *trans* pattern (between 29% and 35% of all genes) as opposed to a *cis* x *trans* pattern (between 9% and 12% of all genes) of regulatory interactions. In contrast, *cis*-only and *trans*-only changes account for a smaller number of differences in gene expression levels at this stage in development. In all comparisons, we found between 15% and 21% of genes changing only in *trans* regulation. There are between 16% and 30% of genes that change only in *cis* regulation between each pair of species compared at this stage in development. Compared to the maternal stage, we found a larger proportion of genes with compensatory changes (between 7% and 10% of all genes) in gene regulation and a smaller proportion of genes that are conserved (between 6% and 8% of all genes) between each pair of species comparisons. The smaller proportion of genes with conserved transcript levels at the zygotic stage compared to the maternal stage is consistent with earlier findings showing maternal transcripts to be more highly conserved between species than zygotic transcripts (Atallah and Lott 2018). Examples of evolved changes include regulators critical to important early zygotic processes, such as gap gene *Kruppel* and pair-rule gene *sloppy paired 1* (Figure 1.4C and D), which are required for segmentation along the anterior-posterior axis (Nüsslein-Volhard and Wieschaus 1980; Grossniklaus *et al.* 1992).

Transcriptional regulation at the maternal stage may be broadly determined by regulation at the level of chromatin, as evidenced in this work and by another study (Omura and Lott 2020). In contrast, regulation at the zygotic stage can be gene or pathway specific and involve transcription only in a spatially localized subset of cells (Jäckle *et al.* 1986; Johnston and Nüsslein-Volhard 1992). As such, if a *trans* regulator changed at the zygotic stage, it may affect genes involved in a specific developmental process. For these reasons, we wanted to ask if genes whose zygotic expression differed between species due to changes solely in *trans* regulation had a specific molecular function or were part of a particular biological process. We used PANTHER (Mi *et al.* 2019) to perform gene ontology (GO) analysis on genes changing only in *trans* regulation in each pairwise comparison of species at stage 5 (see Methods). Identifying GO categories over multiple crosses identifies the types of genes that evolve changes repeatedly over evolution. Shared categories were broad, and included those related to DNA binding, positive regulation of transcription by RNA polymerase II, cell fate determination, and several developmental categories (Figure 1.5). The range in GO categories represented, while broadly important at this developmental timepoint, demonstrate how genes changing in *trans* are distributed across developmental processes. As may be expected for zygotic genes at this stage in development, this finding suggests that changes in *trans* regulators of zygotic genes can affect a broad range of molecular and developmental processes. We also investigated biological process categories unique to each specific cross, these are primarily known developmental processes, and are represented in Figure S1.5 and Table S1.5.

Modes of Inheritance in Hybrids

Misexpression in hybrid offspring has been used to examine regulatory incompatibilities that may contribute to speciation (Michalak 2003; Ortíz-Barrientos *et al.* 2006; Moehring *et al.* 2007; Mack *et al.* 2016). As the maternal and zygotic transcripts examined during embryogenesis showed different patterns of gene expression evolution, we also asked whether there were more hybrid incompatibilities present at one stage than the other by looking at whether these two developmental stages showed different levels of transcript misexpression. One way to identify misexpression in hybrids is to compare the inheritance of transcript levels in hybrids to transcript levels in each parental species. Here, we quantify the total transcript abundance for a gene by summing the levels of both species-specific alleles in the hybrid and comparing this level to the transcript abundance in both parental lines. We used methods developed by Gibson, *et al*., 2004 to define modes of inheritance in our hybrids; as in previous studies (Gibson *et al.* 2004; Landry *et al.* 2005; McManus *et al.* 2010; Coolon *et al.* 2014), we used a conservative fold change of 1.25 ($\log_2$-fold change of 0.32) to define those genes that do not change in the total transcript abundance between genotypes (the conserved category represented in Figure 1.6). Genes with transcript levels that are higher (overdominant) or lower (underdominant) in the hybrid relative to either parental species are categorized as misexpressed. The total transcript abundance in the hybrids can also be more similar to one parent versus the other. Here, we categorize the parental line with expression most similar to the hybrid as the dominant parent.

Expression in the hybrid can also have a level intermediate to both parental species (additive).

Both the maternal and zygotic stages show a large proportion of genes where one species' allele is dominant (Figure 1.6, Table S1.6). In both stages, the species that is dominant in each set of crosses is consistent. A higher proportion of genes that are dominant have *D. simulans*-like expression (in any cross involving *D. simulans*) in comparison to the proportion that have expression more like the other parental line in the cross (see Table S1.6 for percentages). We found that *D. mauritiana* has the least dominance in any cross involving this species. Taken together, our findings indicate that *D. simulans* has the most dominant effect on gene expression at both developmental stages, while *D. mauritiana* has the least dominant effect, with dominance in *D. sechellia* falling between the other two species. While there has been previous work proposing relationships between the proportion of dominance and the physiology of unique species (McManus *et al.* 2010), it is difficult to determine any known factors between these three species that would predict this pattern of relative dominance (species range, effective population size, egg size/maternal investment). It is, however, interesting that while the proportion of changes in *cis* and *trans* vary considerably between stages, both stages have dominance among the largest categories of modes of inheritance, and that the relative patterns of which species are dominant is conserved.

Strikingly, while many genes show conservation of expression levels between parental species and in the hybrids at both developmental stages, we found a much higher percentage of conserved transcript levels between parents and hybrids for genes

that are maternally deposited (Figure 1.6, Table S1.6). We found a high proportion of genes with conserved transcript levels at stage 2 in all crosses, between 15.4% and 31.4% of all genes. In contrast, in stage 5 crosses we found conserved transcript abundance in between 4% and 8% of all genes that are either zygotic-only or are mostly zygotic (see Methods for definitions). While there is a large difference in the percentage of conserved genes between the two stages, our stage 5 analysis is limited to those genes with much higher expression at the zygotic stage in comparison to the maternal stage of development. There may be more genes that are mostly zygotic or zygotic-only that are misregulated at this stage in development relative to all of the genes that are expressed at stage 5.

In contrast to maternal expression patterns, we found more genes that have an additive mode of inheritance or that are misexpressed in the hybrids for zygotic genes (Figure 1.6, Table S1.6). Previous studies indicate that additive inheritance is associated with *cis* regulatory divergence (Lemos *et al.* 2008; McManus *et al.* 2010). This is consistent with our findings that a larger proportion of genes at the zygotic stage have expression divergence due, in part or wholly, to *cis* regulatory changes and that more zygotic genes show an additive pattern of inheritance. Higher levels of misexpression at the zygotic stage, taken together with lower conservation of transcript levels at the zygotic stage, suggests that zygotic genes may contribute more to genetic incompatibilities than maternal genes.

DISCUSSION

In this study, we asked whether evolution of gene regulation differs at different developmental stages. We found striking differences in the proportions of *cis* and *trans* regulatory changes between the stage of embryogenesis where all transcripts are maternally derived, and a stage just a few hours later after the zygotic genome has been activated. Between the species examined, we uncovered an overwhelming number of *trans* regulatory changes resulting in differential maternal transcript levels, whereas a complex mix of *cis*, *trans*, and the combination of the two were responsible for changes in zygotic transcription of mostly zygotic genes (see Methods). Here, we propose that the differences in the patterns of gene regulatory evolution between the stages we examined may be due to fundamental differences in the biological context and regulatory architecture producing the transcriptomes present at these stages.

Maternal transcripts are produced by support cells called nurse cells during oogenesis and are either transported by microtubule-dependent mechanisms or dumped into the oocyte along with the cytoplasmic contents of the nurse cells upon their apoptosis (Kugler and Lasko 2009). Many aspects of maternal provisioning have been well-studied in *D. melanogaster*, including transport of transcripts into the oocyte (Mische *et al.* 2007), localization of transcripts within the oocyte (Theurkauf and Hazelrigg 1998), translational regulation (Salles *et al.* 1994), and degradation of maternal transcripts (Tadros *et al.* 2007; Bushati *et al.* 2008; Laver *et al.* 2015). Surprisingly, how transcription is regulated in the nurse cells is not well understood. Nurse cells are polyploid, and are able to rapidly transcribe a large quantity of RNA that represents a large proportion of the genome (Tadros *et al.* 2007; De Renzis *et al.* 2007;

Thomsen *et al.* 2010; Lott *et al.* 2011; Vastenhouw *et al.* 2019) to provide the oocyte with the large stock of transcripts needed. The oocyte itself is thought to be largely transcriptionally silent (Navarro-Costa *et al.* 2016). What we found here, and what was also found in another study investigating binding motifs in maternal factors across the *Drosophila* genus using computational methods (Omura and Lott 2020), is that maternal transcription is associated with *trans* factors annotated to be insulators and that interact with topologically associated domains (TADs). This provides evidence that maternal transcription may be controlled broadly at the level of chromatin state. In this context, we predict that changes in only a few *trans* factors can be responsible for the bulk of the between-species changes in maternal transcription. Thus, changes in the levels of *trans* regulators at this stage may easily be responsible for changes in transcription level for a large number of genes.

In contrast to the large proportion of regulatory changes in *trans* at the maternal stage, differences in zygotic gene transcription for genes that are mostly zygotic (see Methods) between these species is predominantly explained by a combination of changes in *cis*, *trans*, *cis* + *trans*, and *cis* x *trans*. Zygotic gene transcription for genes without a maternal complement is fundamentally different than maternal gene transcription. Unlike the bulk transcription that takes place in the nurse cells, zygotic gene transcription is precisely regulated at the spatial and temporal level across the embryo with enhancer regions playing a large role in where, when and how much genes are expressed (Haines and Eisen 2018). Due to these fundamental differences in gene regulation, the embryo at the zygotic stage may be more sensitive to changes in gene expression than at the maternal stage. Specifically, changes in *trans* regulation, which

can affect the expression of many genes, may be detrimental to the developing organism at this stage. In contrast, changes in *cis* regulation are gene-specific and may only affect gene expression in a subset of the embryo, which might be a more precise way of fine-tuning zygotic expression. We propose that fundamental differences in the regulatory landscape, and perhaps also the developmental role, of the maternal versus the zygotic stage likely explain why the evolution of gene expression occurs through different mechanisms for transcripts that are maternally deposited and genes that are primarily zygotic.

While this study was directed at understanding the regulatory basis of evolution in gene expression at the maternal and zygotic stages of embryogenesis, it also provides insight into the relative conservation of gene expression, both between species and between parent and hybrid offspring. Here, in both the analysis of regulatory changes and the analysis of modes of inheritance, we found more genes with conserved transcript levels among those that are maternally deposited relative to those that are zygotically transcribed. This is in agreement with previous studies that identified high conservation of maternally deposited transcripts relative to those transcribed zygotically between species (Heyn *et al.* 2014; Atallah and Lott 2018) and indicates that the maternal stage is highly conserved. We observe lower conservation of transcript levels at the zygotic stage. A caveat our gene expression analysis at the zygotic stage is that we had to remove genes that still have a large maternal component at this stage (roughly 50% of total transcript pool at late stage 5 is maternally derived; Lott *et al.* 2014). Thus, our finding is best viewed as genes whose transcripts are primarily zygotic at stage 5, have a higher rate of evolutionary change. Additionally, the large proportion

of genes with conserved transcript levels at the maternal stage may be unexpected considering that there is substantial post-transcriptional regulation of maternally deposited factors (Tadros *et al.* 2007; Rouget *et al.* 2010; Barckmann and Simonelig 2013), so it is not clear that a high degree of conservation at the transcript level should be necessary to maintaining conservation at the protein level. Alternatively, if the maternal genome is primarily regulated at the level of chromatin state, this may be a mechanistic constraint on evolution at the level of gene expression. It may be functionally difficult for a gene located in a region of open chromatin to be repressed, or for a gene in a region of heterochromatin to gain expression. Thus, it may be easier to evolve differences in expression over evolutionary time via post-transcriptional mechanisms for maternal genes. Further study is needed to disentangle conservation at the transcript and protein levels of maternal factors across species.

In addition to the differences in conservation between stages, we also found differences in the patterns of inheritance of gene expression between species at the maternal and zygotic stages of embryogenesis. The zygotic stage has a larger proportion of additive differences, which some previous theory (Gibson *et al.* 2004) and empirical studies (Lemos *et al.* 2008; McManus *et al.* 2010) have suggested may be more likely to be changes in *cis* regulation. This would be consistent both with the increased relative role of *cis* changes at the zygotic stage compared to the maternal stage found here, as well as what is known about zygotic gene regulation more generally (Mannervik 2014). In addition, a larger proportion of changes at the zygotic stage fall into the broad category of misregulation (underdominant, overdominant), which have been proposed to increase with divergence time (Coolon *et al.* 2014) and

may be a potential source of hybrid incompatibility between species (Michalak 2003; Ortíz-Barrientos *et al.* 2006; Moehring *et al.* 2007; Mack *et al.* 2016).

In this study, we found that differences between species in levels of maternally deposited transcripts and zygotically transcribed genes evolve via different patterns of regulatory change. We found that maternal transcript abundance is more conserved but when changes do occur, they occur more frequently through *trans* regulation in comparison to zygotic complements. Regulatory organization, constraints, and developmental processes that are specific to each developmental stage likely play a large role in determining how gene regulation can evolve at these two embryonic timepoints. Further study is needed to characterize the molecular basis of evolved changes in transcript level on a single gene level, and more generally to determine what is controlling the regulatory landscape at each stage in development.

ACKNOWLEDGMENTS

We would like to thank members of the Lott lab for comments on the manuscript, as well as members of the UC Davis fly community for their contributions. We would also like to thank Amanda Crofton and Anthony Le for their assistance with genotyping and RNA extraction and Graham Coop for helpful discussions. This work was supported by the National Institute of General Medical Sciences of the National Institutes of Health grant R01GM111362 and the Floyd and Mary Schwall Fellowship.

REFERENCES

Ali T., R. Renkawitz, and M. Bartkuhn, 2016 Insulators and domains of gene expression. Current Opinion in Genetics & Development 37: 17–26. https://doi.org/10.1016/j.gde.2015.11.009

Ali-Murthy Z., S. E. Lott, M. B. Eisen, and T. B. Kornberg, 2013 An essential role for zygotic expression in the pre-cellular Drosophila embryo. PLoS genetics 9: e1003428. https://doi.org/10.1371/journal.pgen.1003428

Ashburner M., K. G. Golic, and R. S. Hawley, 1989 *Drosophila: A Laboratory Handbook*. Cold Spring Harbor Laboratory Press, 2005.

Atallah J., and S. E. Lott, 2018 Evolution of maternal and zygotic mRNA complements in the early Drosophila embryo, (K. A. Dyer, Ed.). PLoS genetics 14: e1007838. https://doi.org/10.1371/journal.pgen.1007838

Bailey T. L., and C. Elkan, 1994 Fitting a mixture model by expectation maximization to discover motifs in biopolymers

Barckmann B., and M. Simonelig, 2013 Control of maternal mRNA stability in germ cells and early embryos. Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms 1829: 714–724. https://doi.org/10.1016/j.bbagrm.2012.12.011

Baroux C., D. Autran, C. S. Gillmor, D. Grimanelli, and U. Grossniklaus, 2008 The Maternal to Zygotic Transition in Animals and Plants. Cold Spring Harbor Symposia on Quantitative Biology 73: 89–100. https://doi.org/10.1101/sqb.2008.73.053

Bownes M., 1975 A photographic study of development in the living embryo of Drosophila melanogaster. Development 33: 789–801.

Britten R. J., and E. H. Davidson, 1969 Gene Regulation for Higher Cells - a Theory. Science (New York, N.Y.) 165: 349-+. https://doi.org/10.1126/science.165.3891.349

Buchberger, Reis, Lu, and Posnien, 2019 Cloudy with a Chance of Insights: Context Dependent Gene Regulation and Implications for Evolutionary Studies. Genes 10: 492–21. https://doi.org/10.3390/genes10070492

Bushati N., A. Stark, J. Brennecke, and S. M. Cohen, 2008 Temporal reciprocity of miRNAs and their targets during the maternal-to-zygotic transition in Drosophila. Current Biology 18: 501–506. https://doi.org/10.1016/j.cub.2008.02.081

Campos-Ortega J. A., and V. Hartenstein, 2013 *The Embryonic Development of Drosophila melanogaster*. Springer Science & Business Media.

Carroll S. B., 1995 Homeotic Genes and the Evolution of Arthropods and Chordates. Nature 376: 479–485. https://doi.org/10.1038/376479a0

Chan Y. F., M. E. Marks, F. C. Jones, G. Villarreal, M. D. Shapiro, *et al.*, 2010 Adaptive Evolution of Pelvic Reduction in Sticklebacks by Recurrent Deletion of a Pitx1 Enhancer. Science (New York, N.Y.) 327: 302–305. https://doi.org/10.1126/science.1182213

Chen K., J. Johnston, W. Shao, S. Meier, C. Staber, *et al.*, 2013 A global change in RNA polymerase II pausing during the Drosophila midblastula transition. Elife 2: e00861. https://doi.org/10.7554/elife.00861

Clark A. G., M. B. Eisen, D. R. Smith, C. M. Bergman, B. Oliver, *et al.*, 2007 Evolution of genes and genomes on the Drosophila phylogeny. Nature 450: 203–218. https://doi.org/10.1038/nature06341

Coolon J. D., C. J. McManus, K. R. Stevenson, B. R. Graveley, and P. J. Wittkopp, 2014 Tempo and mode of regulatory evolution in Drosophila. Genome Research 24: 797–808. https://doi.org/10.1101/gr.163014.113

Crofton A. E., E. L. Cartwright, A. A. Feitzinger, and S. E. Lott, 2018 Effect of Larval Nutrition on Maternal mRNA Contribution to the Drosophila Egg. G3: Genes|Genomes|Genetics 8: 1933–1941. https://doi.org/10.1534/g3.118.200283

Davidson E. H., and I. S. Peter, 2015 *Genomic control process*.

De Renzis S. D., O. Elemento, S. Tavazoie, and E. F. Wieschaus, 2007 Unmasking Activation of the Zygotic Genome Using Chromosomal Deletions in the Drosophila Embryo. PLOS Biology 5: e117. https://doi.org/10.1371/journal.pbio.0050117

Eichhorn S. W., A. O. Subtelny, I. Kronja, and J. K. Elife, 2016 mRNA poly (A)-tail changes specified by deadenylation broadly reshape translation in Drosophila oocytes and early embryos. cdn.elifesciences.org. https://doi.org/10.7554/elife.16955.001

Farrell J. A., and P. H. O'Farrell, 2014 From Egg to Gastrula: How the Cell Cycle Is Remodeled During the Drosophila Mid-Blastula Transition. Annual Review of Genetics, Vol 48 48: 269–294. https://doi.org/10.1146/annurev-genet-111212-133531

Fischer B. E., E. Wasbrough, L. A. Meadows, O. Randlet, S. Dorus, *et al.*, 2012 Conserved properties of Drosophila and human spermatozoal mRNA repertoires. Proc Royal Soc B Biological Sci 279: 2636–2644. https://doi.org/10.1098/rspb.2012.0153

Gibson G., R. Riley-Berger, L. Harshman, A. Kopp, S. Vacha, *et al.*, 2004 Extensive Sex-Specific Nonadditivity of Gene Expression in Drosophila melanogaster. Genetics 167: 1791–1799. https://doi.org/10.1534/genetics.104.026583

Glaser-Schmitt A., A. Zečić, and J. Parsch, 2018 Gene Regulatory Variation in Drosophila melanogaster Renal Tissue. Genetics genetics.301073.2018-46. https://doi.org/10.1534/genetics.118.301073

Graze R. M., L. M. McIntyre, B. J. Main, M. L. Wayne, and S. V. Nuzhdin, 2009 Regulatory Divergence in Drosophila melanogaster and D. simulans, a Genomewide Analysis of Allele-Specific Expression. Genetics 183: 547–561. https://doi.org/10.1534/genetics.109.105957

Grossniklaus U., R. K. Pearson, and W. J. Gehring, 1992 The Drosophila sloppy paired locus encodes two proteins involved in segmentation that show homology to mammalian transcription factors. Genes & Development 6: 1030–1051. https://doi.org/10.1101/gad.6.6.1030

Haines J. E., and M. B. Eisen, 2018 Patterns of chromatin accessibility along the anterior-posterior axis in the early Drosophila embryo, (J. M. Vaquerizas, Ed.). PLoS genetics 14: e1007367-25. https://doi.org/10.1371/journal.pgen.1007367

Hart C. M., O. Cuvier, and U. K. Laemmli, 1999 Evidence for an antagonistic relationship between the boundary element-associated factor BEAF and the transcription factor DREF. Chromosoma 108: 375–383. https://doi.org/10.1007/s004120050389

Heinz S., C. Benner, N. Spann, E. Bertolino, Y. C. Lin, *et al.*, 2010 Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. Molecular cell 38: 576–589. https://doi.org/10.1016/j.molcel.2010.05.004

Heyn P., M. Kircher, A. Dahl, J. Kelso, P. Tomancak, *et al.*, 2014 The Earliest Transcribed Zygotic Genes Are Short, Newly Evolved, and Different across Species. CellReports 6: 285–292. https://doi.org/10.1016/j.celrep.2013.12.030

