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
Department of Evolution and Ecology, University of California, Davis CA 95616
email: elcartwright@ucdavis.edu, selott@ucdavis.edu

Abstract

During embryogenesis in animals, initial developmental processes are driven entirely by maternally provided gene products that are deposited into the oocyte. The zygotic genome is transcriptionally activated later, when developmental control is handed off from maternal gene products to the zygote during the maternal to zygotic transition (MZT). The MZT is highly regulated and conserved across all animals, and while some details change across model systems where it has been studied, most are too evolutionarily diverged to make comparisons as to how this process evolves. There are differences in maternal gene products and their zygotic complements across *Drosophila* species, so here we used hybrid crosses between sister species of *Drosophila* (*D.* simulans, *D. sechellia*, and *D. mauritiana*) and transcriptomics to determine how gene regulation changes in early embryogenesis across species. We find that regulation of maternal transcript deposition and zygotic transcription evolve through different mechanisms. Changes in transcript levels occur predominantly through differences in *trans* regulation for maternal genes, while changes in zygotic transcription occur through a combination of regulatory changes in *cis*, *trans*, and both *cis* and *trans*. We find that patterns of transcript level inheritance in hybrids relative to parental species differs between maternal and zygotic transcripts; maternal transcript levels are more likely to be conserved but both stages have a large proportion of transcripts showing dominance of one parental species. Differences in the underlying regulatory landscape in the mother and the zygote are likely the primary determinants for how maternal and zygotic transcripts evolve species.

Introduction

Many critical early developmental processes are common across all Metazoans, including axial patterning and germ cell specification. The earliest of these conserved processes are regulated by maternally provided RNA and proteins, which lay the foundation for the rest of development (Vastenhouw, Cao, and Lipshitz 2019) (Tadros and Lipshitz 2009). These maternally derived gene products carry out all initial developmental events prior to activation of the zygotic genome because at the time of fertilization, the zygotic genome is transcriptionally quiescent. Post-transcriptional mechanisms also play an important role in regulating the amount of maternal gene products present (Rouget et al. 2010; Barckmann and

Simonelig 2013; Tadros et al. 2007), which is beneficial as the zygotic genome is not transcriptionally active in the earliest stages of development. 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 (Vastenhouw, Cao, and Lipshitz 2019; Tadros and Lipshitz 2005; Baroux et al. 2008).

While many aspects of early development are broadly similar across species, certain features of the MZT vary, such as the length of time that developing organisms rely solely on maternal factors and the proportion of the genome that is maternally loaded into the oocyte as mRNA (Vastenhouw, Cao, and Lipshitz 2019). Previous work identified that the maternally deposited transcripts and zygotically transcribed genes differ during early development across species of *Drosophila* (Atallah and Lott 2018) and more broadly between zebrafish, flies, and mice (Heyn *et al.* 2014). How differences in gene expression can arise between species in such a highly conserved and tightly regulated early developmental process is unknown. In this study, we characterize the regulatory basis of changes in transcript representation during early development to gain insight into the evolutionary process underlying changes in gene expression, and to understand how transcription is regulated during this critical developmental period.

When considering how regulation of gene transcription evolves between species, one fundamental question is whether differences in expression occur due to regulatory changes in *cis* regulatory elements (such as enhancers or promoters) or *trans*-acting factors (such as transcription factor proteins or miRNAs). Unlike changes in *cis* regulation, which will only affect the allele where the change occurred (assuming the absence of transvection), *trans* regulators can be pleiotropic and affect the expression of many genes. Previous studies implemented the use of genetic hybrids (Wittkopp, Haerum, and Clark 2004; Landry et al. 2005) (McManus *et al.* 2010; Coolon *et al.* 2015) and methods of detecting allele-specific expression (León-Novelo et al. 2014; Graze et al. 2009) to address this question genome-wide. In these investigations, patterns of gene regulatory evolution are determined by comparing transcript levels in hybrids to those in parental lines. Several studies point to differences in *cis* regulation as the primary mechanism of change in transcript abundance between species, while other studies indicate that *trans* changes are more widespread. For example, a majority of *trans* regulatory changes were identified as contributing to tissue specific expression divergence in Malpighian tubules (organs that perform excretory and osmoregulation functions) of different strains of *D. melanogaster* (Glaser-Schmitt, Zečić, and Parsch 2018) and between whole bodies of *D. melanogaster* and *D. sechellia* females (McManus *et al.* 2010). In contrast to the findings in these *Drosophila* studies, more changes in *cis* than in *trans* regulation were identified as contributing to divergent gene expression in the testes of different species of house mice (Mack, Campbell, and Nachman 2016). Additional work in *Drosophila* heads (Graze *et al.* 2009) also indicate *cis* regulatory divergence as the leading contributor to regulatory change between species. These studies indicate that the mechanisms of gene regulatory evolution clearly depend on the system and the developmental period or tissue type examined (Buchberger *et al.* 2019).