Jäckle H., D. Tautz, R. Schuh, E. Seifert, and R. Lehmann, 1986 Cross-regulatory interactions among the gap genes of Drosophila. Nature 324: 668–670. https://doi.org/10.1038/324668a0

Johnston D. S., and C. Nüsslein-Volhard, 1992 The Origin of Pattern and Polarity in the Drosophila Embryo. Cell 68: 201–219. https://doi.org/10.1016/0092-8674(92)90466-p

King M. C., and A. C. Wilson, 1975 Evolution at two levels in humans and chimpanzees. Science (New York, N.Y.) 188: 107–116. https://doi.org/10.1126/science.1090005

Kugler J.-M., and P. Lasko, 2009 Localization, anchoring and translational control of oskar, gurken, bicoid and nanos mRNA during Drosophila oogenesis. Fly 3: 15–28. https://doi.org/10.4161/fly.3.1.7751

Kvon E. Z., O. K. Kamneva, U. S. Melo, I. Barozzi, M. Osterwalder, *et al.*, 2016 Progressive Loss of Function in a Limb Enhancer during Snake Evolution. Cell 167: 633-642.e11. https://doi.org/10.1016/j.cell.2016.09.028

Lachaise D., J. R. David, F. Lemeunier, L. Tsacas, and M. Ashburner, 1986 The Reproductive Relationships of Drosophila sechellia with D. mauritiana, D. simulans, and D. melanogaster from the Afrotropical Region. Evolution 40: 262–271. https://doi.org/10.1111/j.1558-5646.1986.tb00468.x

Landry C. R., P. J. Wittkopp, C. H. Taubes, J. M. Ranz, A. G. Clark, *et al.*, 2005 Compensatory cis-trans evolution and the dysregulation of gene expression in interspecific hybrids of Drosophila. Genetics 171: 1813–1822. https://doi.org/10.1534/genetics.105.047449

Lasko P., 2012 mRNA Localization and Translational Control in Drosophila Oogenesis. Cold Spring Harbor Perspectives in Biology 4: a012294–a012294. https://doi.org/10.1101/cshperspect.a012294

Laver J. D., X. Li, D. Ray, K. B. Cook, N. A. Hahn, *et al.*, 2015 Brain tumor is a sequence-specific RNA-binding protein that directs maternal mRNA clearance during the Drosophila maternal-to-zygotic transition. Genome biology 16: 94. https://doi.org/10.1186/s13059-015-0659-4

Lemos B., L. O. Araripe, P. Fontanillas, and D. L. Hartl, 2008 Dominance and the evolutionary accumulation of cis- and trans-effects on gene expression. Proceedings of the National Academy of Sciences 105: 14471–14476. https://doi.org/10.1073/pnas.0805160105

León-Novelo L. G., L. M. McIntyre, J. M. Fear, and R. M. Graze, 2014 A flexible Bayesian method for detecting allelic imbalance in RNA-seq data. BMC Genomics 15: 920. https://doi.org/10.1186/1471-2164-15-920

Li H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, *et al.*, 2009 The Sequence Alignment/Map format and SAMtools. Bioinformatics 25: 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Li J., and D. S. Gilmour, 2013 Distinct mechanisms of transcriptional pausing orchestrated by GAGA factor and M1BP, a novel transcription factor. The EMBO Journal 32: 1829–1841. https://doi.org/10.1038/emboj.2013.111

Liang J., L. Lacroix, A. Gamot, S. Cuddapah, S. Queille, *et al.*, 2014 Chromatin Immunoprecipitation Indirect Peaks Highlight Long-Range Interactions of Insulator

Proteins and Pol II Pausing. Molecular cell 53: 672–681. https://doi.org/10.1016/j.molcel.2013.12.029

Lim J., M. Lee, A. Son, H. Chang, and V. N. Kim, 2016 mTAIL-seq reveals dynamic poly(A) tail regulation in oocyte-to-embryo development. Genes & Development 30: 1671–1682. https://doi.org/10.1101/gad.284802.116

Liu M. M., J. W. Davey, D. J. Jackson, M. L. Blaxter, and A. Davison, 2014 A conserved set of maternal genes? Insights from a molluscan transcriptome. Int J Dev Biology 58: 501–511. https://doi.org/10.1387/ijdb.140121ad

Lott S. E., J. E. Villalta, G. P. Schroth, S. Luo, L. A. Tonkin, *et al.*, 2011 Noncanonical Compensation of Zygotic X Transcription in Early Drosophila melanogaster Development Revealed through Single-Embryo RNA-Seq. PLOS Biology 9: e1000590. https://doi.org/10.1371/journal.pbio.1000590

Lott S. E., J. E. Villalta, Q. Zhou, D. Bachtrog, and M. B. Eisen, 2014 Sex-specific embryonic gene expression in species with newly evolved sex chromosomes., (H. S. Malik, Ed.). PLoS genetics 10: e1004159. https://doi.org/10.1371/journal.pgen.1004159

Mack K. L., P. Campbell, and M. W. Nachman, 2016 Gene regulation and speciation in house mice. Genome Research gr.195743.115. https://doi.org/10.1101/gr.195743.115

Macmanes M. D., 2014 On the optimal trimming of high-throughput mRNA sequence data. Frontiers in genetics 5: 13. https://doi.org/10.3389/fgene.2014.00013

Mannervik M., 2014 Control of Drosophila embryo patterning by transcriptional co-regulators. Experimental Cell Research 321: 47–57. https://doi.org/10.1016/j.yexcr.2013.10.010

Martin M., 2011 Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal 17: 10–12. https://doi.org/10.14806/ej.17.1.200

Matzat L. H., and E. P. Lei, 2014 Surviving an identity crisis: A revised view of chromatin insulators in the genomics era. Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms 1839: 203–214. https://doi.org/10.1016/j.bbagrm.2013.10.007

McDermott S. R., and R. M. Kliman, 2008 Estimation of Isolation Times of the Island Species in the Drosophila simulans Complex from Multilocus DNA Sequence Data. PLoS ONE 3: e2442. https://doi.org/10.1371/journal.pone.0002442

McManus C. J., J. D. Coolon, M. O. Duff, J. Eipper-Mains, B. R. Graveley, *et al.*, 2010 Regulatory divergence in Drosophila revealed by mRNA-seq. Genome Research 20: 816–825. https://doi.org/10.1101/gr.102491.109

Mi H., A. Muruganujan, X. Huang, D. Ebert, C. Mills, *et al.*, 2019 Protocol Update for large-scale genome and gene function analysis with the PANTHER classification system (v.14.0). Nature Protocols 1–21. https://doi.org/10.1038/s41596-019-0128-8

Michalak P., 2003 Genome-Wide Patterns of Expression in Drosophila Pure Species and Hybrid Males. Molecular Biology and Evolution 20: 1070–1076. https://doi.org/10.1093/molbev/msg119

Mische S., M. Li, M. Serr, and T. S. Hays, 2007 Direct Observation of Regulated Ribonucleoprotein Transport Across the Nurse Cell/Oocyte Boundary, (S. Wente, Ed.). Molecular Biology of the Cell 18: 2254–2263. https://doi.org/10.1091/mbc.e06-10-0959

Moehring A. J., K. C. Teeter, and M. A. F. Noor, 2007 Genome-Wide Patterns of Expression in Drosophila Pure Species and Hybrid Males. II. Examination of Multiple-Species Hybridizations, Platforms, and Life Cycle Stages. Molecular Biology and Evolution 24: 137–145. https://doi.org/10.1093/molbev/msl142

Navarro-Costa P., A. McCarthy, P. Prudêncio, C. Greer, L. G. Guilgur, *et al.*, 2016 Early programming of the oocyte epigenome temporally controls late prophase I transcription and chromatin remodelling. Nat Commun 7: 12331. https://doi.org/10.1038/ncomms12331

Nolte V., R. V. Pandey, R. Kofler, and C. Schlötterer, 2013 Genome-wide patterns of natural variation reveal strong selective sweeps and ongoing genomic conflict in Drosophila mauritiana. Genome Research 23: 99–110. https://doi.org/10.1101/gr.139873.112

Nüsslein-Volhard C., and E. Wieschaus, 1980 Mutations affecting segment number and polarity in Drosophila. Nature 287: 795–801. https://doi.org/10.1038/287795a0

Omura C. S., and S. E. Lott, 2020 The conserved regulatory basis of mRNA contributions to the early Drosophila embryo differs between the maternal and zygotic genomes. Plos Genet 16: e1008645. https://doi.org/10.1371/journal.pgen.1008645

Ortíz-Barrientos D., B. A. Counterman, and M. A. F. Noor, 2006 Gene expression divergence and the origin of hybrid dysfunctions. Genetica 129: 71–81. https://doi.org/10.1007/s10709-006-0034-1

Paris M., J. E. Villalta, M. B. Eisen, and S. E. Lott, 2015 Sex Bias and Maternal Contribution to Gene Expression Divergence in Drosophila Blastoderm Embryos., (A.

G. Clark, Ed.). PLoS genetics 11: e1005592. https://doi.org/10.1371/journal.pgen.1005592

Perry M. W., A. N. Boettiger, and M. Levine, 2011 Multiple enhancers ensure precision of gap gene-expression patterns in the Drosophila embryo. Proceedings of the National Academy of Sciences 108: 13570–13575. https://doi.org/10.1073/pnas.1109873108

Prud'homme B., N. Gompel, A. Rokas, V. A. Kassner, T. M. Williams, *et al.*, 2006 Repeated morphological evolution through cis -regulatory changes in a pleiotropic gene. Nature 440: 1050–1053. https://doi.org/10.1038/nature04597

Quinlan A. R., and I. M. Hall, 2010 BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26: 841–842. https://doi.org/10.1093/bioinformatics/btq033

R Core Team, 2017 *R: A language and environment for statistical computing*.

Ramírez F., V. Bhardwaj, L. Arrigoni, K. C. Lam, B. A. Grüning, *et al.*, 2018 High-resolution TADs reveal DNA sequences underlying genome organization in flies. Nature Communications 9: 1–15. https://doi.org/10.1038/s41467-017-02525-w

Rouget C., C. Papin, A. Boureux, A.-C. Meunier, B. Franco, *et al.*, 2010 Maternal mRNA deadenylation and decay by the piRNA pathway in the early Drosophila embryo. Nature 467: 1128-U144. https://doi.org/10.1038/nature09465

Salles F. J., M. E. Lieberfarb, C. Wreden, J. P. Gergen, and S. Strickland, 1994 Coordinate initiation of Drosophila development by regulated polyadenylation of maternal messenger RNAs. Science (New York, N.Y.) 266: 1996–1999. https://doi.org/10.1126/science.7801127

Schulz K. N., and M. M. Harrison, 2019 Mechanisms regulating zygotic genome activation. Nature Reviews Genetics 1–14. https://doi.org/10.1038/s41576-018-0087-x

Small S., A. Blair, and M. Levine, 1992 Regulation of Even-Skipped Stripe-2 in the Drosophila Embryo. Embo Journal 11: 4047–4057. https://doi.org/10.1002/j.1460-2075.1992.tb05498.x

Supek F., M. Bošnjak, N. Škunca, and T. Šmuc, 2011 REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms, (C. Gibas, Ed.). PLoS ONE 6: e21800-9. https://doi.org/10.1371/journal.pone.0021800

Tadros W., A. L. Goldman, T. Babak, F. Menzies, L. Vardy, *et al.*, 2007 SMAUG Is a Major Regulator of Maternal mRNA Destabilization in Drosophila and Its Translation

Is Activated by the PAN GU Kinase. Developmental Cell 12: 143–155. https://doi.org/10.1016/j.devcel.2006.10.005

Tadros W., and H. D. Lipshitz, 2009 The maternal-to-zygotic transition: a play in two acts. Development 136: 3033–3042. https://doi.org/10.1242/dev.033183

Theurkauf W. E., and T. I. Hazelrigg, 1998 In vivo analyses of cytoplasmic transport and cytoskeletal organization during Drosophila oogenesis: characterization of a multi-step anterior localization pathway. Dev Camb Engl 125: 3655–66.

Thomsen S., S. Anders, S. C. Janga, W. Huber, and C. R. Alonso, 2010 Genome-wide analysis of mRNA decay patterns during early Drosophila development. Genome biology 11: R93. https://doi.org/10.1186/gb-2010-11-9-r93

Torres-Oliva M., I. Almudi, A. P. McGregor, and N. Posnien, 2016 A robust (re-)annotation approach to generate unbiased mapping references for RNA-seq-based analyses of differential expression across closely related species. BMC Genomics 17: 392. https://doi.org/10.1186/s12864-016-2646-x

Trapnell C., A. Roberts, L. Goff, G. Pertea, D. Kim, *et al.*, 2012 Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nature Protocols 7: 562. https://doi.org/10.1038/nprot.2012.016

Van der Auwera G. A. V. der, M. O. Carneiro, C. Hartl, R. Poplin, G. del Angel, *et al.*, 2013 Current Protocols in Bioinformatics. Curr Protoc Bioinform Ed Board Andreas D Baxevanis Et Al 43: 11.10.1-11.10.33. https://doi.org/10.1002/0471250953.bi1110s43

Vastenhouw N. L., W. X. Cao, and H. D. Lipshitz, 2019 The maternal-to-zygotic transition revisited. Development 146: dev161471-20. https://doi.org/10.1242/dev.161471

Wittkopp P. J., B. K. Haerum, and A. G. Clark, 2004 Evolutionary changes in cis and trans gene regulation. Nature 430: 85–88. https://doi.org/10.1038/nature02698

Xu G., J. Lyu, Q. Li, H. Liu, D. Wang, *et al.*, 2020 Adaptive evolution of DNA methylation reshaped gene regulation in maize. Biorxiv 2020.03.13.991117. https://doi.org/10.1101/2020.03.13.991117

Zabidi M. A., C. D. Arnold, K. Schernhuber, M. Pagani, M. Rath, *et al.*, 2015 Enhancer–core-promoter specificity separates developmental and housekeeping gene regulation. Nature 518: 556–559. https://doi.org/10.1038/nature13994

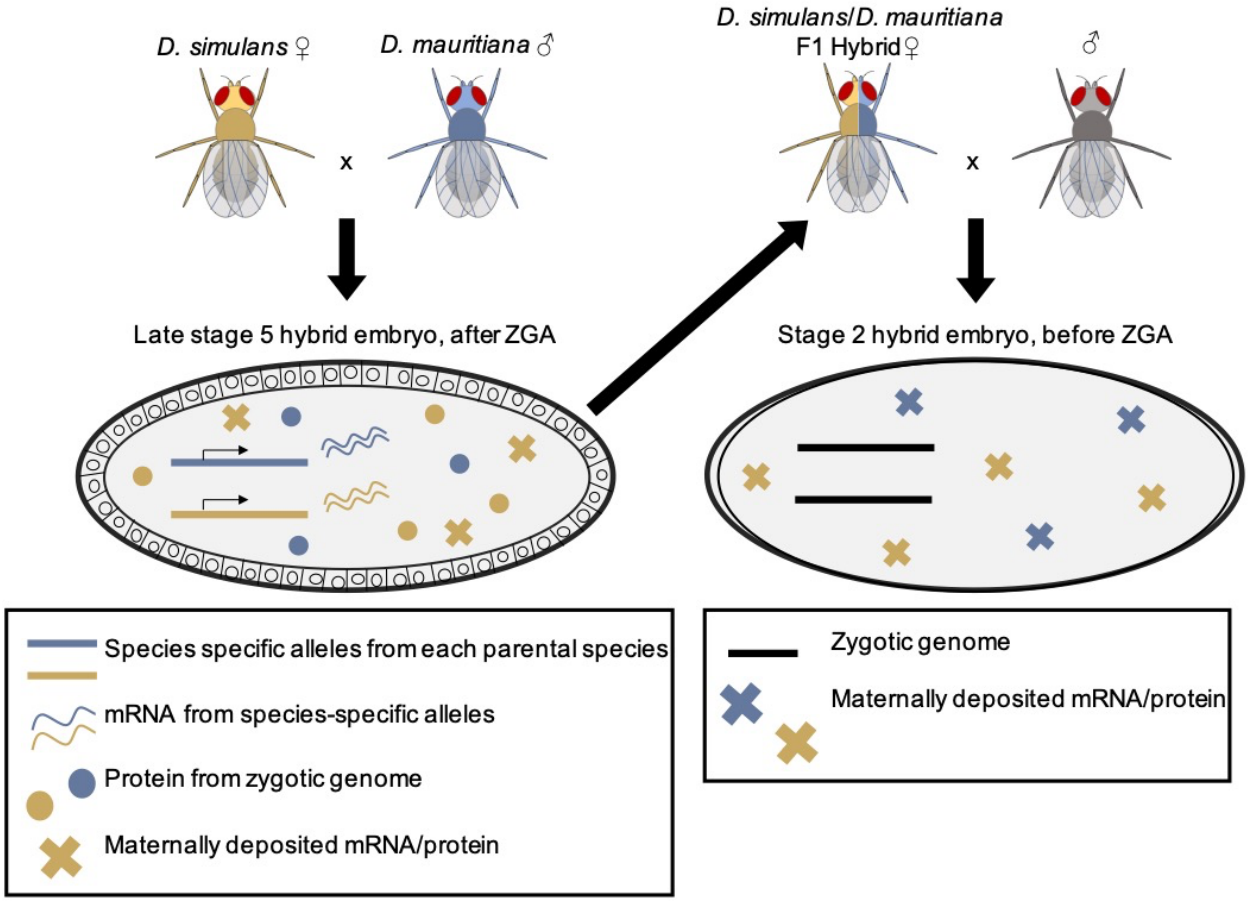


Figure 1.1: Crosses to produce hybrid embryos for the zygotic and maternal stages. To look at changes in regulation for zygotic genes, hybrid stage 5 embryos (left) were produced by crossing two parental species and collecting their eggs at the appropriate stage (late stage 5). To look at regulatory changes in maternal transcript deposition, F1 hybrid mothers were mated to males and stage 2 embryos were collected (right). In both cases, transcription is coming from a F1 hybrid genome, either that of the zygote (left) which is measured after zygotic genome activation (late stage 5) or the mother (right) which is measured when all the transcripts in the embryo are maternally deposited (stage 2).

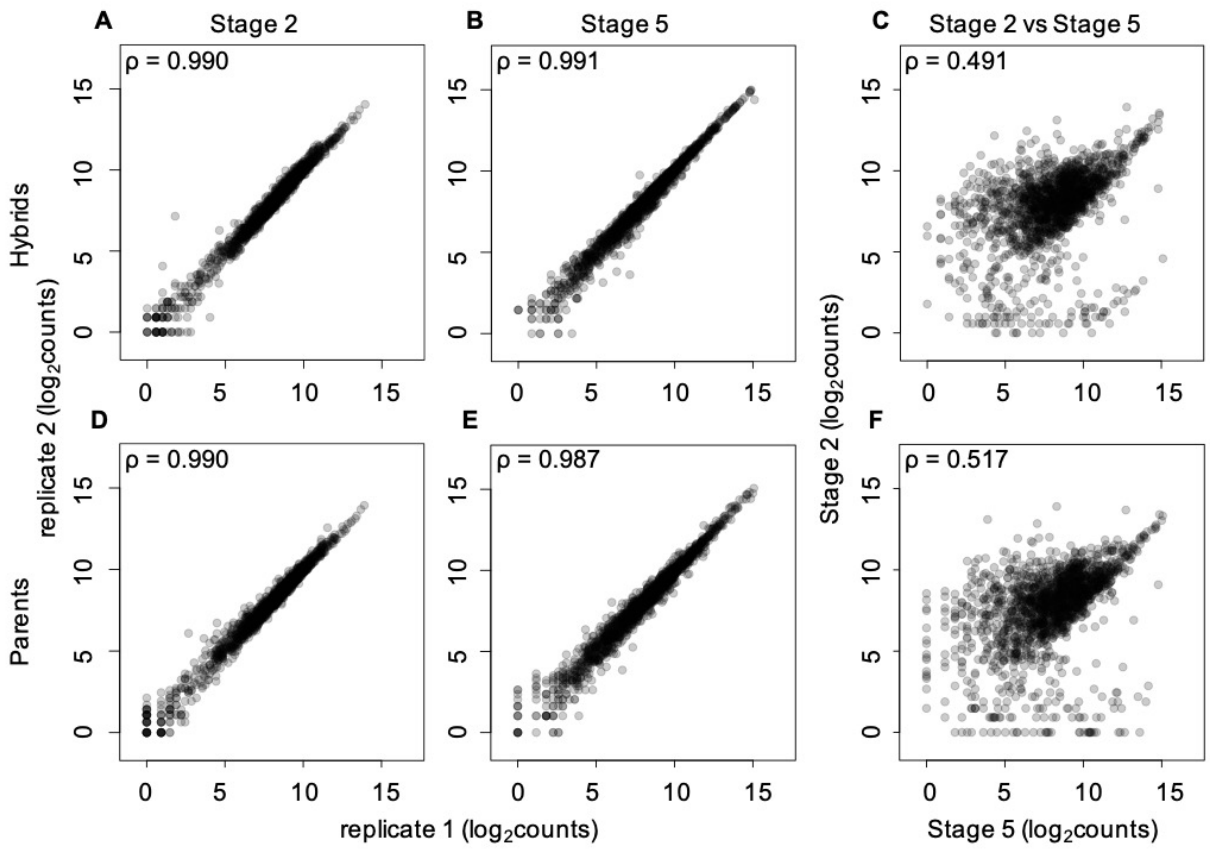


Figure 1.2: Hybrid and parental species single embryo transcript levels are highly reproducible. (A,B,D,E) Spearman's rank correlation coefficients are high when counts from replicate transcriptomes of the same stage and genotype are compared. Correlation coefficients are similarly high in parental species (D,E) and when comparing replicates from hybrid crosses (A,B). (C,F) Samples from different stages and the same genotype have much lower correlations, indicating a large difference in transcriptomes between the maternal and zygotic stages.

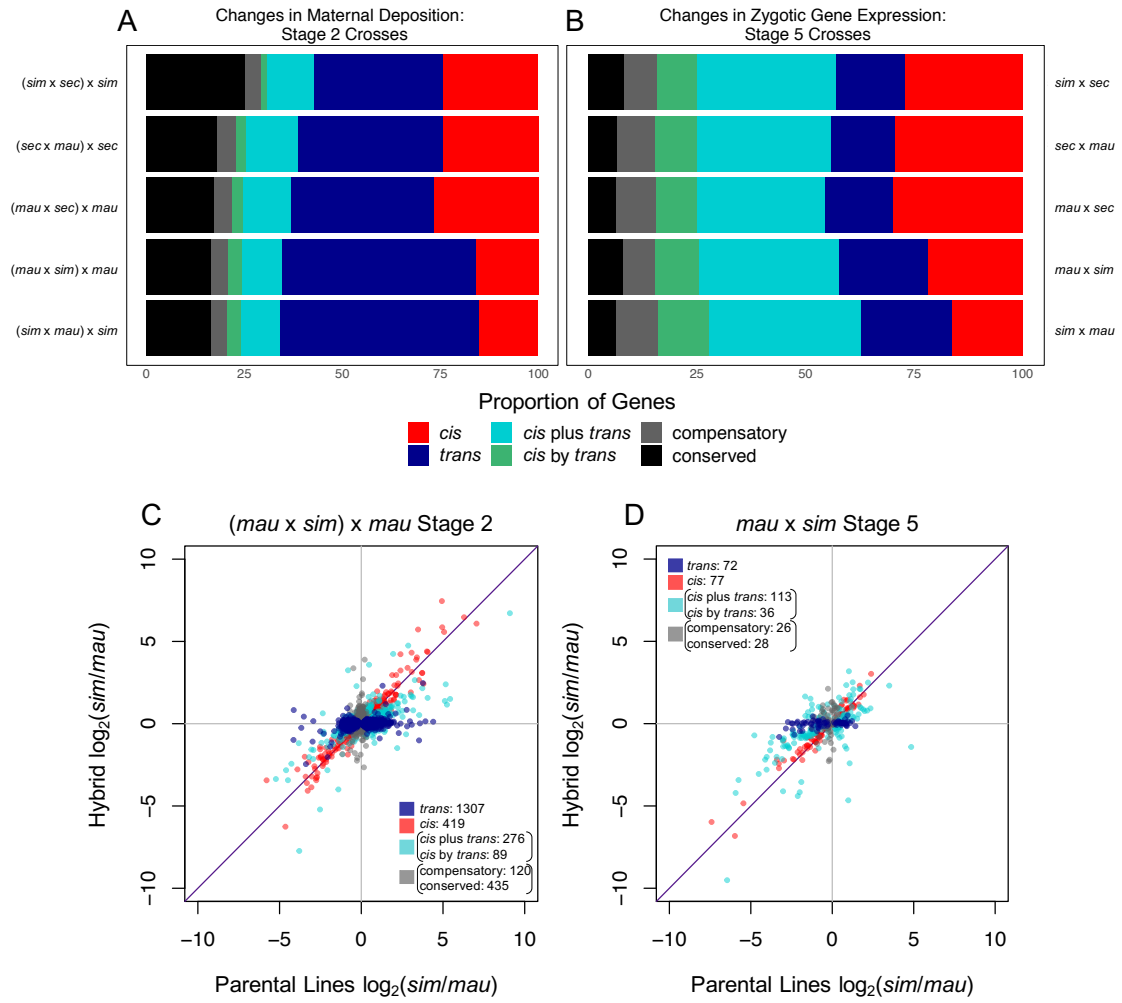


Figure 1.3: Different types of evolved regulatory changes dominate in maternal transcript deposition vs. zygotic transcription. Proportion of genes that fall into categories of regulatory change for each cross are shown for both the maternal transcript deposition (A) and zygotic gene transcription (B), for mostly-zygotic genes. Transcript level ratios between parental lines and within hybrids at stage 2 (C) and stage 5 (D) describe regulatory changes between *D. mauritiana* and *D. simulans* in one direction of crosses (for the rest of the crosses, see Figures S1.2 and S1.3).

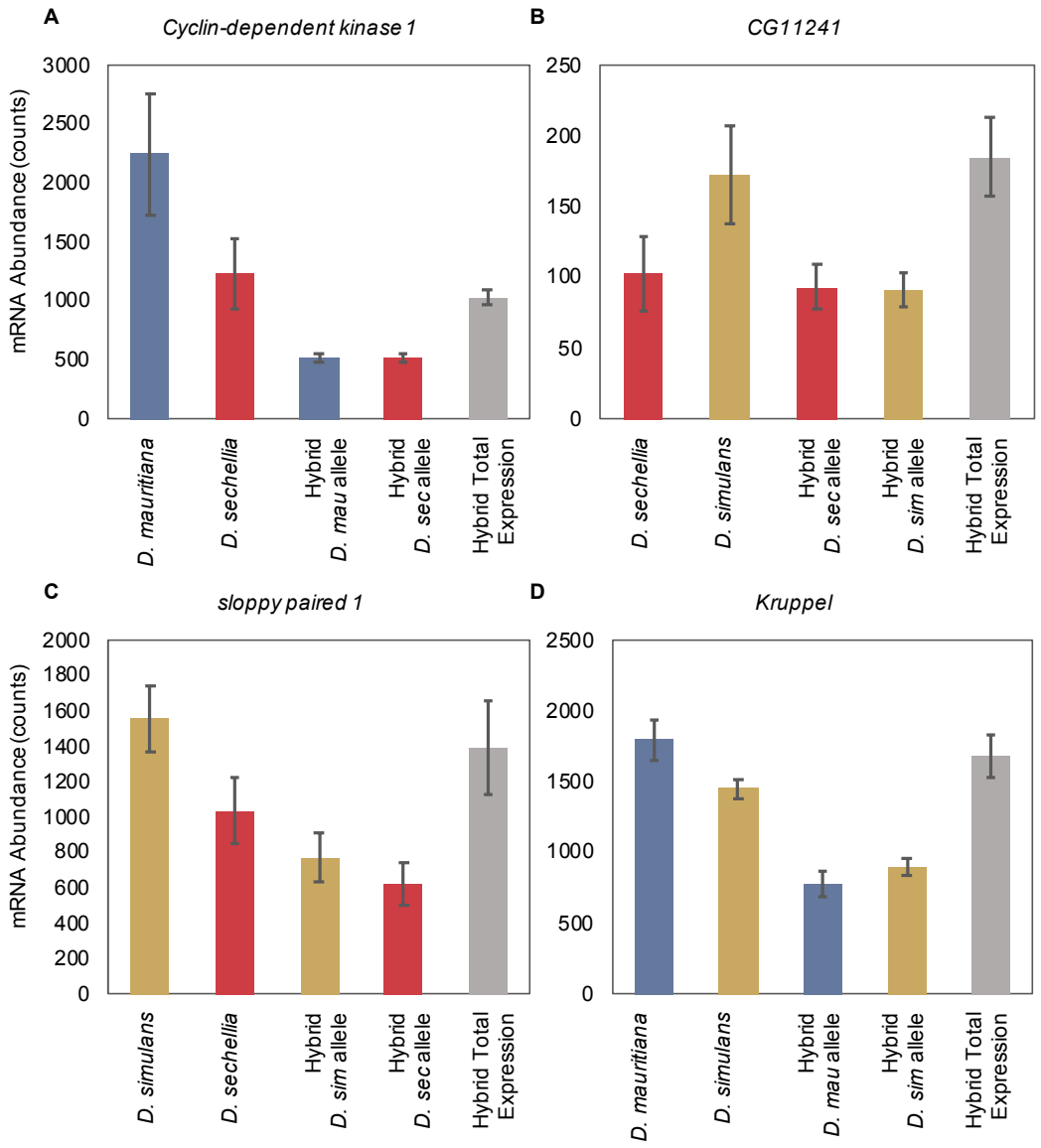


Figure 1.4: Examples of the type of regulatory changes observed, for individual genes. Transcript abundance, shown in counts, for each gene is plotted for the total mRNA abundance in both parental lines and for each parental allele within the hybrid; error bars shown represent the standard deviation. Total transcript abundance in the hybrid (the summation of levels from parental alleles in the hybrid) is shown as the last bar on the right in each graph. (A) Maternal transcript deposition of *Cyclin-dependent kinase 1 (Cdk1),* a critical cell cycle regulator in early development, changes in *trans* regulation between *D. mauritiana* and *D. sechellia*. Hybrid mRNA abundance is from the (*mau* x *sec*) x *mau* cross. (*Cdk1* also changes in *trans* regulation in the reciprocal cross comparison, (*sec* x *mau*) x *sec*.) (B) Maternal transcript deposition of *CG11241*, a gene of currently unknown function, changes in *trans* regulation between *D. sechellia* and *D. simulans*. Hybrid mRNA abundance is from the

(*sim* x *sec*) x *sim* cross. (C) At stage 5 in development, *sloppy paired 1,* a critical pair-rule segmentation gene, changes in regulation through a combination of *cis* and *trans* regulatory changes (*cis* + trans) between *D. simulans* and *D. sechellia*. Hybrid expression is shown for the *sim* x *sec* cross. *Sloppy paired 2* also changes in *cis* regulation between these two species. (D) At stage 5, *Kruppel*, a gap gene crucial to segmentation changes in regulation through a combination of *cis* and *trans* regulatory elements (*cis* x *trans*) between *D. mauritiana* and *D. simulans*. Here, expression in the hybrid is from the *sim* x *mau* cross but in the reciprocal cross (*mau* x *sim*), *Kruppel* also changes in *cis* x *trans* regulation.

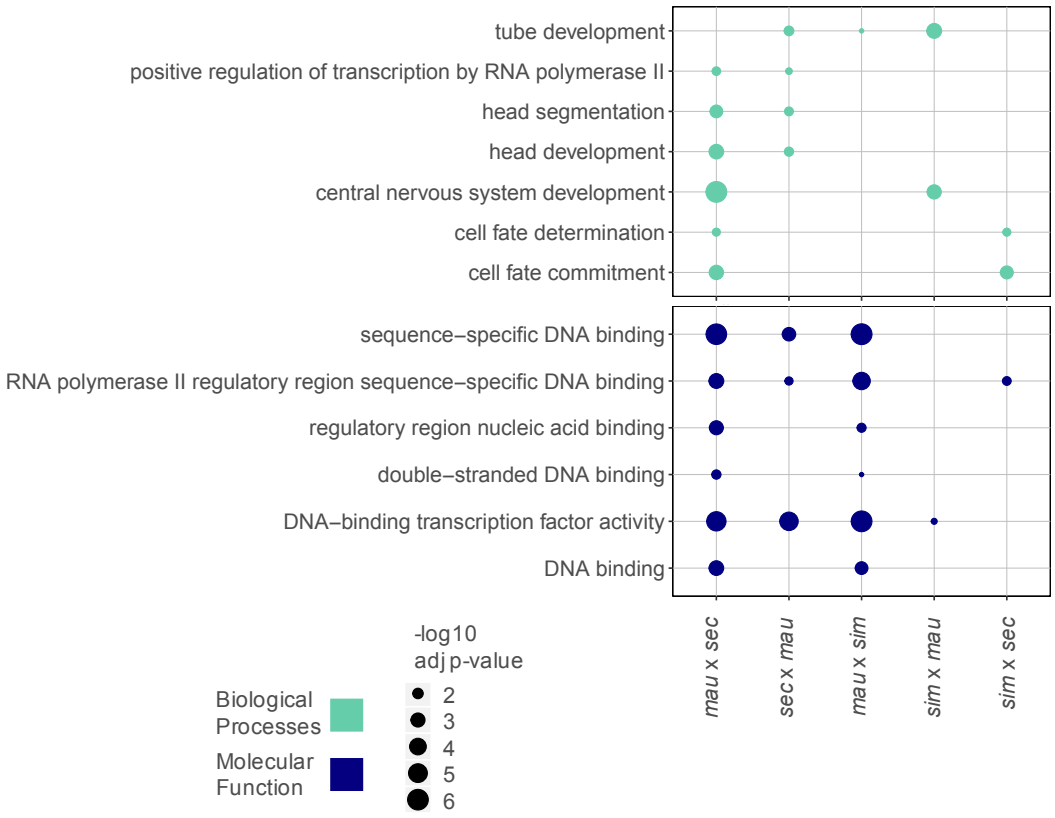


Figure 1.5: Gene ontology (GO) analysis identifies transcription factors that act in developmental processes as types of genes that change zygotically. Significantly enriched GO terms are listed for zygotically transcribed genes that change in *trans* regulation between each pair of species compared. Genes represented in this analysis are categorized as mostly zygotic (see Methods). Terms are listed for Biological Processes and Molecular Function categories and only terms that appear in more than one cross are shown in this figure. Terms unique to a specific cross are listed in Figure S1.5. Biological process categories identified relate to development, molecular function categories identify functions consistent with DNA binding and regulation.

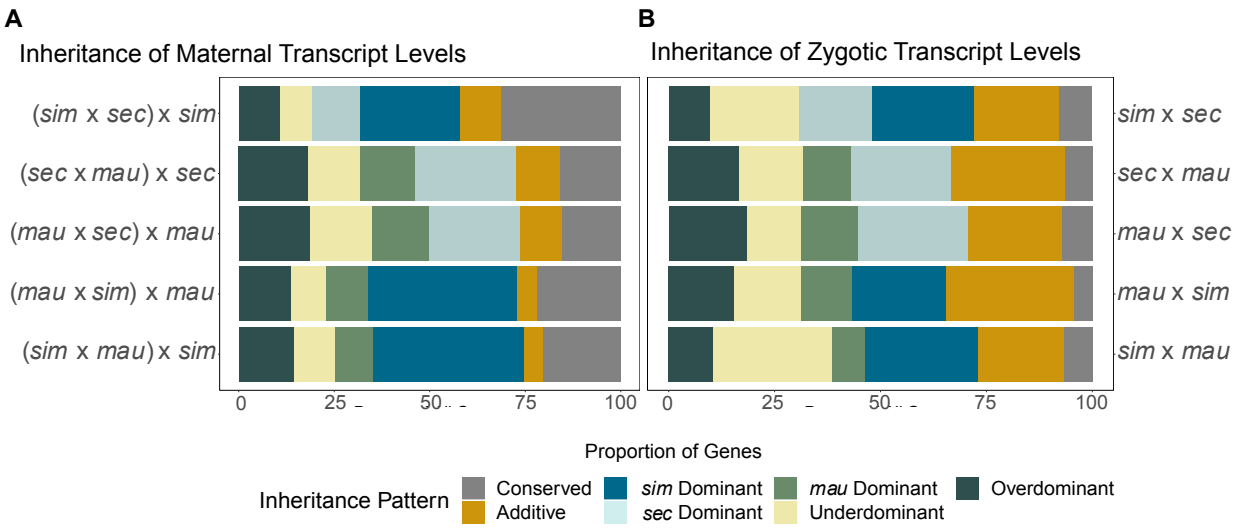


Figure 1.6: Patterns of inheritance show dominance of particular parental genomes at both stages. A) Shows patterns of inheritance for stage 2, over all genes and all crosses. B) Shows patterns of inheritance for stage 5, for mostly zygotic genes (see Methods) and all crosses. The maternal stage (A) shows a higher proportion of conserved genes than the zygotic stage (B). Both stages show a high degree of dominance for *D. simulans* for crosses involving that species, and for *D. sechellia* in crosses with *D. mauritiana*, forming the general dominance pattern of *D. simulans* > *D. sechellia* > *D. mauritiana.* There is a greater proportion of additive inheritance for the zygotic stage (B) than the maternal stage (A).

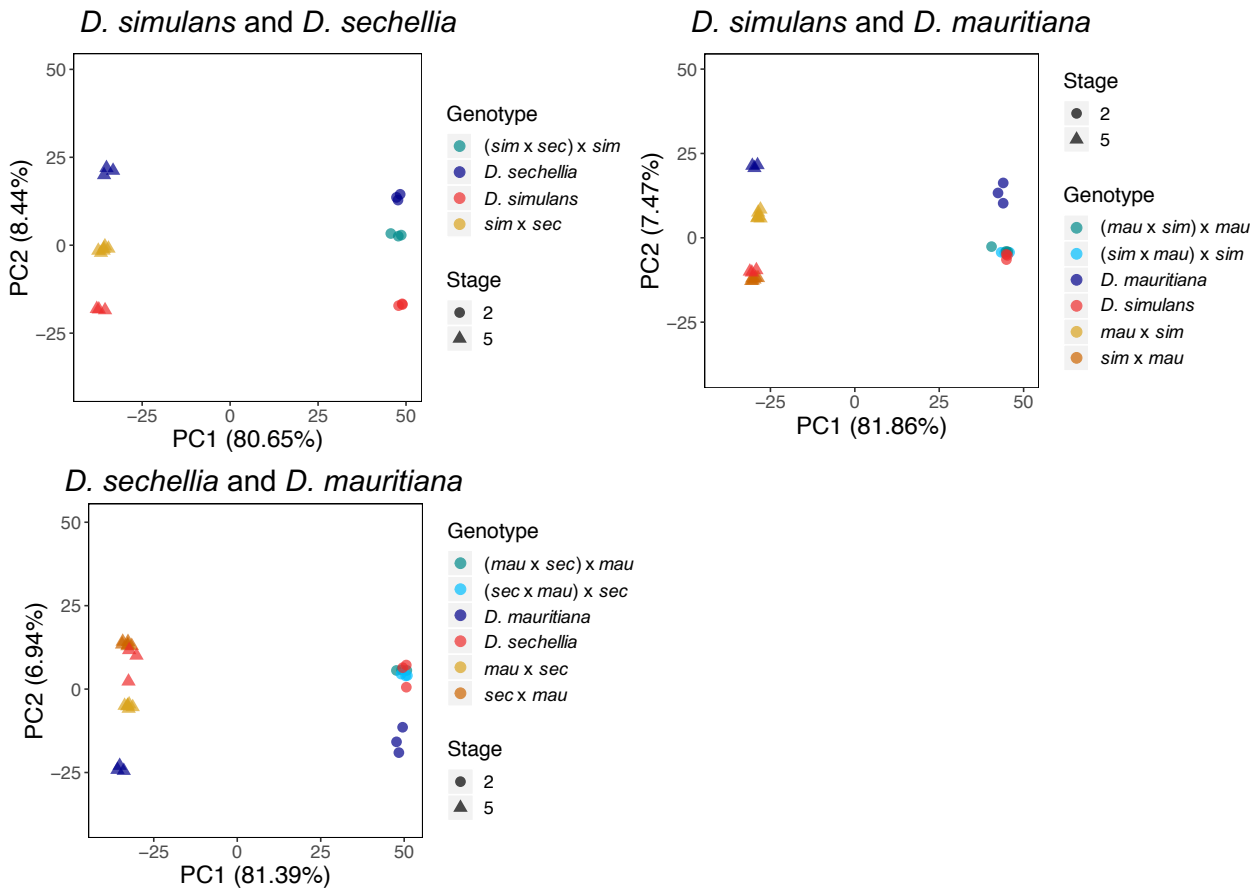


Figure S1.1: PCA plots for transcript abundance in all crosses. Samples of each stage, 2 or 5, cluster together. Samples of each genotype also cluster together, parental samples and hybrids. Proportion of variance explained by each principal component is listed on each axis.