To study the regulatory basis of differences in transcript levels between species during early development, we focus 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 before and after zygotic genome activation (ZGA) (Atallah and Lott 2018). We looked at whether alterations in gene expression occur due to changes in *cis*, in *trans*, or in a combination of the two.

We find that patterns of gene regulatory changes between these species are distinct across developmental stages, when comparing hybrids and parental lines of the species *D. simulans*, *D. sechellia*, and *D. mauritiana*. Differences in maternal transcript deposition 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. Even though it may be surprising to find many differences in transcript abundance at the maternal stage due to changes in *trans*, as changes in *trans* regulators are more likely to have pleiotropic effects, our results suggest that the maternal stage may have unique features that require gene regulation to evolve via *trans* changes. We find more genes with conserved gene expression between species at the maternal stage as compared to the zygotic stage. These species are very closely related and conservation of gene expression is expected, however we find a bigger proportion of genes at the maternal stage in comparison to the zygotic stage that are misexpressed in the hybrids relative to the parental lines. Additionally, many genes in the hybrid have a dominant mode of inheritance, where expression in the hybrid at a specific developmental stage is more similar to one parental species than the other, at the same developmental stage. Overall, we find distinct patterns of gene regulatory changes at the two embryonic timepoints, before and after ZGA, indicating changes in gene regulation vary based on the developmental context.

Results

In order to determine the regulatory basis of changes in maternal transcript deposition and zygotic gene expression between species, we performed a series of crosses between closely related species followed by RNA-Seq on resulting embryos (Figure 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)). 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 (Campos-Ortega and Hartenstein 2013; Bownes 1975)). While the zygotic genome is fully activated at this developmental stage, maternal transcripts are not entirely degraded at this time 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 species were crossed to produce F1 females, whose embryos were collected at a stage when all the transcripts in the egg are

maternal in origin (stage 2). The female F1s were produced from a hybrid cross (e.g. *D. simulans* females crossed to *D. mauritiana* males produce (*sim* x *mau*) F1 females), which were then crossed to males that were the same species as the females in the initial cross (here, *D. simulans*) to produce hybrids stage 2 embryos. This example cross will be denoted as (*sim* x *mau*) x *sim*, with the first term of the cross (*sim* x *mau*) indicating the maternal F1 genotype, and the second term of the cross (*sim*) indicating the paternal genotype (also see Figure 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). 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 can be highly reproducible, despite replicate samples representing both biological and technical replicates (Lott et al. 2014; Lott et al. 2011; Paris et al. 2015; Atallah and Lott 2018). Our current study extends this to include replicates of F1 and F2 crosses 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 2, A,B,D,E, Table S1). For example, stage 2 samples of the (*mau* x *sim*) x *mau* and (*sim* x *mau*) x *sim* hybrid crosses, have correlation coefficients that range from 0.965 to 0.995 (Table S1). Stage 5 hybrids from the *mau* x *sim* cross have equally high correlation coefficients, ranging from 0.980 to 0.996 (Table S1). Similarly, correlation coefficients of *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), with one notable exception. When we compare 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. 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, have high correlation coefficients, ranging from 0.955 to 0.972. In contrast, *sim* x *mau* stage 5 hybrid embryos, when compared to *D. mauritiana* embryos, have lower correlation coefficients, ranging from 0.863 to 0.887. In this particular comparison, the lower correlation coefficients is 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 samples of different stages yields strikingly lower correlation coefficients (Figure 2, C,F, Table S1), emphasizing the turnover of transcripts between these stages. When comparing stage 2 hybrids from crosses

with *D. mauritiana* and *D. simualans* to stage 5 hybrids from the same cross, correlation coefficients range from 0.483 to 0.573 (Table S1). The correlation coefficients are lower between stages, indicating that the pool of transcripts present at the maternal stage is different from that at the zygotic stage of development.