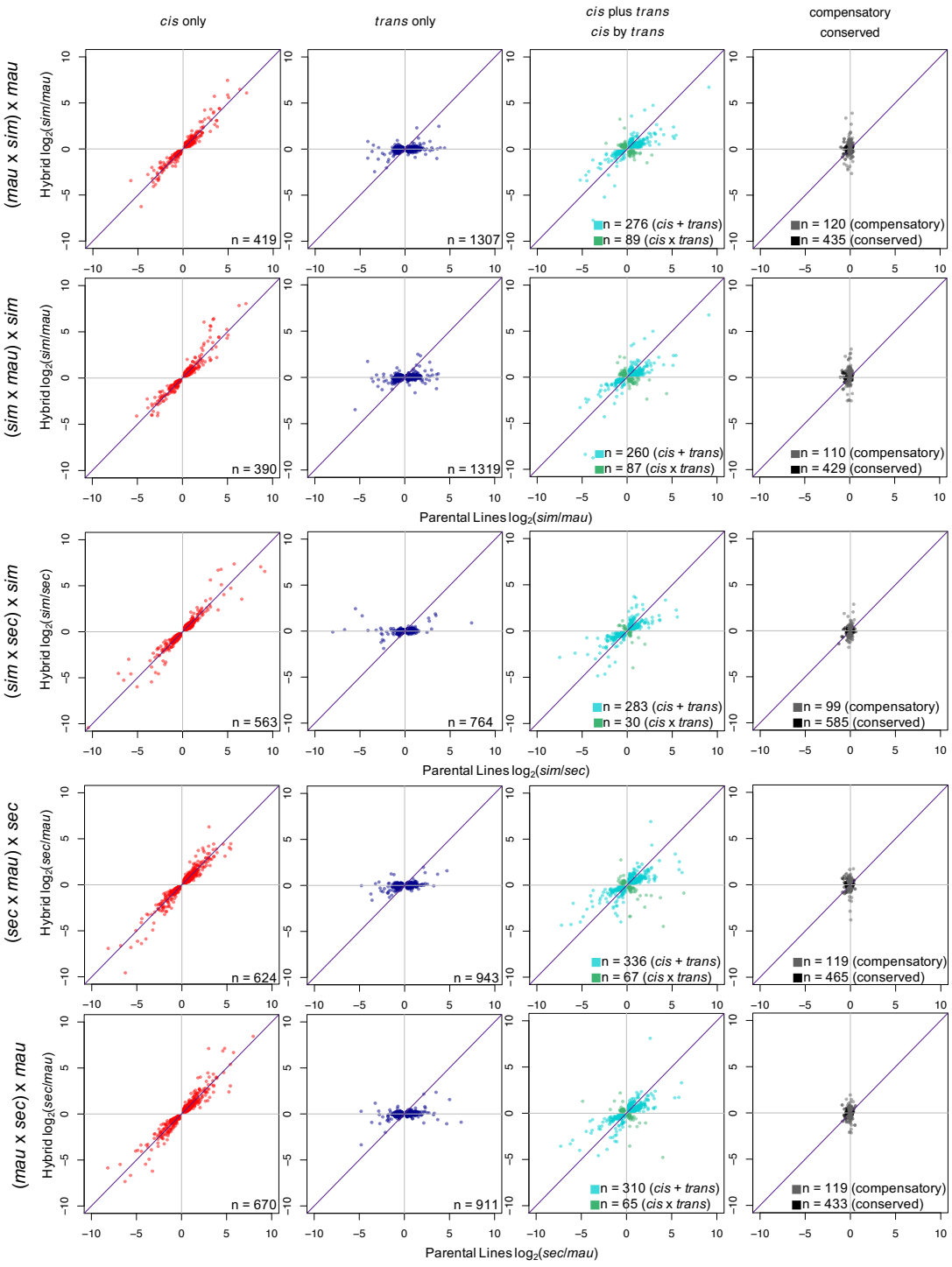


Figure S1.2: Regulatory changes in all pairwise comparisons for maternally deposited transcripts. Transcript level ratios between parental lines and within hybrids at stage 2 describe regulatory changes between species in each set of crosses. The number of genes in each category of regulatory change (n=) is listed in each plot. For definitions of categories of changes and criteria, see Methods.

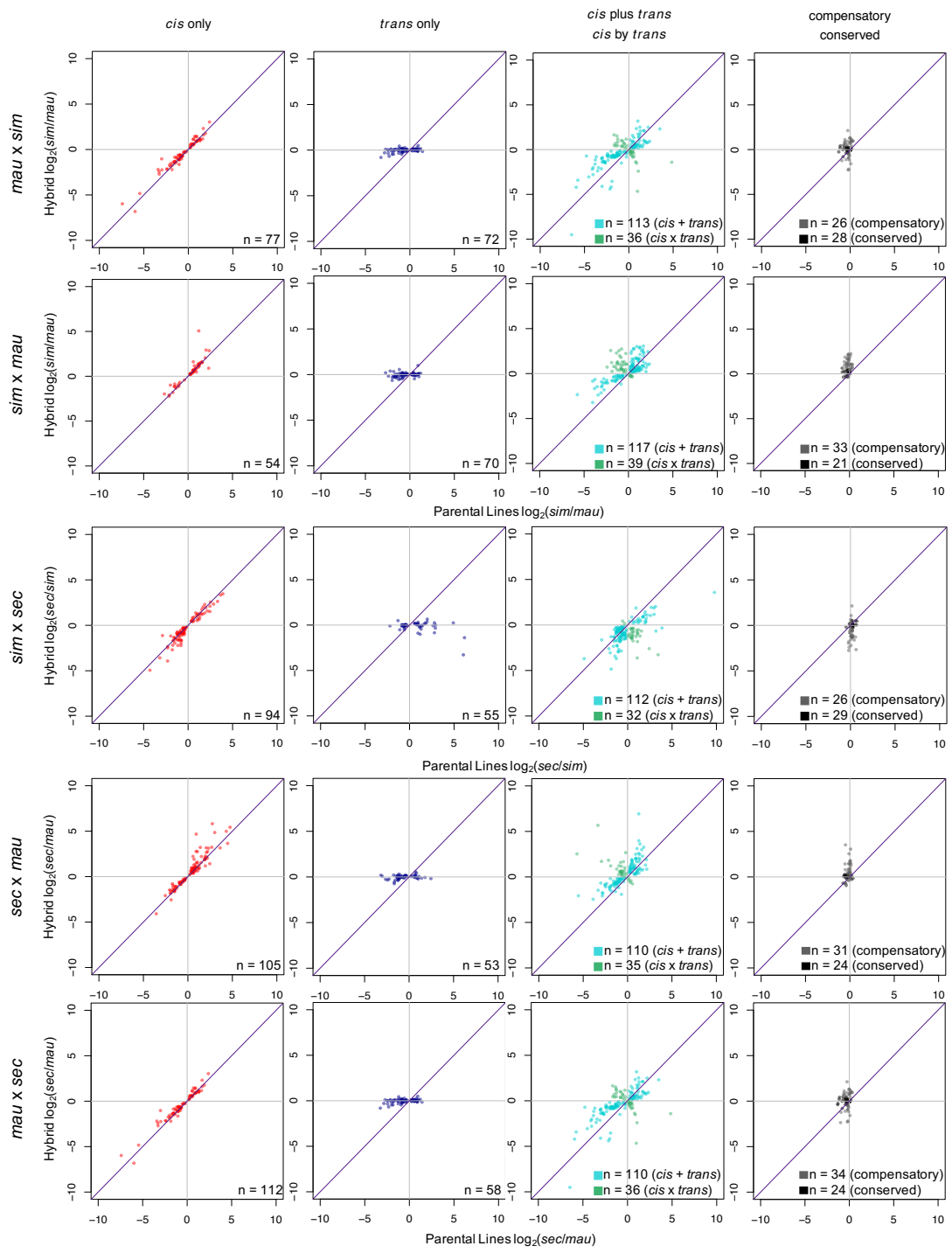


Figure S1.3: Regulatory changes in all pairwise comparisons for mostly zygotic genes. Transcript level ratios between parental lines and within hybrids at for mostly zygotic genes (see Methods) at stage 5 describe regulatory changes between species in each set of crosses. The number of genes in each category of regulatory change (n=) is listed in each plot. For definitions of categories of changes and criteria, see Methods.

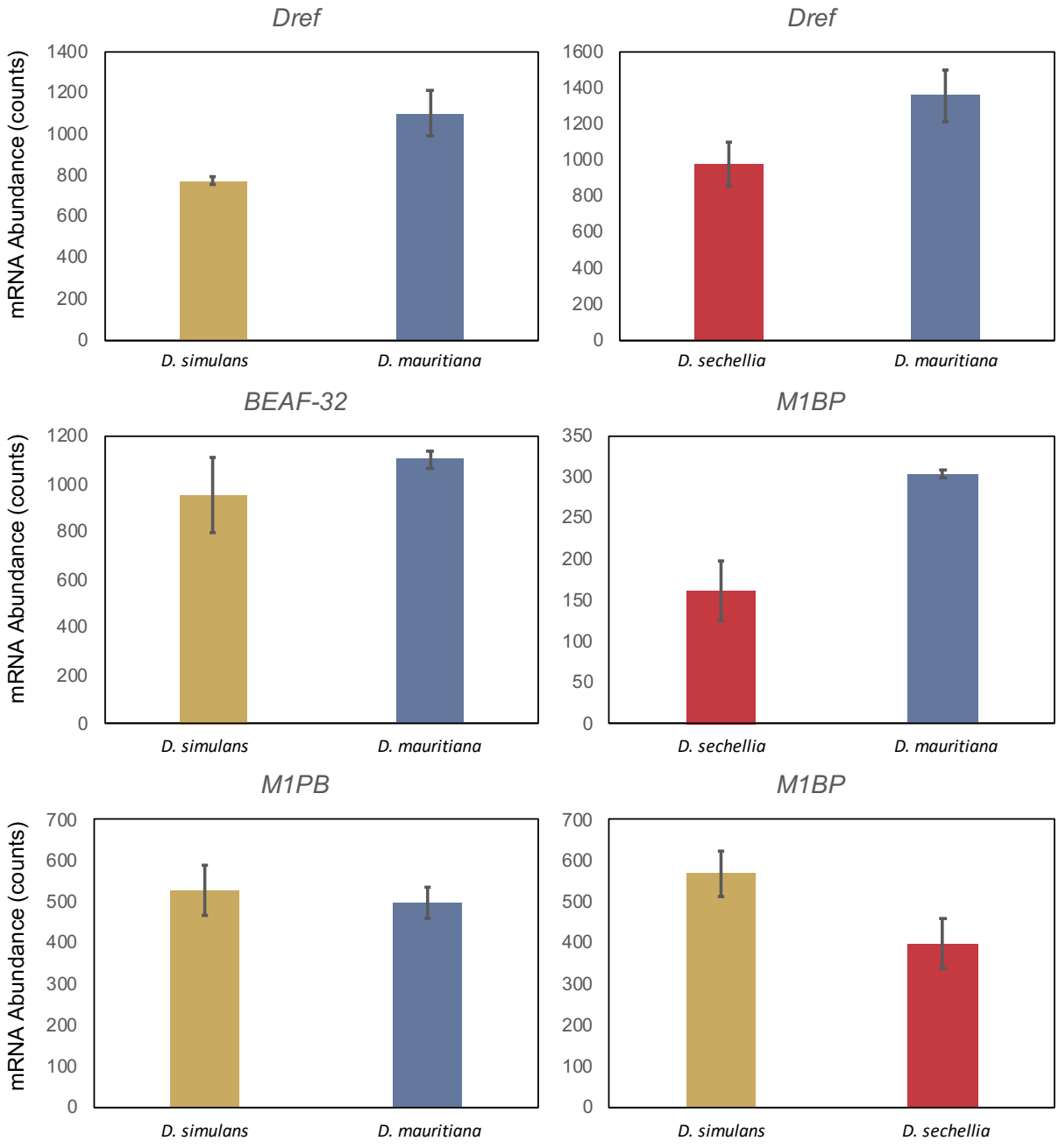


Figure S1.4: Transcript abundance from parental lines at stage 2 demonstrates differential maternal deposition of *M1BP*, *Dref* and *BEAF-32*. Counts for *D. simulans*/*D. mauritiana* and *D. simulans*/*D. sechellia* comparisons are averages across replicates from alignment to the *D. simulans genome*. Counts for *D. sechellia*/*D. mauritiana* comparison are averages across replicates from alignments to the *D. sechellia* genome. Error bars represent standard deviations. Count data for the same species and gene may differ across comparisons due to the genome used for alignment in each comparison and normalization of counts within a comparison.

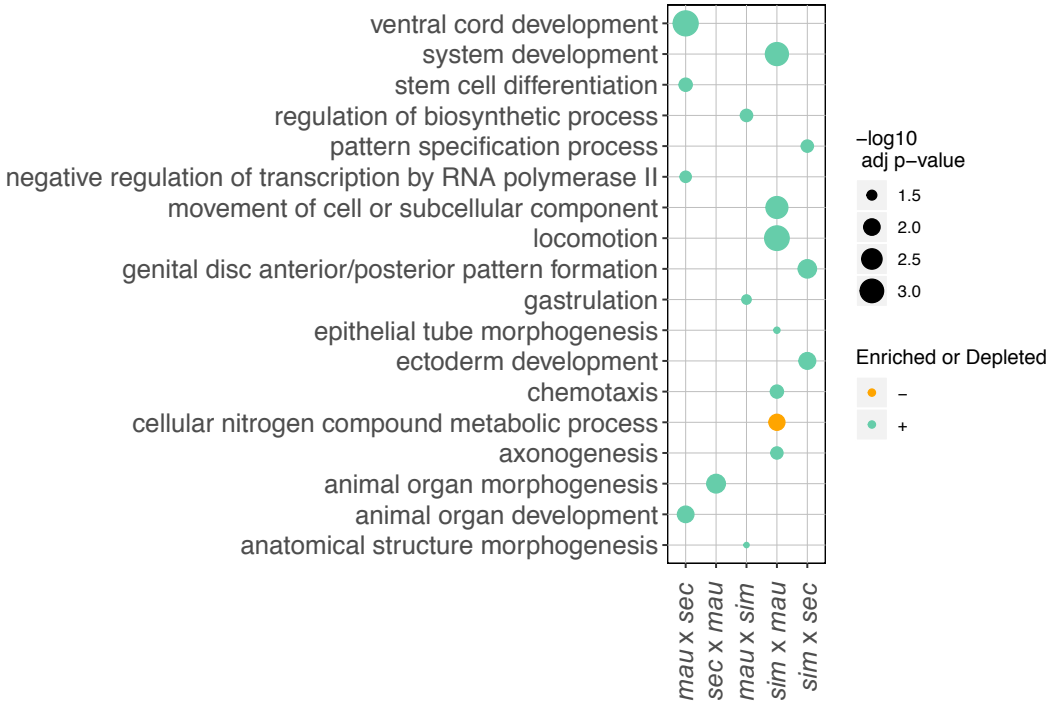


Figure S1.5: Gene ontology (GO) analysis for categories unique to a specific cross show enrichment for specific developmental processes. Significantly enriched GO terms are listed for zygotically transcribed genes that change in *trans* regulation between each pair of species compared. Again, zygotically transcribed genes are limited to those that are mostly zygotic (see Methods), in comparison to the maternal stage of development. Terms are listed for the biological processes category. Gene categories identified uniquely in a single cross primarily represent specific types of developmental processes, and may indicate evolved differences in parental genomes in these processes.

Target Gene	Sequence
L_ORY	aatacaactcaggagcgggacaatttt
R_ORY	tcgtaccatttgcaatccgactag
L_kl3	gaacgcgcatccattttattct
R_kl3	tcgaaaagcccacgacaggtattt
L_ftz	accaaccccgtgaagaagctgaagtaca
R_ftz	cgtgtgtgatgcctacctgatgccaaagt

Table S1.1: Primers for genes *ORY*, *kl3* (both on the Y chromosome) and *ftz* (control locus, on 3R) that were used for genotyping stage 5 embryos as male or female.

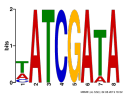
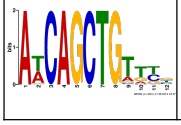
Motif	Predicted Binding Proteins	Enriched in upstream regions of *D. sechellia* genes that are conserved in expression (E-value)	Enriched in upstream regions of *D. sechellia* genes that are changing in expression (E-value)	Enriched in upstream regions of *D. simulans* genes that are conserved in expression (E-value)	Enriched in upstream regions of *D. simulans* genes that are changing in expression (E-value)
	Dref, BEAF-32	(*sec* x *mau*) x *sec* (1.7e-009) (*sim* x *sec*) x *sim* (9.6e-011)	(*mau* x *sec*) x *mau* (2.2e-023) (*sec* x *mau*) x *sec* (3.5e-024) (*sim* x *sec*) x *sim* (2.7e-011)	(*sim* x *sec*) x *sim* (1.7e-006)	(*mau* x *sim*) x *mau* (2.5e-010) (*sim* x *mau*) x *sim* (2.8e-032) (*sim* x *sec*) x *sim* (2.2e-032)
	M1BP	(*mau* x *sec*) x *mau* (2.8e-006) (*sec* x *mau*) x *sec* (3.0e-009) (*sim* x *sec*) x *sim* (1.1e-006)	(*mau* x *sec*) x *mau* (7.7e-008) (*sec* x *mau*) x *sec* (5.5e-028) (*sim* x *sec*) x *sim* (1.8e-015)		(*sim* x *mau*) x *sim* (2.3e-008)
	crp, salivary gland-expressed bHLH, similar to Deadpan, E-box, nau (only in *sec* x *mau* and *sim* x *sec* comparison in *D. sechellia* upstream regions)	(*mau* x *sec*) x *mau* (6.6e-007) (*sec* x *mau*) x *sec* (6.6e-003) (*sim* x *sec*) x *sim* (7.5e-004)			

Table S1.4: Enriched motifs found upstream of maternally deposited genes. Sequences 500bp upstream were extracted for genes in *D. simulans* and *D. sechellia* that change in regulation or that are conserved in each pairwise comparison. Motifs that were significantly enriched in analysis using MEME and HOMER are listed in the table and predicted binding proteins discovered using Tomtom and Homer are also described. E-values generated by MEME indicating the enrichment of each motif compared to background in each cross are also listed. The position weight matrix represented is a representative example of the discovered motifs.

Table S6: Number of genes with each pattern of inheritance in hybrids							
	Additive	*sim* dominant	*sec* dominant	*mau* dominant	Underdominant	Overdominant	Conserved
(*sim* x *mau*) x *sim*	88	735		182	197	270	376
(*mau* x *sim*) x *mau*	97	733		210	168	260	412
(*sec* x *mau*) x *sec*	237		540	292	280	371	322
(*mau* x *sec*) x *mau*	220		489	302	329	382	313
(*sim* x *sec*) x *sim*	180	450	215		141	187	536
sim x *mau*	51	68		20	71	27	17
mau x *sim*	77	57		31	40	40	11
sec x *mau*	70		62	30	39	44	17
mau x *sec*	58		69	35	34	49	19
sim x *sec*	49	59	42		51	24	19

Table S1.6: Number of genes within each category of inheritance pattern in the hybrids. Stage 2 hybrids are shown in the first five rows and stage 5 hybrids are shown in the last five rows of the table.

CHAPTER 2:

Dynamics of Maternal Transcript Degradation Across Species of *Drosophila*

ABSTRACT

Post-transcriptional mechanisms of regulation have been shown to have an especially important role during the earliest stages of development, as transcription is not yet active at the time of fertilization. Here, we investigate how regulation at the level of transcript degradation occurs over the course of early embryogenesis, during a time when the degradation of maternal transcripts is necessary for the proper transfer of developmental control to the zygote. While several mechanisms of maternal transcript degradation are well understood, they do not account for degradation of all maternal transcripts, and little is known about the transcripts that remain after the zygotic genome is active. To get a more complete understanding of the regulation of maternal transcript degradation, we perform RNA-sequencing on embryos from 7 developmental stages, from a time when only maternal transcripts are present to a time when we expect that all should be degraded. We extend this analysis to look at four species of *Drosophila* with vastly different developmental times, evolutionary divergence times, and of different climates and geographic origins. We compare the trajectories of maternal transcript degradation across species and find that while the proportions of maternal-only transcripts that degrade at each developmental stage are similar across species, there are differences in the identities of maternal-only transcripts between species.

INTRODUCTION

In all animals, maternally deposited RNA and proteins drive all aspects of early embryogenesis before the zygotic genome is activated (Tadros and Lipshitz 2009; Vastenhouw et al. 2019). Prior to transcription from the zygotic genome, maternal

transcripts are post-transcriptionally regulated at the levels of translation, localization, and degradation (Hamm and Harrison 2018), which are integral to setting up the body plan in the developing embryo. During the maternal to zygotic transition (MZT) these maternal factors are degraded as the zygotic genome is activated and takes control over the rest of development. The transfer of developmental control from the maternal to the zygotic genome is a highly conserved and regulated process (Tadros and Lipshitz 2009; Vastenhouw et al. 2019). The timely removal of maternally deposited transcripts is important for proper activation of the zygotic genome and progression through development (Tadros et al. 2007; Benoit et al. 2009).