While samples of the same stage but different genotypes have similar transcript abundance, they are still distinguishable by genotype. Samples of the same stage are highly correlated by their Spearman's rank correlation coefficients but stage-matched embryos separate out by genotype in PCA analysis (Figure S1) by the second principal component. The second principle component of this PCA analysis accounts for between 6.94-8.44% of the variance in the three sets of comparisons (the first principal component corresponds to developmental stage and explains between 80.65% and 81.86% of the variance). Although samples of the same stage have similar transcript abundances, as evidenced by correlation coefficients, they are most similar to samples of the same genotype, as seen through PCA analysis.

Regulatory changes at the maternal stage of development

Changes in gene expression can occur through many mechanisms: alterations in chromatin state, differences in *cis* or in *trans* regulation, or through post-transcriptional modifications. Here, we examine how *cis* and *trans* regulation evolves to differentially affect transcript levels during the maternal to zygotic transition, across species of *Drosophila*. Changes in *cis* regulation 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 occur via changes in factors that bind to the DNA, such as transcription factor proteins. Changes in *trans* regulation affect gene expression independent of allele-specificity.

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) to determine allelic imbalance between parental lines and within hybrid embryos. Reads from each sample were aligned to both genomes of the parental species used in the cross. In these alignments, we identified sites with fixed differences between lines of species used to determine the parental species of origin for each read (see Methods). We used the results from both directions of mapping, meaning mapping to both parental genomes, to determine mapping bias and allelic imbalance. When reads differentially mapped to the two parental genomes, we identified these genes as biased in their mappings and removed them from the analysis (see Methods). Genes that were not biased in their mapping and had a read count of >5 reads were retained for analysis.

Transcript levels between each parental species were compared using PG models to identify differential maternal transcript deposition between the parental lines. Similarly, transcript levels of species-specific alleles in the stage 2 embryos produced by hybrid mothers were compared to determine differential deposition of maternal transcripts into the embryo. We compared the two sets of analyses to determine the regulatory basis of 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, based on comparisons made with confidence intervals of the bias parameters generated through the PG models (see Methods).

We find that most regulatory changes underlying differentially maternally deposited transcripts occurred in *trans* between each pair of species examined (Figure 3A and C, Figure S2), where a change in a transcription factor or other *trans*-acting regulatory element affects both alleles equally (shown if Figure 4A and B). In all pairs of comparisons, the proportion of *trans* changes was higher than any other category of changes. Interestingly, 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 24.2%) of regulatory changes between species at the maternal stage changed only in *cis* regulation. Slightly fewer regulatory changes occur due to a combination of *cis* and *trans* acting factors (between 13.4% and 23.1% in all comparisons). Most genes that change in *cis* and *trans* 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. We also find a percentage of genes with conserved levels of maternal transcript deposition between species, between 16.5% 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 find 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 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 representative trans changes 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. Examples of individual genes with changes attributed to *trans* regulatory differences at the maternal stage are represented in Figure 4A and B.

As *trans* regulatory changes can affect numerous genetic loci, we focused on identifying *trans* regulators that may differentially affect the maternal transcript deposition of many genes between the species studied. In order to identify binding sites of *trans* factors that may differentially regulate maternally deposited transcripts, we used a computational approach to search for overrepresented motifs among genes that change in maternal transcript deposition and among those that have conserved transcript levels (see Methods). We looked at the upstream regions of all genes that change in regulation, not only those with differences in *trans* regulation, because genes with changes in *cis* regulation may have had changes that affect the binding of the same *trans* regulators. We used the upstream regions of genes in *D. simulans* and *D. sechellia* because the *D. mauritiana* genome is not as well annotated compared to the other two species in this study. Using HOMER and MEME (Heinz et al. 2010; Bailey and Elkan 1994) we find motifs associated with insulator binding in the upstream regions of genes changing in regulation (*trans*, *cis*, *cis* + *trans*, *cis* x *trans*, and compensatory) as well as in the genes expressed at this stage that do not change in regulation (conserved; see Methods). We find motifs associated with insulator binding in the upstream regions of genes changing in regulation (*trans*, *cis*, *cis* + *trans*, *cis* x *trans*, and compensatory) as well as in the genes

expressed at this stage that do not change in regulation (conserved) (Table S2). The Dref/BEAF-32 binding site (BEAF-32 and Dref bind overlapping DNA sequences, (Hart, Cuvier, and Laemmli 1999)) is the most significantly enriched (Table S2), and these factors are annotated as insulators (Ali, Renkawitz, and Bartkuhn 2016; Matzat and Lei 2014) and known to be present at topologically associated domain (TAD) boundaries (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 S2). M1BP is involved in transcriptional regulation and RNA polymerase II pausing at the promoter of genes (J. Li and Gilmour 2013), which may also be associated with chromatin state (Ramírez *et al.* 2018). Transcript abundance data from our study indicates that Dref, BEAF-32, and M1BP are differentially maternally deposited in several between-species comparisons (Figure S4), 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 enriched in all maternal 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 (mostly zygotic, see Methods).