Maternal transcripts that are deposited into the oocyte during oogenesis are targeted for degradation by RNA binding proteins (Tadros et al. 2007; Benoit et al. 2009; Laver et al. 2015b) and miRNAs (Bushati et al. 2008) at specific times during development. In *Drosophila*, a system with excellent resources and a long history of study in the mechanisms of early development, some RNA binding proteins are maternally deposited to direct degradation and translational repression of maternal transcripts in an early wave of degradation (Tadros et al. 2007; Benoit et al. 2009; Laver et al. 2015b). Other factors, including the microRNAs in the miR-309 cluster in *D. melanogaster* (Bushati et al. 2008), are produced by the zygotic genome and direct maternal transcript decay later in development. Some factors, including the RNA binding protein Brain Tumor (Laver et al. 2015a; b), act during both the early and later waves of maternal transcript degradation. These maternal and zygotic waves of degradation represent a regulatory partnership with feedback between the genomes during the handoff of developmental control. In *D. melanogaster*, RNA binding proteins

including Smaug and Brain Tumor and miRNAs in the mir-309 cluster are each known to target a unique subset of maternally deposited transcripts at specific times during development (Laver et al. 2015b). Pumilio, another RNA binding protein, shares some overlapping targets with Brain Tumor but also targets a unique set of transcripts not associated with Brain Tumor (Laver et al. 2015b). While several of these factors that facilitate maternal mRNA degradation are well characterized, it is estimated that they only target a subset of all maternal transcripts deposited into the oocyte (Laver et al. 2015b). Research indicates that additional miRNAs may be involved in maternal transcript clearance (Thomsen et al. 2010) but more work is needed to determine what additional mechanisms are responsible for degradation of the remaining maternal transcripts. Further, maternal transcript degradation has only been characterized up until about the time of gastrulation, at which point the zygotic genome is active but many maternal transcripts still remain (De Renzis et al. 2007; Thomsen et al. 2010; Lott et al. 2014; Vastenhouw et al. 2019). The timing of degradation for transcripts still present after gastrulation, the mechanism of their removal, and whether this group of more stable transcripts play an important role in development is not known. Here, we ask when all maternal transcripts are degraded by examining maternal mRNA levels across 7 developmental timepoints, until no maternal transcripts remain.

Prior studies of maternal transcript degradation in *Drosophila* largely focused on the species *D. melanogaster* (Tadros et al. 2007; De Renzis et al. 2007; Bushati et al. 2008; Benoit et al. 2009; Thomsen et al. 2010; Laver et al. 2015b). The function of several regulators of maternal transcript degradation in this model species have been characterized and the timing of degradation has been studied, although less is known

about the dynamics of decay after the MZT. The process of maternal transcript degradation has not been characterized in other species of *Drosophila* yet we know that there are qualitative and quantitative differences in maternal transcript deposition across the *Drosophila* phylogeny (Atallah and Lott 2018). Transcripts that evolve maternal deposition presumably must also evolve signals to target them for degradation during the MZT. In this study, we examine four species of *Drosophila* with a range of evolutionary divergence times (~5 - 47 million years; Suvorov et al. 2022) and of differing geographic origin (species with cosmopolitan, tropical, alpine, and temperate origins; Kuntz and Eisen 2014). These species also vary over two-fold in the amount of time that each takes to complete embryonic development, from approximately 14 hours in *D. yakuba* to approximately 32 hours in *D. virilis* at 24°C (Markow et al. 2008). *D. melanogaster* and *D. persimilis* fall in the middle of this range with embryonic development times of about 22 hours and about 26 hours, respectively (Markow et al. 2008). We use comparative transcriptomics to investigate whether the maternal mRNAs in species with longer development times, such as *D. virilis*, are stable for a longer period of absolute time than more rapidly developing species, like *D. yakuba*. For each species, we created a transcriptomic time course of 7 stages during embryonic development from stage 2, when all transcripts are maternal in origin, to stage 12, when we expect all maternal transcripts to have degraded. We used morphology to identify developmental stages, in order to obtain comparable staging between species with different developmental times. We chose to use stages based on morphology rather than absolute time after egg laying because developmental timing scales

proportionately with the total time for embryonic development across species of *Drosophila* and across temperatures (Kuntz and Eisen 2014).

Here, we look further into development after the zygotic genome is activated and the pool of transcripts in the embryo is composed of those from the zygotic genome and any remaining maternal transcripts. In order to differentiate reads originating from the maternal or zygotic genomes, we used crosses between genetically distinct lines of the same species. Each line has different single nucleotide polymorphisms (SNPs), when compared to the genome line, that we can use to assign reads uniquely to the alleles originating from the maternal or paternal genome. We find a subset of maternal-only transcripts that do not appear to be zygotically expressed until after their maternal counterparts are degraded. Among these maternal-only transcripts, about one half to one third still persist after widespread zygotic genome activation while over half are degraded by the end of stage 5. There isn't significant enrichment for genes related to a particular biological process or function when comparing transcripts that degrade before and after the zygotic genome is fully activated. This may indicate that degraded transcripts are spread over biological functions. Additionally, only a small proportion of the maternal-only transcripts are maternal-only across all four species while a larger percentage are unique to a specific species. Future analysis will investigate the degradation signals that may be present in these unique maternal-only transcripts and compare them across species. Through this analysis, we will determine how the stability of maternal transcripts compares across species of *Drosophila*, until the last maternal transcripts are degraded.

RESULTS

We looked at four species of *Drosophila* with vastly different developmental times: *D. melanogaster*, *D. persimilis*, *D. virilis*, and *D. yakuba*, to investigate the trajectory of maternal transcript degradation across embryogenesis. These species have more than two-fold variation in development time (Markow et al. 2008), and represent a range of divergence times (~5 - 47 million years; Suvorov et al. 2022). We collected single embryos from 7 developmental timepoints over the course of embryogenesis including stages: 2, 4, the very end of stage 5, 7, a timepoint between stages 8 and 9, 10, and 12 (Bownes' stages; Bownes 1975; Campos-Ortega and Hartenstein 2013). At stage 2, all of the transcripts present are maternal, the end of stage 5 (late stage 5) represents a time after widespread zygotic genome activation, and by stage 10 or 12, we expect all maternal transcripts to be degraded (Lott et al. 2014). Sampling this wide range of timepoints will permit investigation of degradation across the entire period of embryonic development where maternal transcripts are present. In order to differentiate transcripts originating from the maternal versus the zygotic genome, we crossed two lines from each species (see methods), with each line being genetically distinct from the genome reference lines. We used morphological features unique to each timepoint in development to precisely stage each embryo prior to collection (see methods).

Transcriptomes of single embryos cluster by developmental stage

Across species, gene expression is similar at each developmental stage. Clustering analysis reveals that in each species, replicates from the same timepoint

cluster with one another in both principal component analysis (PCA; Figure 2.1) and hierarchical clustering (Figures 2.2 and S2.1). There are only three instances across all samples where a sample clustered more closely with an adjacent stage than with other replicates of the same stage. In the hierarchical clustering for *D. virilis* samples, one of the stage 10 samples falls outside of the cluster of the other two stage 10 replicates and outside of the stage 12 replicates (Figure S2.1). Despite this, the Spearman's rank correlation coefficients from read counts in counts per million (CPM) are still very high across stage 10 replicates and range from 0.97-0.99. Similarly, among the *D. persimilis* samples in hierarchical clustering, one stage 10 sample clusters more closely with the stage 12 replicates than the other stage 10 replicates and one of the stage 8-9 samples clusters most closely with the stage 7 replicates (Figure S2.1). Again, correlation among replicates remains high with Spearman's rank correlation coefficients around 0.99 for the stage 8-9 replicates and between 0.93 and 0.98 for the stage 10 replicates. PCA for the *D. persimilis* samples indicates that stage 8-9 replicates and stage 10 replicates differ more for PC2 than replicates of other stages (Figure 2.1). We decided to include all replicates in the analysis because while not all the samples clustered perfectly among replicates, they still cluster when looking at the PCA and have high Spearman's rank correlation coefficients. Stage 8-9 and stage 10 are longer developmental stages than the earlier stages, so greater variance of samples in *D. virilis* and *D. persimilis*, more slowly developing species, may indicate sampling slightly earlier or later in each stage. The rest of the samples cluster as replicates of the same stage and have higher Spearman's rank correlation coefficients among replicates of the same stage than between samples of different stages. The data from hierarchical clustering, PCA, and

Spearman's rank correlation coefficients indicate that the collection of single embryos at the seven developmental timepoints is highly reproducible across species.

Maternal-only genes differ across species

To investigate patterns of degradation across species, we focused first on maternally deposited transcripts that are not later transcribed by the zygotic genome, which we will call maternal-only genes. Thus far, we classified maternal-only genes as those that are expressed at stage 2, the stage when only maternal transcripts are present, and that do not increase in expression before they are completely degraded at a later stage (Figures 2.3, S2.2, S2.3, and S2.4). Maternally deposited genes are those with 5 or more CPM (counts per million) across all three stage 2 samples. We define degradation as having a significant decrease in expression between stage 2 and a later stage (as determined by edgeR, see Methods) and as having less than 5 CPM across all three samples of a later stage. Genes were filtered out if they had a significant increase in transcript level at any point in time after stage 2, before being completely degraded, as this was likely a result of zygotic transcription (see Methods). Future allele-specific analysis will validate that these are indeed maternal-only, and will take into account when the maternal component of zygotically expressed genes degrades. Of all of the genes that are maternally deposited in any of the four species, between ~11-15% of genes are maternal-only by our definition (see methods). Roughly the same percentage, ~10-13%, of all maternally deposited genes are maternal-only when restricted to genes that have a single ortholog in each species. For all comparisons

across species, we restricted the analysis to genes with one identified ortholog in each species.

In each species, we detect maternally-deposited transcripts being degraded throughout the six stages following stage 2 (Figures 2.3, S2.2, S2.3, and S2.4). The dynamics of maternal-only transcript degradation appear most similar between the closest related species, *D. melanogaster* and *D. yakuba*, where both species have approximately the same proportion of maternal-only genes with degraded transcripts at each developmental stage (Figure 2.4, Tables 2.1 and 2.2). *D. persimilis* and *D. virilis* have fewer genes with transcripts that degrade by stage 4, but by late stage 5 the proportion of maternal-only genes with degraded transcripts approximately matches that of *D. yakuba* and *D. melanogaster* (Table 2.2). Across all species, roughly 66-71% of the maternal-only genes no longer have transcripts present at the end of stage 5 (Table 2.2). For all maternally deposited transcripts, not just those that are maternal-only, prior work estimates a range in the percentage that are degraded by this time in development (De Renzis et al. 2007; Thomsen et al. 2010; Lott et al. 2014; Vastenhouw et al. 2019). In comparison to prior work, the range here, 66-71% of all maternal-only transcripts, falls just outside of those estimates where the highest was about 60%. This may be because the analysis here only focuses on a specific subset of maternal transcripts. Although we expect that all transcripts will be zygotic in origin at stage 12, a relatively small proportion of maternal-only genes have transcripts that remain at this stage in development, ~3-7% across species, and these will be validated with the allele-specific data.

While the relative proportions of genes that are maternal-only remain consistent across species and there are similar proportions of maternal-only transcripts that degrade at each stage across two or more species, there is not a large overlap in the identities of maternal-only genes that are common among all four species. This is consistent with similar findings showing that maternal-only genes have higher transcriptomic divergence across species of *Drosophila* when compared to zygotic-only genes (Atallah and Lott 2018). Only 119 genes, between ~18-23% of the total number of maternal-only genes for that species, are shared across all four species (Figure 2.5). In contrast, ~29-40% of maternal-only genes, between 184-208 genes, are unique to that species (Figure 2.5).

Maternal-only genes that degrade early versus later in development do not show clear evidence of gene ontology enrichment

Gene ontology (GO) analysis was used to investigate whether maternal-only transcripts that degrade earlier or later in development are functionally similar or part of the same biological processes. We defined transcripts that degrade by stages 4 and late 5 as degrading early while those that degrade by stages 7, 8-9, 10 and 12 were classified as degrading late. Early decay transcripts were compared to maternal-only transcripts that degrade later in development to see if transcripts that degrade earlier are functionally similar to one another when compared to those that degrade later in embryogenesis. We also compared transcripts that degrade later during development to those that degrade earlier to investigate whether transcripts that degrade later during

development are functionally similar to one another when compared to transcripts that degrade earlier in development.

For most species, and for most GO categories, there was no significant enrichment for GO terms in any of these comparisons. The only exceptions included comparing the target set of early decay transcripts to the background set of late decaying transcripts in *D. virilis*. Here, there are fewer genes than expected in the biological processes categories of response to abiotic stimuli and response to radiation (Table 2.3). Additionally, when comparing late decay transcripts to the background of early decaying transcripts in *D. persimilis*, there is an enrichment for biological processes terms regionalization and axis specification (Table 2.3). Also in *D. persimilis*, when using the early decay transcript list as the target set and the late decay transcript list as the background, there are fewer genes than expected for the biological processes categories regionalization and axis specification (Table 2.3). The same categories show up as depleted and enriched in these comparisons, depending on the direction of the target and background list. For *D. virilis*, this reciprocal pattern was not seen when comparing the target set of the late decay transcripts to the background of the early decay transcripts. This is likely due to reduced significance of marginal cases, resulting from reduced power when the target set is smaller relative to the background in this direction of the comparison. In this direction, the comparison did not result in any significant categories of enrichment or depletion. The overall lack of enrichment categories across species indicates that maternal-only transcripts that degrade at different times during development may be spread across biological functions. There

may also be a lack of significant enrichment because the list of maternal-only genes is relatively small and there may not be enough statistical power.

Comparisons were also made between the early degrading set of maternal-only transcripts or the late degrading set of maternal-only transcripts to a background list that included all maternally deposited transcripts except for those in the target set. The background thus reflects all maternal genes with a zygotic component as well as those maternal-only genes that degrade at a different time. There is enrichment for categories when comparing the transcripts that degrade by late stage 5 to a background of all maternally deposited transcripts, with the exclusion of the target set in the background. Categories related to anion or ion transport and transporter activity showed enrichment among early degraded transcripts across all species. Significant GO categories for all species that fall under Molecular Function are listed in Table 2.4. Additionally, a representative example of what we find across species is shown in Table 2.5 which lists all significant GO categories for Biological Processes in *D. melanogaster*. There were very few categories that showed enrichment of Biological Processes, Molecular Functions, or Cellular Components among transcripts that were degraded early in development. Most significant categories had fewer transcripts than expected and were depleted in the target set (Table 2.5). As maternal-only transcripts are a small proportion of all maternally deposited transcripts, depletion may simply represent the types of genes that are likely to also have zygotic transcription. There is a similar trend for transcripts that degrade later in development although there are fewer significant categories, overall. Most of the significant GO categories are indicative of fewer transcripts than expected in the group of transcripts degrading at stages 7, 8-9, 10, and

12, relative to the background list. The general lack of cohesive signal for categories of maternal-only genes being degraded early or late supports that these maternal-only transcripts degraded early in development may be spread over many biological functions and are not enriched for any specific biological process or function, or that we lack the statistical power to detect enrichment with the relatively small number of maternal-only genes.

DISCUSSION

The handoff of developmental control from the maternal to the zygotic genome is a critical time during the early embryonic development of all animals (Tadros and Lipshitz 2009; Vastenhouw et al. 2019). Coordination between the two genomes involves the decay of maternal transcripts as the zygotic genome is activated. Clearance of maternal transcripts is critical, and allows for the spatial patterning of zygotic gene expression in specific domains (De Renzis et al. 2007) and for the expression of zygotic-specific isoforms (Atallah and Lott 2018). While the degradation of maternal transcripts is integral to the MZT, it has only been investigated in *D. melanogaster* up until about the time of gastrulation. Prior research indicates that many maternal transcripts still remain at this time but how and when they degrade is not known. Here, we confirm that many maternal transcripts likely still persist in the embryo after gastrulation, across species. In this part of the analysis we investigated maternal-only genes that are not likely to be re-expressed by the zygotic genome until after the maternal counterparts are degraded. Future studies using allele-specific analysis will be done using line-specific SNPs called by GATK (see methods). Allele-specific analysis

will allow us to confirm that these transcripts originate from only the maternal genome as well as to investigate the trajectory of degradation for maternal transcripts that are also zygotically expressed.

From this initial analysis of maternal-only transcripts, we find that while many are shared between species, there is a large set of maternal-only transcripts that are unique to each species. This is consistent with previous studies that found higher transcriptomic differences for maternal-only genes, across species (Atallah and Lott 2018). It will be interesting to compare genes that may be maternal-only in one species but are maternally deposited and not fully degraded before being zygotically transcribed in other species to compare whether they have similar timing of degradation across species.

We find that across all species, maternal-only transcripts that decay earlier in development do not appear enriched for any specific biological process or function when compared to transcripts that degrade later in development. This may be due to genes being distributed over biological processes or due to not having enough genes in each category to have statistical power to run GO analysis. By incorporating maternal genes that are zygotically expressed in the GO analysis, the number of genes will increase and there may be enough power to run this analysis.

In addition to looking at the identity of maternal-only transcripts across species, we also find that maternal-only transcripts have largely similar proportions that degrade at each developmental stage in different species. *D. melanogaster* and *D. yakuba*, which have the most recent divergence time, have very similar proportions of maternal-only transcripts that degrade at each developmental stage. Interestingly, while the

proportion of maternal-only transcripts that degrade by stage 4 differs across species, ranging from ~2-20%, by late stage 5, ~66-71% of all maternal-only transcripts have degraded in all species. By the time that the zygotic genome is fully activated, roughly the same proportion of maternal transcripts have degraded in each species even though it preliminarily appears that fewer transcripts degrade by stage 4 in *D. virilis* and *D. persimilis*. These two species have a longer developmental time so future analysis looking at the trajectories of all genes, not only the ones that are maternal-only, will give a more complete picture of whether there are fewer transcripts that degrade early in these species that have slower development. It will also be interesting to investigate whether there are signatures of degradation that differ in the pool of transcripts degrading before stage 4 versus stage 5 and whether these signals differ across species.

Future work will look for enrichment of motifs that may potentially act as binding sites for miRNAs or RNA binding proteins in the UTRs of maternally deposited transcripts. By investigating this over developmental time, we can determine whether transcripts that degrade earlier or later in development have similar or different signatures that may target them for degradation. Further, we can compare enriched motifs across species to investigate whether similar sequences persist in maternal transcripts across evolutionary time.

MATERIALS AND METHODS

Species' stocks and sample acquisition

Single embryos from 7 different developmental stages were collected from crosses between two lines of each of the following species: *D. melanogaster*, *D. persimilis*, *D. virilis*, and *D. yakuba*. The lines that were chosen for each species were different from the genome reference lines used in the 12 Genomes study (Clark et al. 2007). We inbred lines of *D. persimilis*, *D. virilis*, and *D. yakuba* when we received them from the National *Drosophila* Species Stock Center by mating virgin females to their brothers for eight generations.

For embryo collection, crosses were established using two lines from each species such that the maternal and paternal lines were genetically distinct from each other and from the genome line of each species. *D. melanogaster* DGRP-307 females were crossed to *D. melanogaster* DGRP-357 males. *D. melanogaster* lines were part of the *Drosophila* Genetic Reference Panel (Huang et al. 2014). Lines for other species, *D. persimilis*, *D. virilis*, and *D. yakuba*, were obtained from the National *Drosophila* Species Stock Center. *D. persimilis* 14011-0111.50 females were crossed to *D. persimilis* 14011-0111.35 males. *D. virilis* 15010-1051.118 females were crossed to *D. virilis* 15010-1051.85 males. *D. yakuba* 14021-0261.54 females were crossed to *D. yakuba* 14021-0261.40 males. All stocks and crosses were kept in vials on standard cornmeal media at 25°C. Crosses were established using 4-5 virgin females of one line and 4-5 males from the other line of the same species. Eggs were collected from 4-14 day old females after flies were transferred to new vials and allowed to lay for lengths of time ranging from ~4 hours to overnight, depending on what developmental stage was collected. Overnight laying was used to collect later staged embryos. Collected eggs were washed with water and dechorionated using 50% bleach. After dechorionation, the

eggs were again rinsed with water and then quickly transferred into a small pool of halocarbon oil on a microscope slide. Embryos were staged at the correct developmental stages using a microscope (Zeiss Axio Imager M2). Staging was done based on morphological features of the embryos. Stage 2 was identified when the cytoplasm retracts from the vitelline membrane at the anterior and posterior poles of the embryo but before the pole cells have migrated to the posterior pole. Stage 4 is marked by pole cells migrating to the posterior and nuclei moving to the periphery of the embryo, but cellularization has not yet begun. Late stage 5 embryos were identified as a time when cellularization is complete but gastrulation has not started. Stage 7 embryos mark the time when gastrulation is complete and the pole cells become invaginated. The next timepoint sampled was between stages 8 and 9, during germband elongation. Stage 10 was marked by stomodeal invagination and stage 12 is when germband shortening occurs and there is clear segmentation (Ashburner et al. 1989; Campos-Ortega and Hartenstein 2013).

Once staged, embryos were quickly transferred to a piece of parafilm (Bemis) with a paintbrush, rolled to remove excess halocarbon oil, and gently placed into a drop of TRIzol (Ambion) before rupturing with a needle. Once the contents of the embryo dissolved in the TRIzol, the sample was collected into a 1.5mL microcentrifuge tube and frozen at -80°C until extraction. RNA extractions for each sample were carried out from a volume of 1mL TRIzol using glycogen as a carrier (following manufacturer instructions). RNA was quantified using a Qubit 2.0 Fluorometer (Invitrogen) and approximately 80-120 ng of RNA was extracted from each sample. RNA quality was validated using an Agilent Bioanalyzer. TURBO DNA-free (Invitrogen) was used on

each sample, as per manufacturer instructions, following RNA extraction to remove any DNA.

Genotyping was conducted for samples at all developmental timepoints after stage 2 (for stages 4 through 12) to determine the sex of the embryo. We did not genotype stage 2 embryos as the zygotic genome is not yet active at this developmental stage (Ali-Murthy et al. 2013; Kwasnieski et al. 2019) and thus the transcriptome is entirely maternally derived. From the genotyping, three female samples from each developmental stage after stage 2 were selected for sequencing, along with three stage 2 samples. Female samples were selected because it allows for allele-specific data to distinguish maternal from zygotic expression for genes on the X chromosome. For genotyping, DNA was extracted along with the RNA and then amplified using a whole genome amplification kit (Illustra GenomePhi v2, GE Healthcare). We designed sex specific primers for genes on the Y chromosome for each species to identify male embryos (Table 2.6). Two sets of primers were used to identify the genes on the Y chromosome and one set of primers was used to identify an autosomal gene. Some primers worked to identify genes in multiple species. One set of primers was used for *kl-2* and another was used for *WDY* in both *D. melanogaster* and *D. yakuba* on the Y chromosome. A single set of primers was also used for *ftz* for *D. melanogaster* and *D. yakuba*. For *D. virilis*, primers were taken from Paris, et al., 2015 for *kl-2* and *ORY* on the Y chromosome and for *Adh* as a control (Paris et al. 2015). *D. persimilis* primers that were originally designed for *Adh* in *D. pseudoobscura* by Paris, et al., 2015 (Paris et al. 2015) were used as a control while Y specific primers for *CG12218Y-Ψ* (also originally designed for *D. pseudoobscura*) were used from Carvalho and Clark, 2005

(Carvalho and Clark 2005). Only one set of Y-specific primers was used for *D. persimilis*.

Library Preparation and Sequencing

Libraries were both prepared and sequenced at the DNA Technologies and Expression Analysis Cores at the UC Davis Genome Center. mRNA-seq libraries were generated using poly-A enrichment and 150bp paired-end sequencing was done on one lane of the NovaSeq.

RNAseq Alignment

Reads were trimmed using Cutadapt (version 1.7.1; Martin 2011) and then aligned to the corresponding species' reference genome using HISAT2 (version 2.2.1; Kim et al. 2019). For alignment and subsequent analysis, the following reference genomes and corresponding annotation files from FlyBase were used: dmel-all-chromosome-r6.36.fasta, dmel-all-r6.36.gtf, dper-all-chromosome-r1.3.fasta, dper-all-r1.3.gtf, dvir-all-chromosome-r1.07.fasta, dvir-all-r1.07.gtf, dyak-all-chromosome-r1.05.fasta, dyak-all-r1.05.gtf. Alignment sam files were converted to bam files and then sorted and indexed using Samtools (version 1.11; Li et al. 2009).

Ortholog table for comparison across species

An ortholog table was used to compare genes across species. The table was generated using the "dmel_orthologs_in_drosophila_species_fb_2021_01.tsv" table from FlyBase and consists of genes in *D. melanogaster* and the corresponding

orthologs in *D. persimilis*, *D. virilis*, and *D. yakuba*. If there was not an ortholog listed or if multiple orthologs existed for the *D. melanogaster* gene, nothing was listed for that species. For this analysis, we restricted comparisons across species to those genes that had orthologs in all four species.

Variant Calling

GATK (version 4.2.0; Van der Auwera and O'Connor 2020) and Picard (version 2.25.5; Picard Toolkit 2019) tools were used to call variants using the RNA-seq reads aligned to the reference genome for each species. The following programs were used: AddorReplaceReadGroups, MarkDuplicates, SplitNCigarReads, and HaplotypeCaller in GVCF mode. CombineGVCFs was used to combine GVCFs from different samples of the same developmental stage and indels were removed by using SelectVariants. The single file was then run through GenotypeGVCFs for joint genotyping with the following filters: QD < 2.0, FS > 30.0, and DP < 5.0. While the method for GATK joint genotyping is relatively new, Brouard, et al. validated using the joint calling method with RNAseq reads as an effective tool for genotyping (Brouard et al. 2019).

Differential Expression Analysis

Differential expression analysis was performed in R (version 4.1.1; R Core Team 2021). featureCounts (Liao et al. 2014), part of the Rsubread package (version 2.6.4), was used to count the reads aligning to each gene with the parameter countMultiMappingReads set to FALSE. edgeR (version 3.34.1; Robinson et al. 2010) was then used to perform differential expression analysis using the generalized linear

models and quasi-likelihood F-test functionalities with an FDR < 0.05. For any given stage, a cutoff of 5 or more CPM (counts per million) across all three replicates was used as a threshold for genes categorized as expressed while genes with a CPM of less than 5 were considered not expressed in this analysis. A threshold of less than 5 CPM across all three samples was required for a gene to be categorized as degraded by any stage (later than stage 2) in this analysis. A threshold of 5 CPM was chosen based on the distribution of counts in the maternal samples. Additional thresholds between 1 and 10 CPM were tested and did not change the examined results. Maternal-only genes were classified as being expressed at stage 2 and completely degraded at a subsequent stage, without increasing in expression at any timepoint in between. Differentially expressed genes with a significant increase in expression before being fully degraded were removed from the “maternal-only” category as an increase in expression can be attributed to transcription from the zygote. Significant increases in expression were determined in edgeR. Without allele-specific data, at this point in the analysis, we cannot determine what proportion of the transcript level is due to zygotic transcription versus remaining maternal transcripts for genes that increase in expression levels. Here, maternal-only genes were required to have a significant decrease in expression between stage 2 and a later stage and to not have a significant increase in expression before they degraded, as determined by being differentially expressed in edgeR.

Correlation analysis, clustering, and PCA

Correlation analysis, clustering, and PCA were also performed in R (version 4.1.1; R Core Team 2021). Spearman's rank correlation coefficients were calculated for all pairwise comparisons between samples, within a species, using the *corrr* package in R. PCA was also done in R, using the *prcomp* function. Hierarchical clustering was done using the *hclust* function.

Gene Ontology

Gene Ontology (GO) analysis was performed using Panther (Mi et al. 2020). GO Biological Processes complete, GO Molecular Function complete, and GO Cellular Component complete were analyzed using the statistical overrepresentation test. The Fisher's exact test type and the Bonferroni correction for multiple testing were used for GO analysis. Analysis was run for each species using the *D. melanogaster* orthologous genes. Analysis was first run two ways for each species, once using a target list made up of transcripts that degrade by stages 4 and late 5 and a background list of transcripts that degrade by stages 7, 8-9, 10 and 12. In another set of analysis, the target list was composed of transcripts that degrade by stages 7, 8-9, 10 and 12 while the background list consisted of transcripts that degrade by stages 4 and late 5. Analysis was also run using either transcripts that degrade by stages 4 and late 5 or that degrade by stages 7, 8-9, 10, and 12 as the target sets. The background lists were composed of all maternally deposited transcripts but with the target set removed from them.

REFERENCES

Ali-Murthy Z., S. E. Lott, M. B. Eisen, and T. B. Kornberg, 2013 An essential role for zygotic expression in the pre-cellular Drosophila embryo. PLoS genetics 9: e1003428. https://doi.org/10.1371/journal.pgen.1003428

Ashburner M., K. G. Golic, and R. S. Hawley, 1989 *Drosophila: A Laboratory Handbook*. Cold Spring Harbor Laboratory Press, 2005.

Atallah J., and S. E. Lott, 2018 Evolution of maternal and zygotic mRNA complements in the early Drosophila embryo, (K. A. Dyer, Ed.). PLoS genetics 14: e1007838. https://doi.org/10.1371/journal.pgen.1007838

Benoit B., C. H. He, F. Zhang, S. M. Votruba, W. Tadros, *et al.*, 2009 An essential role for the RNA-binding protein Smaug during the Drosophila maternal-to-zygotic transition. Development 136: 923–932. https://doi.org/10.1242/dev.031815

Bownes M., 1975 A photographic study of development in the living embryo of Drosophila melanogaster. Development 33: 789–801.

Brouard J.-S., F. Schenkel, A. Marete, and N. Bissonnette, 2019 The GATK joint genotyping workflow is appropriate for calling variants in RNA-seq experiments. J Anim Sci Biotechno 10: 44. https://doi.org/10.1186/s40104-019-0359-0

Bushati N., A. Stark, J. Brennecke, and S. M. Cohen, 2008 Temporal reciprocity of miRNAs and their targets during the maternal-to-zygotic transition in Drosophila. Current Biology 18: 501–506. https://doi.org/10.1016/j.cub.2008.02.081

Campos-Ortega J. A., and V. Hartenstein, 2013 *The Embryonic Development of Drosophila melanogaster*. Springer Science & Business Media.

Carvalho A. B., and A. G. Clark, 2005 Y Chromosome of D. pseudoobscura Is Not Homologous to the Ancestral Drosophila Y. Science 307: 108–110. https://doi.org/10.1126/science.1101675

Clark A. G., M. B. Eisen, D. R. Smith, C. M. Bergman, B. Oliver, et al., 2007 Evolution of genes and genomes on the Drosophila phylogeny. Nature 450: 203–218. https://doi.org/10.1038/nature06341

De Renzis S., O. Elemento, S. Tavazoie, and E. F. Wieschaus, 2007 Unmasking Activation of the Zygotic Genome Using Chromosomal Deletions in the Drosophila Embryo. PLOS Biology 5: e117. https://doi.org/10.1371/journal.pbio.0050117

Hamm D. C., and M. M. Harrison, 2018 Regulatory principles governing the maternal-to-zygotic transition: insights from Drosophila melanogaster. Royal Society Open Biology. https://doi.org/10.1098/rsob.180183

Huang W., A. Massouras, Y. Inoue, J. Peiffer, M. Ràmia, et al., 2014 Natural variation in genome architecture among 205 Drosophila melanogaster Genetic Reference Panel lines. Genome Research 24: 1193–1208. https://doi.org/10.1101/gr.171546.113

Kim D., J. M. Paggi, C. Park, C. Bennett, and S. L. Salzberg, 2019 Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol 37: 907–915. https://doi.org/10.1038/s41587-019-0201-4

Kuntz S. G., and M. B. Eisen, 2014 Drosophila Embryogenesis Scales Uniformly across Temperature in Developmentally Diverse Species. PLoS genetics 10: e1004293. https://doi.org/10.1371/journal.pgen.1004293

Kwasnieski J. C., T. L. Orr-Weaver, and D. P. Bartel, 2019 Early genome activation in Drosophila is extensive with an initial tendency for aborted transcripts and retained introns. Genome Res 29: 1188–1197. https://doi.org/10.1101/gr.242164.118

Laver J. D., A. J. Marsolais, C. A. Smibert, and H. D. Lipshitz, 2015a *Regulation and Function of Maternal Gene Products During the Maternal-to-Zygotic Transition in Drosophila*. Elsevier Inc.

Laver J. D., X. Li, D. Ray, K. B. Cook, N. A. Hahn, et al., 2015b Brain tumor is a sequence-specific RNA-binding protein that directs maternal mRNA clearance during the Drosophila maternal-to-zygotic transition. Genome biology 16: 94. https://doi.org/10.1186/s13059-015-0659-4

Li H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, et al., 2009 The Sequence Alignment/Map format and SAMtools. Bioinformatics 25: 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Liao Y., G. K. Smyth, and W. Shi, 2014 featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 30: 923–930. https://doi.org/10.1093/bioinformatics/btt656

Lott S. E., J. E. Villalta, Q. Zhou, D. Bachtrog, and M. B. Eisen, 2014 Sex-Specific Embryonic Gene Expression in Species with Newly Evolved Sex Chromosomes, (H. S. Malik, Ed.). PLoS Genetics 10: e1004159. https://doi.org/10.1371/journal.pgen.1004159

Markow T. A., S. Beall, and L. M. Matzkin, 2008 Egg size, embryonic development time

and ovoviviparity in Drosophila species. J Evolution Biol 22: 430–434. https://doi.org/10.1111/j.1420-9101.2008.01649.x

Martin M., 2011 Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal 17: 10–12. https://doi.org/10.14806/ej.17.1.200

Mi H., D. Ebert, A. Muruganujan, C. Mills, L.-P. Albou, et al., 2020 PANTHER version 16: a revised family classification, tree-based classification tool, enhancer regions and extensive API. Nucleic Acids Res 49: gkaa1106-. https://doi.org/10.1093/nar/gkaa1106

Paris M., J. E. Villalta, M. B. Eisen, and S. E. Lott, 2015 Sex Bias and Maternal Contribution to Gene Expression Divergence in Drosophila Blastoderm Embryos., (A. G. Clark, Ed.). PLoS genetics 11: e1005592. https://doi.org/10.1371/journal.pgen.1005592

“Picard Toolkit.” 2019. Broad Institute, GitHub Repository. https://broadinstitute.github.io/picard/; Broad Institute

R Core Team, 2021 *R: A language and environment for statistical computing*.

Robinson M. D., D. J. McCarthy, and G. K. Smyth, 2010 edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26: 139–140. https://doi.org/10.1093/bioinformatics/btp616

Suvorov A., B. Y. Kim, J. Wang, E. E. Armstrong, D. Peede, et al., 2022 Widespread introgression across a phylogeny of 155 Drosophila genomes. Curr Biol 32: 111-123.e5. https://doi.org/10.1016/j.cub.2021.10.052

Tadros W., A. L. Goldman, T. Babak, F. Menzies, L. Vardy, et al., 2007 SMAUG Is a Major Regulator of Maternal mRNA Destabilization in Drosophila and Its Translation Is Activated by the PAN GU Kinase. Developmental Cell 12: 143–155. https://doi.org/10.1016/j.devcel.2006.10.005

Tadros W., and H. D. Lipshitz, 2009 The maternal-to-zygotic transition: a play in two acts. Development 136: 3033–3042. https://doi.org/10.1242/dev.033183

Thomsen S., S. Anders, S. C. Janga, W. Huber, and C. R. Alonso, 2010 Genome-wide analysis of mRNA decay patterns during early Drosophila development. Genome biology 11: R93. https://doi.org/10.1186/gb-2010-11-9-r93

Van der Auwera G. A. V. der, and B. D. O’Connor, 2020 *Genomics in the Cloud*. O’Reilly Media, Inc.

Vastenhouw N. L., W. X. Cao, and H. D. Lipshitz, 2019 The maternal-to-zygotic transition revisited. Development 146: dev161471-20. https://doi.org/10.1242/dev.161471

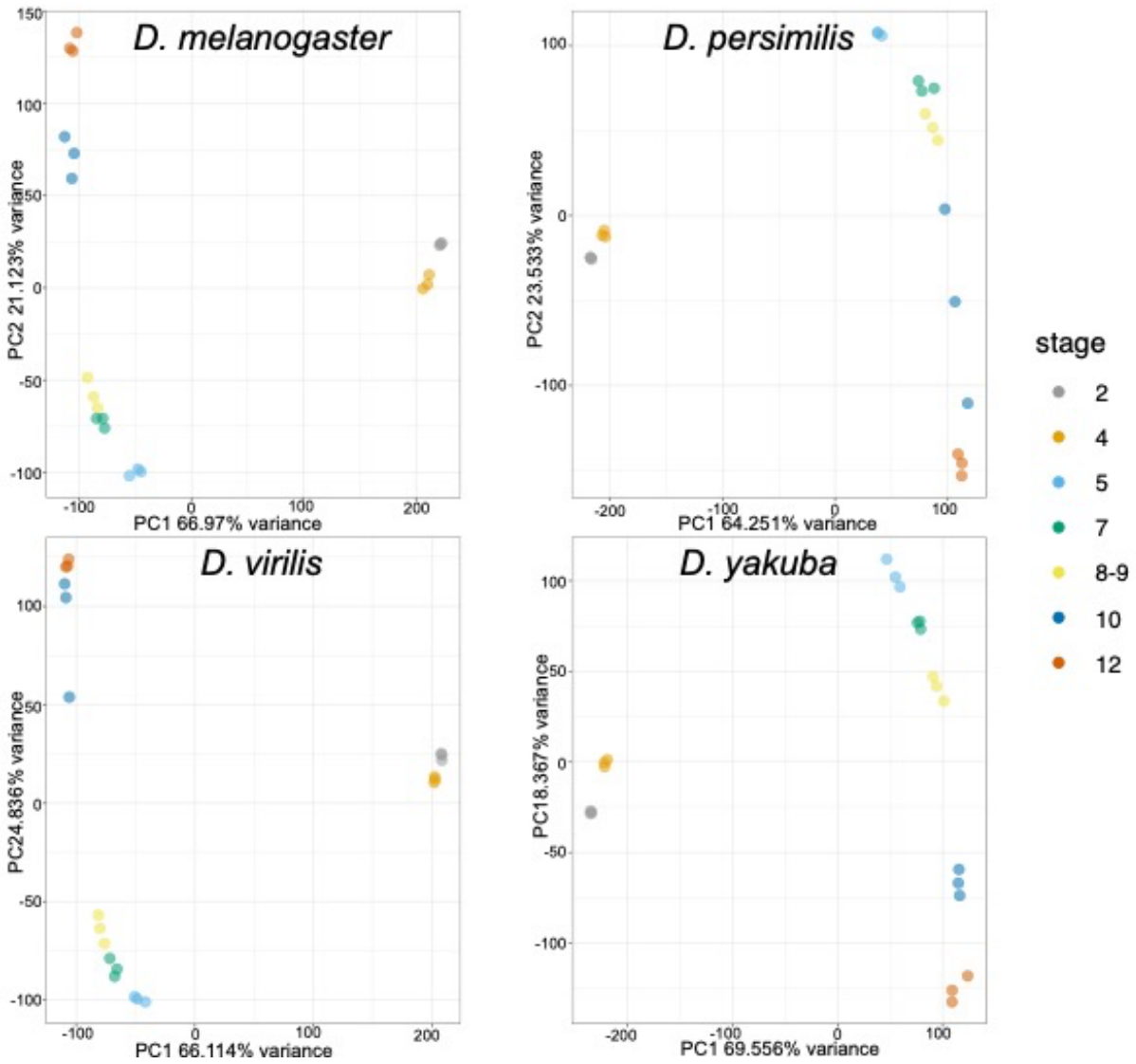


Figure 2.1: Samples group by developmental stage in PCA. For each species, replicates of the same stage cluster more closely with one another than with samples of other stages with few exceptions. Stage 10 *D. virilis* and *D. persimilis* samples separate on PC2. Samples of similar stages, those close in chronology, also cluster more closely with each other than with more distant stages. Stages 2 and 4 cluster more closely along PC1 and separate from the later developmental stages. Then, along PC2, stages late 5, 7, and 8-9 cluster more closely and stage 10 and 12 cluster with one another.

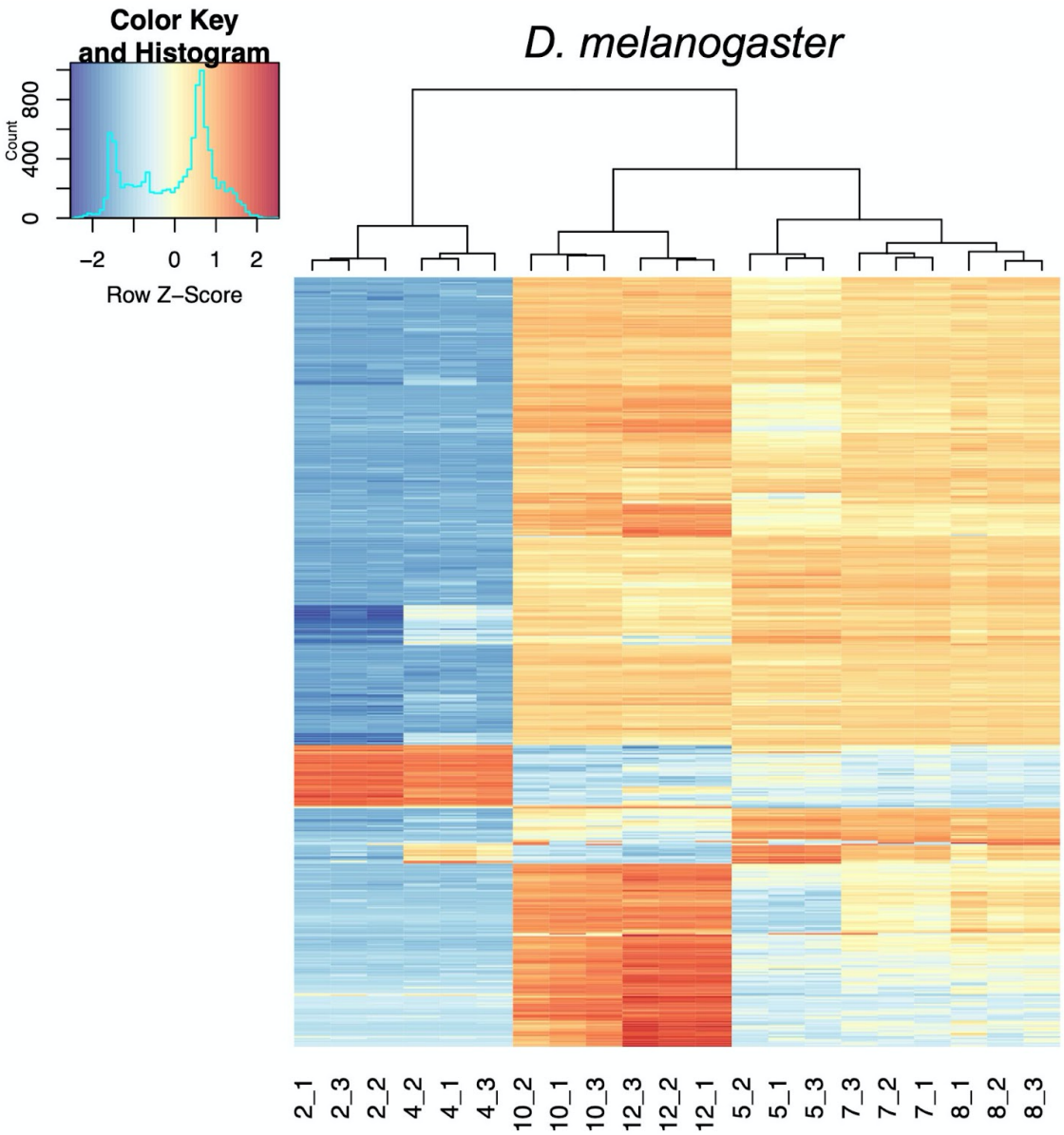


Figure 2.2: Samples group by developmental stage in hierarchical clustering analysis. Replicates of the same stage cluster in hierarchical clustering of *D. melanogaster* samples using the top 500 most variable transcripts. (See Figure S2.1 for hierarchical clustering of *D. persimilis*, *D. virilis*, and *D. yakuba* samples.) Samples are shown at the bottom of the heatmap with the first number representing the stage and the second representing the replicate number. Secondary to samples clustering by stage, samples closer in developmental time also cluster more closely with one another.

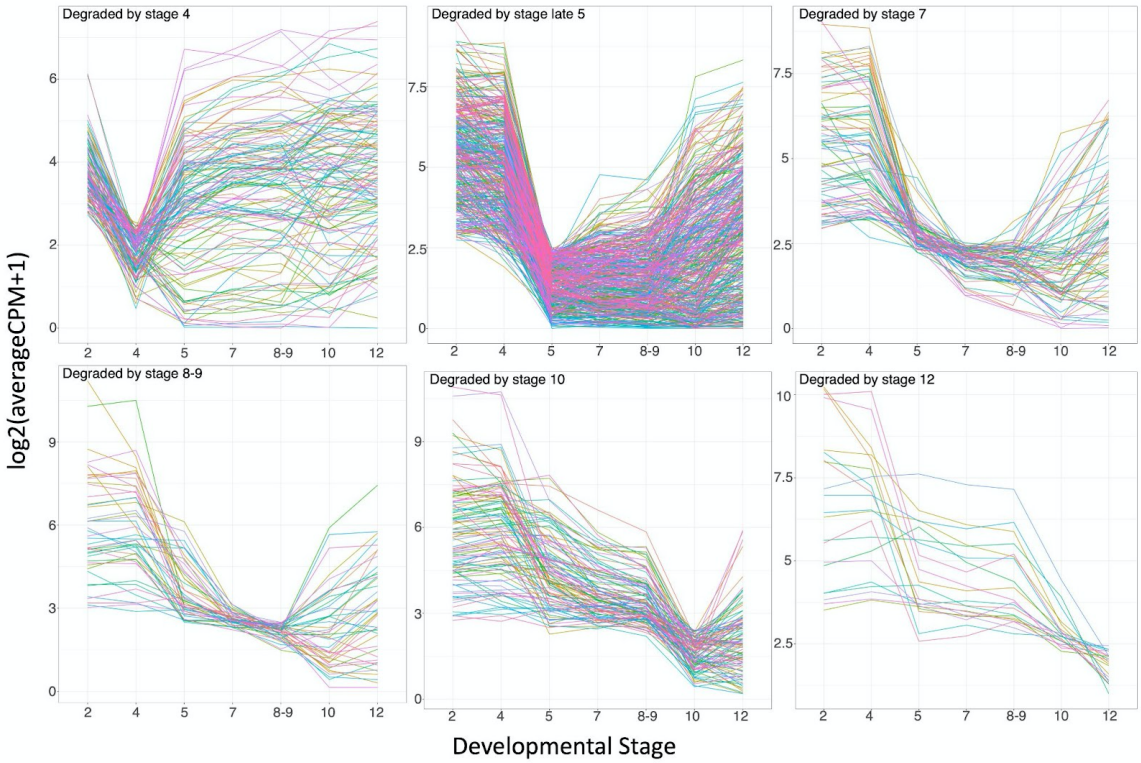


Figure 2.3: Transcript levels of maternal-only genes, grouped by when they degrade. Transcripts for maternal-only genes in *D. melanogaster* degrade across all stages sampled. (See figures S2.2, S2.3, and S2.4 for transcripts in *D. persimilis*, *D. virilis*, and *D. yakuba*). Maternal-only genes are those with transcripts that are maternally deposited and are not transcribed by the zygotic genome or are fully degraded before being expressed by the zygote. Each graph depicts the group of transcripts that are maternally deposited and degraded by a specific developmental stage, as determined by having less than 5 CPM across all replicates. The largest group of maternal-only transcripts degrades between stage 4 and stage late 5. Fewer transcripts degrade at later stages but maternal transcripts still persist until stage 12.

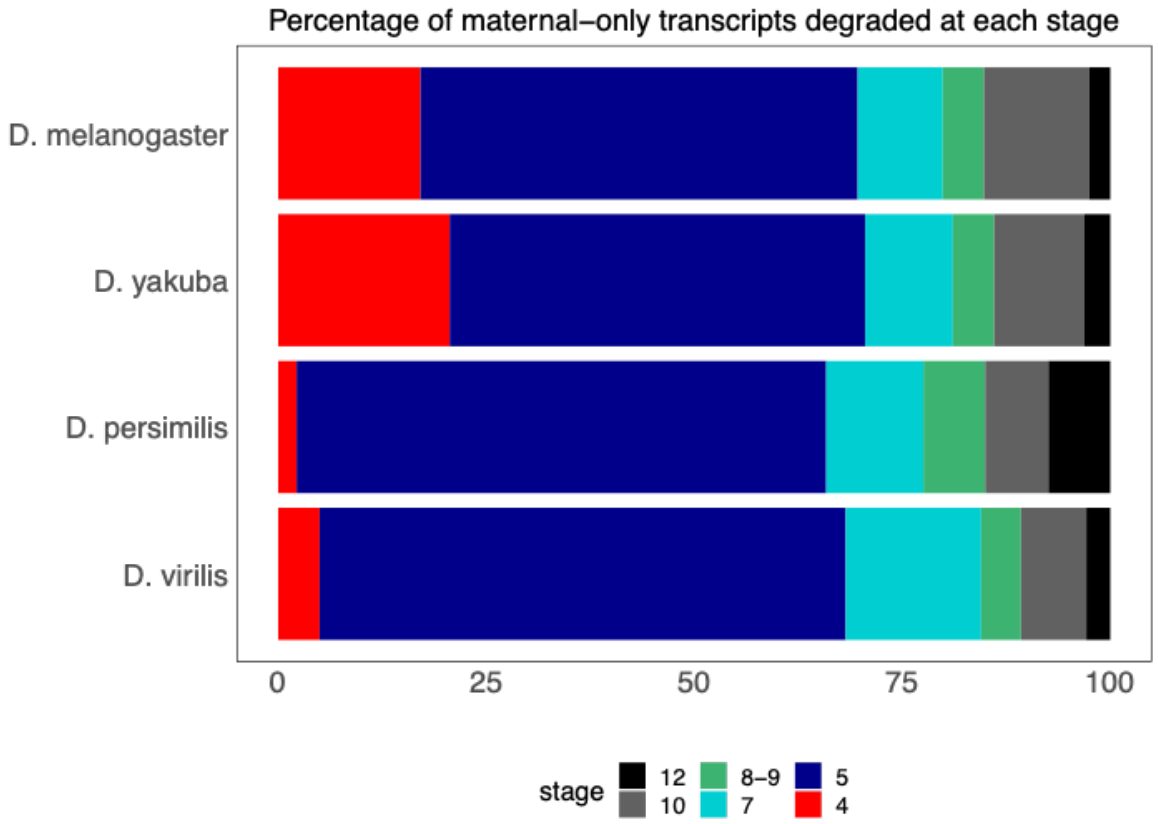


Figure 2.4: Percentages of maternal-only transcripts that degrade at each stage of development in each species. (See Tables 2.1 and 2.2 for raw numbers and percentages.) Maternal-only genes are those with maternally deposited transcripts that are fully degraded and are not transcribed by the zygotic genome or are degraded before their zygotic counterpart is expressed (see Methods). Transcripts are categorized as degraded by a particular stage when they fall below a threshold of 5 CPM, across all three replicates. The proportions shown are representative of genes with orthologs across all four species.

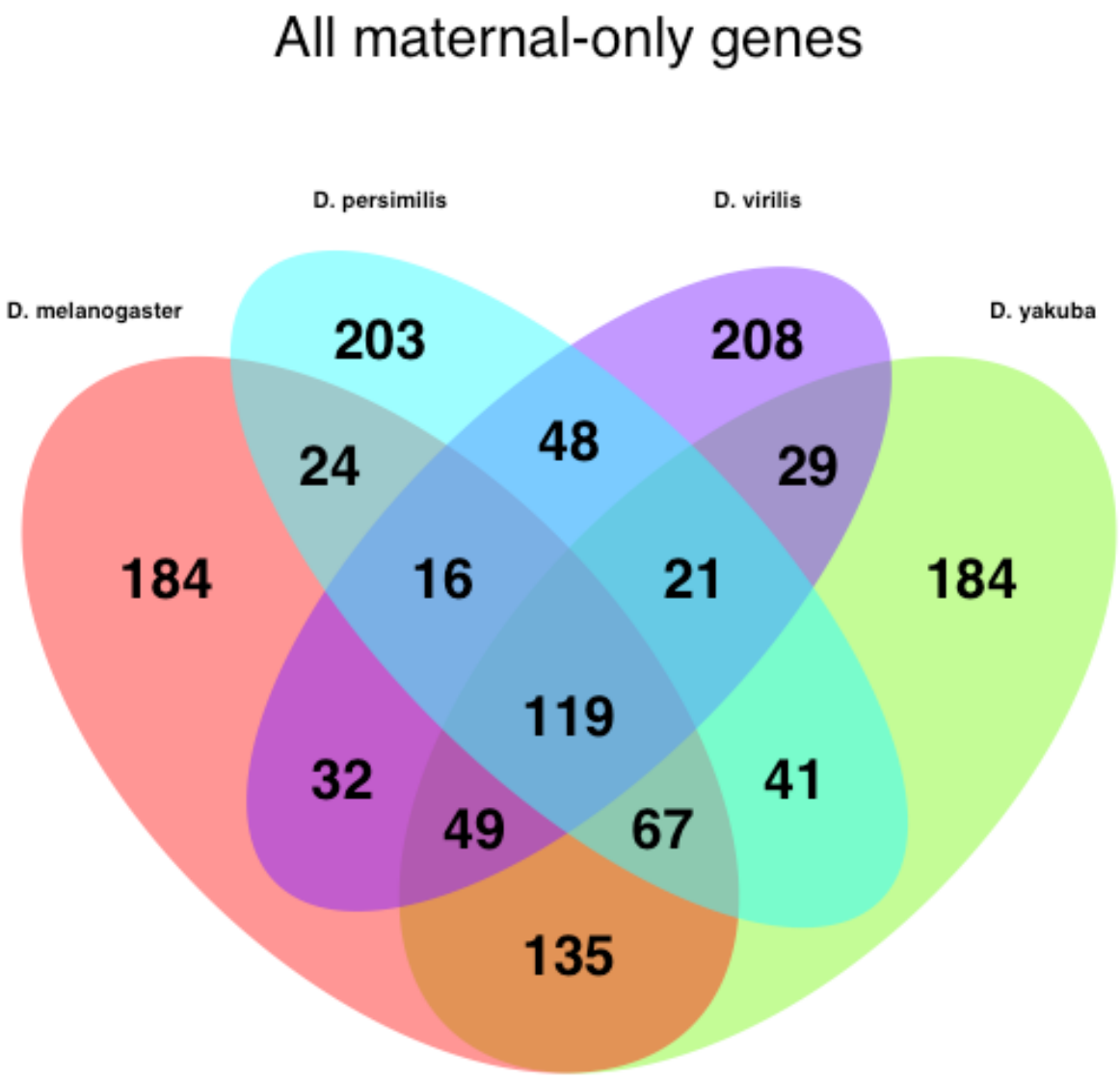


Figure 2.5: Maternal-only genes are often species-specific. Venn diagram showing overlap of genes with maternal-only transcript representation across species. Maternal-only genes represent transcripts that degrade at any stage examined in this analysis.

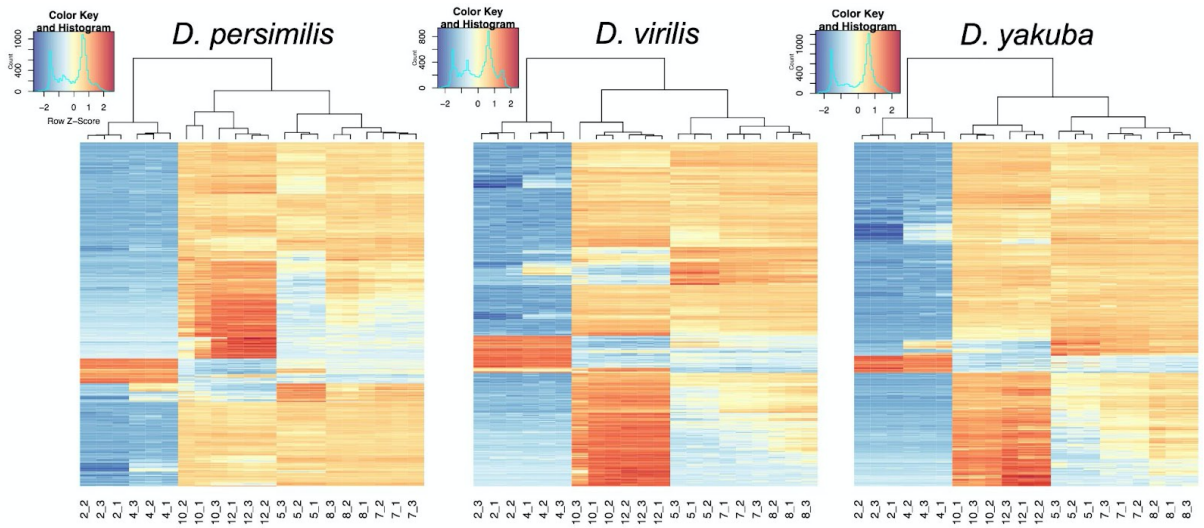


Figure S2.1: Samples cluster by developmental stage in hierarchical clustering based on the top 500 most variable genes in *D. persimilis, D. virilis,* and *D. yakuba*. Hierarchical clustering shows that samples of the same developmental stage cluster, with few exceptions. For stage 10 *D. virilis* and *D. persimilis* samples, one stage 10 sample clusters outside of the rest and in the case of *D. persimilis*, clusters more closely to stage 12 samples. Additionally, one stage 8 sample clusters more closely with stage 7 samples in *D. persimilis*. All other samples cluster with replicates of the same developmental stage. Developmental stages that are closer together in time also cluster with one another.

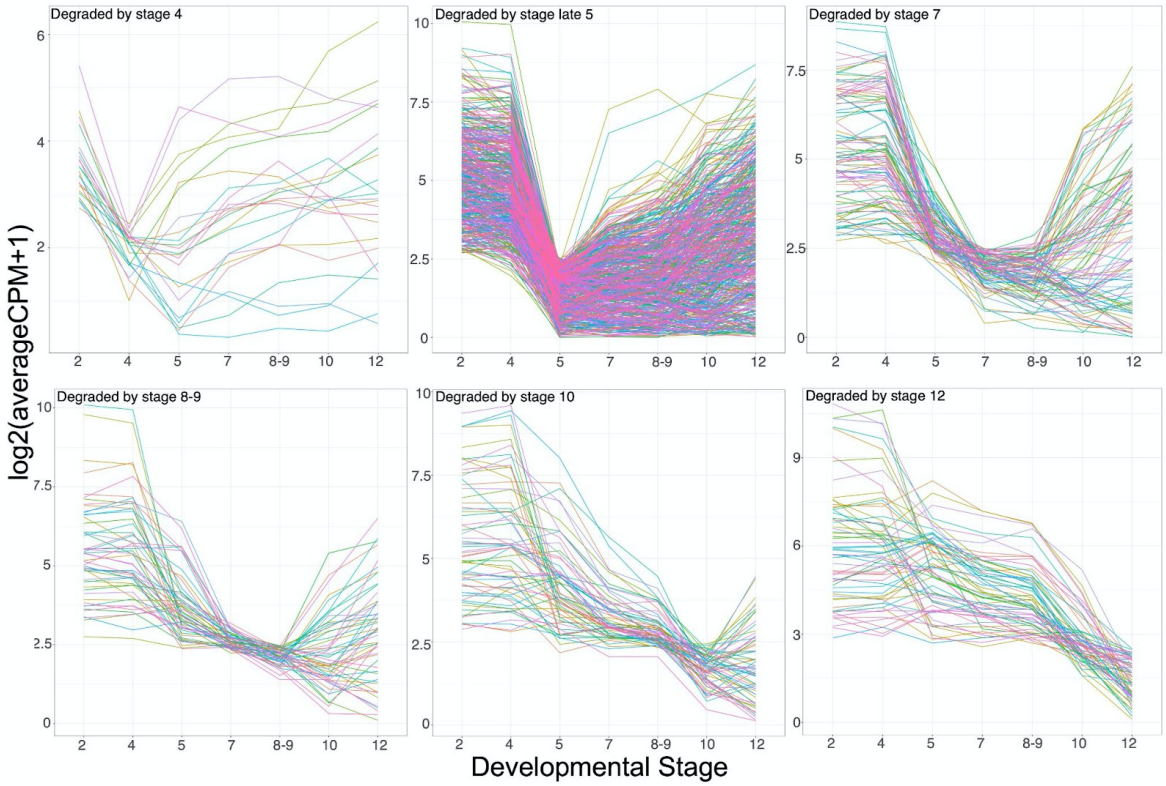


Figure S2.2: Maternal-only transcripts grouped by when they degrade across development in *D. persimilis.* This is the version of Figure 2.3 for *D. persimilis*. See Figure 2.3 caption for details.

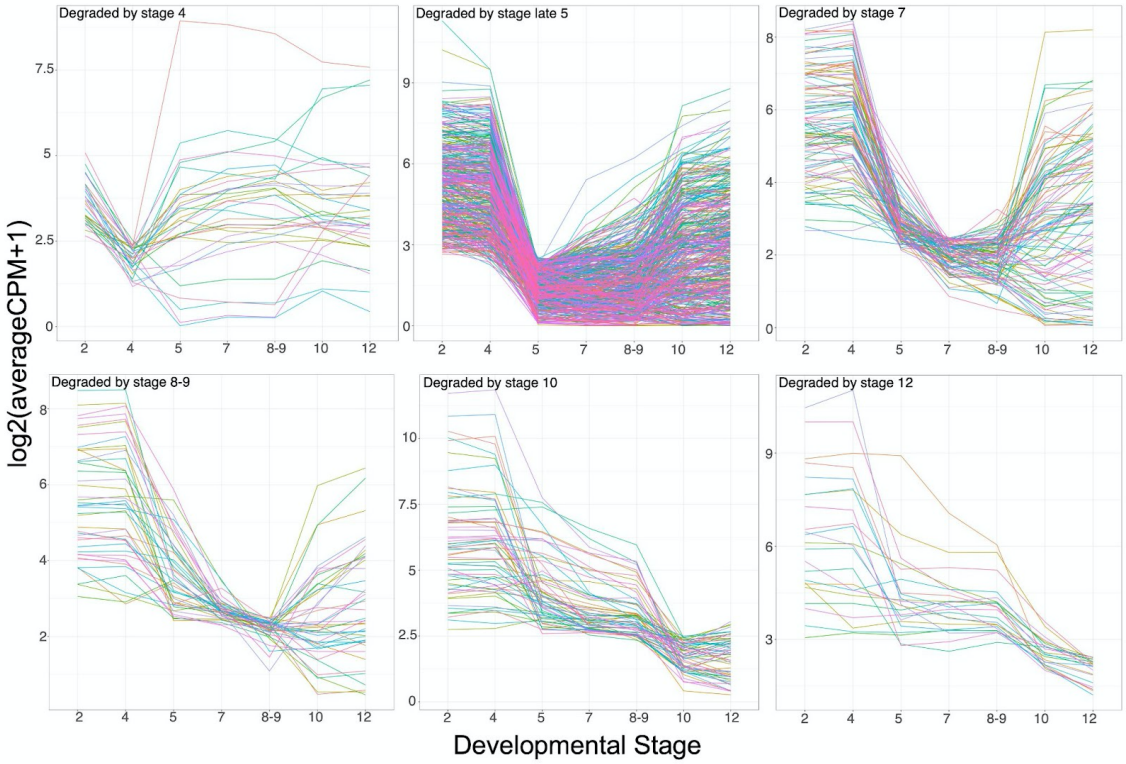


Figure S2.3: Maternal-only transcripts grouped by when they degrade across development in *D. virilis*. This is the version of Figure 2.3 for *D. virilis*. See Figure 2.3 caption for details.

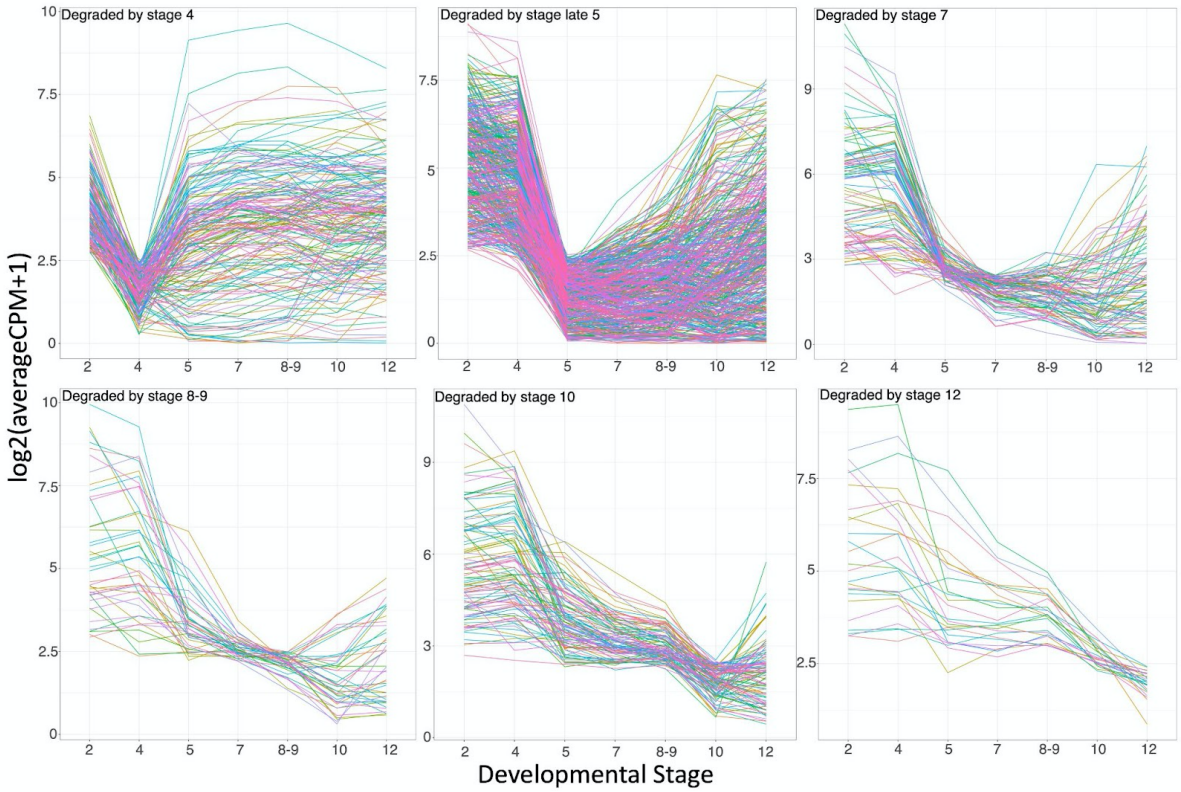


Figure S2.4: Maternal-only transcripts grouped by when they degrade across development in *D. yakuba*. This is the version of Figure 2.3 for *D. yakuba*. See Figure 2.3 caption for details.

Species	Maternal-only transcripts degraded by stage 4	Maternal-only transcripts degraded by stage late 5	Maternal-only transcripts degraded by stage 7	Maternal-only transcripts degraded by stage 8-9	Maternal-only transcripts degraded by stage 10	Maternal-only transcripts degraded by stage 12	Maternal-only transcripts degraded across all stages	All maternally deposited transcripts
D. mel	107	329	64	31	79	16	626	4970
D. yak	133	322	68	32	70	20	645	5096
D. per	12	343	63	40	41	40	539	5242
D. vir	26	330	85	25	41	15	522	5261

Table 2.1: Number of maternal-only transcripts that degrade by each developmental stage as well as the total number of maternally deposited transcripts in each species. Maternal-only genes are those where maternal transcripts are degraded and not expressed by the zygote later in development, or are fully degraded before they are zygotically transcribed (see Methods). Degraded transcripts are categorized by which stage they fall below the threshold of 5 CPM, across all three replicates. Here, numbers represent those genes with orthologs in all four species.

Species	Proportion of maternally deposited transcripts that are maternal-only	Proportion of maternal-only transcripts degraded by stage 4	Proportion of maternal-only transcripts degraded by stage late 5	Proportion of maternal-only transcripts degraded by stage 7	Proportion of maternal-only transcripts degraded by stage 8-9	Proportion of maternal-only transcripts degraded by stage 10	Proportion of maternal-only transcripts degraded by stage 12
D. mel	0.1259557344	0.1709265176	0.5255591054	0.1022364217	0.04952076677	0.1261980831	0.02555910543
D. yak	0.1265698587	0.2062015504	0.4992248062	0.1054263566	0.0496124031	0.1085271318	0.03100775194
D. per	0.1028233499	0.02226345083	0.6363636364	0.1168831169	0.07421150278	0.07606679035	0.07421150278
D. vir	0.09922068048	0.04980842912	0.632183908	0.162835249	0.04789272031	0.0785440613	0.02873563218

Table 2.2: Proportion of maternal-only genes that are degraded at each stage in development. Maternal-only genes are those where the maternal transcripts are degraded and are not zygotically transcribed or are fully degraded before they are zygotically expressed (see Methods). Degraded transcripts are categorized by stage based upon when they fall below the threshold of 5 CPM, across all three replicates. The proportions are representative of genes that have orthologs in all four species.

Species	GO function	Category	GO Term	Bonferroni adjusted p-value	up_down	Comparison
D. vir	BP	response to abiotic stimulus	GO:0009628	4.25E-02	-	early vs. late
D. vir	BP	response to radiation	GO:0009314	3.95E-02	-	early vs. late
D. per	BP	regionalization	GO:0003002	2.74E-02	-	early vs. late
D. per	BP	axis specification	GO:0009798	2.08E-02	-	early vs. late
D. per	BP	axis specification	GO:0009798	3.16E-02	+	late vs. early
D. per	BP	regionalization	GO:0003002	4.18E-02	+	late vs. early

Table 2.3: Significant GO categories for comparisons of early and late decaying maternal-only transcripts. All maternal-only transcripts that degrade by late stage 5 (called early in this table) were compared to maternal-only transcripts that degrade after late stage 5 (called late in the table). The reciprocal comparison of using late genes as the target set and the early genes as the background was also done. Enriched categories are indicated with a “+” sign while depleted categories are signified by a “-” sign in the table. Categories for all significant GO terms fell under the category of Biological Processes (BP).

Species	GO function	Category	GO Term	Bonferroni adjusted p-value	up_down
D. mel	MF	transmembrane transporter activity	GO:0022857	6.00E-03	+
D. mel	MF	transporter activity	GO:0005215	3.60E-03	+
D. mel	MF	binding	GO:0005488	5.18E-03	-
D. mel	MF	protein binding	GO:0005515	6.11E-06	-
D. mel	MF	RNA binding	GO:0003723	4.39E-04	-
D. per	MF	ligand-gated channel activity	GO:0022834	1.28E-02	+
D. per	MF	ligand-gated ion channel activity	GO:0015276	1.28E-02	+
D. per	MF	secondary active transmembrane transporter activity	GO:0015291	4.77E-04	+
D. per	MF	active transmembrane transporter activity	GO:0022804	1.61E-02	+
D. per	MF	transmembrane transporter activity	GO:0022857	4.21E-07	+
D. per	MF	ion transmembrane transporter activity	GO:0015075	1.71E-03	+
D. per	MF	transporter activity	GO:0005215	7.09E-06	+
D. per	MF	binding	GO:0005488	7.94E-05	-
D. per	MF	protein binding	GO:0005515	1.75E-03	-
D. per	MF	nucleic acid binding	GO:0003676	3.70E-02	-
D. per	MF	RNA binding	GO:0003723	6.54E-05	-
D. vir	MF	transmembrane transporter activity	GO:0022857	4.63E-05	+
D. vir	MF	transporter activity	GO:0005215	1.15E-03	+
D. vir	MF	binding	GO:0005488	2.02E-06	-
D. vir	MF	protein binding	GO:0005515	1.11E-02	-
D. vir	MF	nucleic acid binding	GO:0003676	6.04E-05	-
D. vir	MF	RNA binding	GO:0003723	3.06E-04	-

D. yak	MF	lyase activity	GO:0016829	2.95E-02	+
D. yak	MF	transmembrane transporter activity	GO:0022857	4.55E-02	+
D. yak	MF	transporter activity	GO:0005215	2.65E-02	+
D. yak	MF	binding	GO:0005488	3.72E-06	-
D. yak	MF	protein binding	GO:0005515	1.30E-06	-

Table 2.4: Significant GO categories among transcripts that are degraded early. Comparisons were made between transcripts that degrade early, by the end of stage 5, to the background list of all maternally-deposited transcripts, excluding those that are in the target set. This list contains only the categories that fall under Molecular Function (MF). Enriched categories are indicated with a “+” sign while depleted categories are signified by a “-” sign in the table.

Category	GO Term	Bonferroni adjusted p-value	up_down
anion transport	GO:0006820	4.20E-02	+
organic acid metabolic process	GO:0006082	4.94E-02	+
lipid metabolic process	GO:0006629	6.03E-03	+
small molecule metabolic process	GO:0044281	1.64E-02	+
biological_process	GO:0008150	3.52E-02	-
cellular process	GO:0009987	2.79E-03	-
regulation of cellular process	GO:0050794	8.02E-03	-
nitrogen compound metabolic process	GO:0006807	1.94E-03	-
regulation of biological process	GO:0050789	2.57E-04	-
protein metabolic process	GO:0019538	1.49E-02	-
cellular protein metabolic process	GO:0044267	2.67E-02	-
cellular macromolecule metabolic process	GO:0044260	1.06E-04	-
cellular component organization	GO:0016043	3.34E-05	-
cellular component organization or biogenesis	GO:0071840	6.69E-07	-
macromolecule metabolic process	GO:0043170	1.08E-09	-
positive regulation of cellular process	GO:0048522	5.99E-03	-
cell differentiation	GO:0030154	1.66E-02	-
cellular developmental process	GO:0048869	1.27E-02	-
organelle organization	GO:0006996	4.68E-04	-
system development	GO:0048731	2.13E-03	-
cellular component biogenesis	GO:0044085	8.56E-03	-
negative regulation of cellular process	GO:0048523	2.55E-03	-
nucleic acid metabolic process	GO:0090304	4.49E-04	-

macromolecule biosynthetic process	GO:0009059	4.60E-02	-
negative regulation of biological process	GO:0048519	6.63E-05	-
animal organ development	GO:0048513	3.58E-03	-
RNA metabolic process	GO:0016070	9.06E-04	-
tissue development	GO:0009888	7.41E-03	-
gene expression	GO:0010467	4.08E-06	-
regulation of cellular component organization	GO:0051128	4.44E-02	-
nervous system development	GO:0007399	4.31E-03	-
RNA processing	GO:0006396	4.64E-02	-
epithelium development	GO:0060429	2.36E-03	-
generation of neurons	GO:0048699	1.87E-02	-
neurogenesis	GO:0022008	5.17E-03	-
movement of cell or subcellular component	GO:0006928	8.40E-03	-
sensory organ development	GO:0007423	4.86E-02	-

Table 2.5: Significant GO terms under the category of Biological Processes for maternal-only transcripts that degrade by the end of stage 5 in *D. melanogaster*. All of the categories are for enriched or depleted among transcripts in *D. melanogaster* that are degraded by the end of stage 5 when compared to a background list of all genes with maternally deposited transcripts in *D. melanogaster*, minus those in the target list. Enriched categories are indicated with a “+” sign while depleted categories are signified by a “-” sign in the table.

Species	Gene	Primer Name	Primer Sequence	Y or control?	Fragment length
D. vir	kl-2	vir_kl2_L	tagaccaagcgatatacccaatccttcg	Y	634
	kl-2	vir_kl2_R	cgctcatgtactagctgatttcgtggac		
D. vir	Adh	vir_Adh_L	TCACCAGATTGAGCGTACTATTGCGGTA	control	410
	Adh	vir_Adh_R	AGTCGTTTGAGTGGGATGCTCAAGAAGT		
D. vir	Ory	vir_ory_L	agttcttaatggaggctgagcggttttt	Y	218
	Ory	vir_ory_R	gtcgccaagggtatcccttagtttttgt		
D. per	Adh	pse_Adh_L	GTGACCATCACCTTCTATCCCTACGATG	control	421
	Adh	pse_Adh_R	TGACACCAGTAATGGGAGCCAGTTTCTA		
D. per	CG12218 Y-Ψ	pse_7Y_L	GCAGTCGAACCAGTGCAAT	Y	410
		pse_7Y_R	GTGCGGGCAATGGATAAT		
D. mel and D. yak	kl-2	mel_yak_kl2_L	tggaatcaatcgaactcccttagaagt	Y	670
	kl-2	mel_yak_kl2_R	taagtccacggctattacaaatgatcc		
D. mel and D. yak	WDY	mel_yak_WDY_L	ATTATAGTTTGGGATCCTTGGACAG	Y	182
	WDY	mel_yak_WDY_R	GCGTTGTTATAGTTCCAGATTTTCAGGG		
D. mel and D. yak	ftz	mel_ana_yak_ftz_L	CGTGAAGAAGCTGAAGTACACCCC	control	361
	ftz	mel_ana_yak_ftz_R	AATTCGATGATTGATCTCCTGGC		

Table 2.6: Primers used to identify Y chromosome and autosomal genes in each species for genotyping. Primers for *D. virilis*, both for the Y chromosome and control, were from Paris, et al. 2015 as were the control primers for *D. persimilis*. Primers used to identify the Y chromosome gene in *D. persimilis* are from Carvalho and Clark, 2005. All *D. persimilis* primers used here were originally designed for *D. pseudoobscura* and fragment lengths listed are also those for *D. pseudoobscura*.

CONCLUSION AND FUTURE DIRECTIONS

The work in this dissertation addresses how gene regulation during early development of *Drosophila* can evolve across species at the levels of transcription and post-transcriptional processes. Studies previously indicated that maternal transcript deposition and zygotic gene transcription, while driving the coordination of highly regulated and conserved developmental processes, do evolve (Atallah and Lott 2018). The gene regulatory changes that occurred across species and resulted in differences in transcript levels were not previously known. Additionally, other work focusing on gene regulatory evolution, seems to indicate that how evolution happens is dependent upon the system that is under study. Some studies point to more *cis* regulatory changes underlying differences in transcript abundance (McManus et al. 2010; Glaser-Schmitt et al. 2018) while others find more changes in *trans* regulation (Graze et al. 2009; Mack et al. 2016). Here, we chose to focus on studying a time in early development that is highly conserved and represents contributions from two genomes with different regulatory contexts.

In Chapter 1, we address how maternally deposited transcripts and zygotically expressed genes change in regulation across closely related species. By comparing the transcriptomes representing two timepoints in early development, one in which only maternal transcripts are present and the other when the zygotic genome is activated, we find that transcript levels change via different mechanisms at each developmental stage. The results here highlight how the maternal and zygotic genome have different regulatory environments. Differences between species in maternal transcript deposition were largely due to changes in *trans* regulation. In contrast, differences in transcript

levels for mostly zygotic genes were due to changes in *cis*, *trans*, and the combined action of *cis* and *trans*. This study also points to several regulators that may be involved in changes in maternal transcript deposition between species. Through this study and others (Omura and Lott 2020) DNA sequences upstream of maternally deposited genes were found to be enriched for motifs associated with Dref, BEAF-32, and M1BP binding sites. Dref and BEAF-32 are known to be associated with topologically associated domains and are annotated as insulators (Liang et al. 2014; Matzat and Lei 2014; Ali et al. 2016; Ramírez et al. 2018). These findings point to regulators of chromatin state that may regulate transcription during oogenesis and may change between species, resulting in the differences we see in maternal transcript deposition. Future work can follow up on the abundance and function of these proteins during oogenesis, across Drosophila species.

Next, we turned our focus to post-transcriptional regulation during the same period of early development. While much is known about the degradation of maternal transcripts up until the point of gastrulation in *D. melanogaster*, there are many maternal transcripts left that have not yet been degraded. Prior work has not focused on the trajectory or mechanism of the decay of transcripts present at these later stages. Nor do we have any knowledge about how maternal transcript degradation may evolve. The second chapter focuses on comparing maternal transcript degradation across species, to investigate how it may change between species with different developmental times and a range of evolutionary divergence. We focus on maternal-only transcripts, those that are maternally transcribed and degrade before they are produced by the zygote. Across the species in this study, we find that maternal-only transcripts differ in identity

with a small fraction being shared across all four species. This is despite the fact that the larger group of all maternally deposited transcripts is highly conserved (Atallah and Lott 2018). In the future, we will follow up by using motif enrichment analysis to ask what machinery targets these maternal-only transcripts for degradation. We will also add to this study by incorporating maternally deposited transcripts that are not fully degraded before their zygotic counterpart is expressed and comparing their degradation across species. While the identities of maternal-only transcripts differ across species, we found here that roughly the same proportion of maternal-only transcripts are degraded by the end of stage 5, in all species. This indicates that transcripts are stable for longer absolute periods of time in species that take longer to develop. Here, we add to what is currently known about maternal transcript degradation in *Drosophila* by finding that many maternal transcripts may persist until stage 12 across species and finding that maternal-only transcripts degrade in roughly the same proportions across species, at each developmental stage. We do, however, find that a smaller proportion of maternal-only transcripts degrade by stage 4 in species with longer development times. In species with longer development times, this means that these transcripts may be more stable and is something that can be followed up on in the future by extending the analysis to all maternally deposited transcripts, beyond those that are maternal-only.

In total, the work in this dissertation highlights how regulation of gene expression can evolve, at the transcriptional and post-transcriptional levels, across the two genomes that control the critical developmental stage at the very beginning of embryonic development.

REFERENCES

Ali T., R. Renkawitz, and M. Bartkuhn, 2016 Insulators and domains of gene expression. Current Opinion in Genetics & Development 37: 17–26. https://doi.org/10.1016/j.gde.2015.11.009

Atallah J., and S. E. Lott, 2018 Evolution of maternal and zygotic mRNA complements in the early Drosophila embryo, (K. A. Dyer, Ed.). PLoS genetics 14: e1007838. https://doi.org/10.1371/journal.pgen.1007838

Glaser-Schmitt A., A. Zečić, and J. Parsch, 2018 Gene Regulatory Variation in Drosophila melanogaster Renal Tissue. Genetics genetics.301073.2018-46. https://doi.org/10.1534/genetics.118.301073

Graze R. M., L. M. McIntyre, B. J. Main, M. L. Wayne, and S. V. Nuzhdin, 2009 Regulatory Divergence in Drosophila melanogaster and D. simulans, a Genomewide Analysis of Allele-Specific Expression. Genetics 183: 547–561. https://doi.org/10.1534/genetics.109.105957

Liang J., L. Lacroix, A. Gamot, S. Cuddapah, S. Queille, *et al.*, 2014 Chromatin Immunoprecipitation Indirect Peaks Highlight Long-Range Interactions of Insulator Proteins and Pol II Pausing. Molecular cell 53: 672–681. https://doi.org/10.1016/j.molcel.2013.12.029

Mack K. L., P. Campbell, and M. W. Nachman, 2016 Gene regulation and speciation in house mice. Genome Research gr.195743.115. https://doi.org/10.1101/gr.195743.115

Matzat L. H., and E. P. Lei, 2014 Surviving an identity crisis: A revised view of chromatin insulators in the genomics era. Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms 1839: 203–214. https://doi.org/10.1016/j.bbagrm.2013.10.007

McManus C. J., J. D. Coolon, M. O. Duff, J. Eipper-Mains, B. R. Graveley, et al., 2010 Regulatory divergence in Drosophila revealed by mRNA-seq. Genome Research 20: 816–825. https://doi.org/10.1101/gr.102491.109

Omura C. S., and S. E. Lott, 2020 The conserved regulatory basis of mRNA contributions to the early Drosophila embryo differs between the maternal and zygotic genomes. Plos Genet 16: e1008645. https://doi.org/10.1371/journal.pgen.1008645

Ramírez F., V. Bhardwaj, L. Arrigoni, K. C. Lam, B. A. Grüning, et al., 2018 High-resolution TADs reveal DNA sequences underlying genome organization in flies. Nature Communications 9: 1–15. https://doi.org/10.1038/s41467-017-02525-w