While we find that most genes change in *trans* regulation between *D. simulans* and its sister species at the maternal stage of development, we see different patterns of regulatory changes after zygotic genome activation. At the zygotic stage, differences in gene regulation between the three species examined occur mostly in both *cis* and *trans*, where changes in both types of regulatory elements affect the transcript level of single genes (Figure 3B and D, Figure S3). Using the framework outlined by previous studies (Coolon et al. 2014; McManus et al. 2010) (Landry et al. 2005), we define changes in *cis* + *trans* as cases where changes in *cis* and in *trans* affect gene expression in the same direction. Here, the allele with higher expression in the hybrid comes from the parental line with the higher level of expression. *Cis* x *trans* interactions are described as cases where the changes in *cis* and in *trans* have opposing effects on gene expression and result in expression patterns where the allele with lower expression in the hybrid comes from the parental line with the higher level of expression. For zygotic genes at stage 5, changes in both *cis* and *trans* regulatory elements (either *cis* + *trans* or *cis* x *trans* interactions) account for expression differences in between 39.0% and 46.3% of genes in our between-species comparisons. We also see a higher proportion of these interactions occurring in a *cis* + *trans* pattern (between 29.4% and 34.7% of all genes) as opposed to a *cis* x *trans* pattern (between 9.1% and 11.6% of all genes) of regulatory interactions. *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 find between 14.8% and 20.8% of genes changing only

in *trans* regulation. There are between 16.3% and 29.9% 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 find a larger proportion of genes with compensatory changes (between 7.5% and 10.4% of all genes) in gene regulation and a smaller proportion of genes that are conserved (between 6.2% and 8.4% of all genes) between each pair of species comparisons. The smaller number 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*, which are required for segmentation along the anterior-posterior axis (Figure 4C and D).

In contrast to what we have found for regulation at the maternal stage, where transcription may be broadly determined by chromatin boundaries, 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, Cell, 1992, n.d.). 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 changes in zygotic *trans* regulation affected particular types or categories of genes. 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). The significantly enriched GO categories for molecular function and biological process categories that are shared amongst two or more crosses are shown in Figure 5. Identifying GO categories over multiple crosses identifies the types of genes that evolve changes repeatedly over evolution. Shared categories include those related to DNA binding, positive regulation of transcription by RNA polymerase II, cell fate determination, and several developmental categories. We also find biological process categories unique to a specific cross (Figure S5). Again, many of these categories are related to particular developmental processes, consistent with what is known about regulation of transcription at the zygotic stage.

Modes of Inheritance in Hybrids

Misexpression in hybrid offspring has been used to examine regulatory incompatibilities that may contribute to speciation (Mack, Campbell, and Nachman 2016; Moehring, Teeter, and Noor 2007; Michalak 2003). Total expression levels in the hybrid (the sum of the expression of the two species-specific alleles) can be similar to one parental species or the other, or they can deviate from both parental levels entirely. Genes with much higher or lower expression levels than either parental species are considered misexpressed, and can be sources of regulatory incompatibilities between two species. We use methods developed by Gibson, et al., 2004 to define conserved, misexpressed, dominant, and additive inheritance, although many studies use somewhat different metrics to categorize modes of inheritance, making it difficult to compare findings across the field. As in previous studies (Gibson et al. 2004; Landry et al. 2005; McManus et al. 2010; Coolon et al. 2015), we use a conservative fold change of 1.25 ($\log_2$-fold change of 0.32) to define those genes that do not change in transcript abundance between genotypes. Yet even with this conservative cutoff, we find a substantial proportion of genes with conserved transcript levels at each developmental timepoint. Cases where transcript levels

are higher (overdominant) or lower (underdominant) in the hybrid relative to either parental species are categorized as misexpressed. In other cases, the total transcript abundance in the hybrids is more similar to one parent versus the other. Here, we categorize the parental line with expression most similar to the hybrid as dominant. Expression in the hybrid can also have a level intermediate to both parental species (additive).

Strikingly, while many genes show conservation of expression levels between parental species and in the hybrids at both developmental stages, we find a much higher percentage of conserved transcript levels between parents and hybrids for genes that are maternally deposited (Figure 6). We find 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 find conserved transcript abundance in between 4.3% and 7.8% 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.

Of the genes that do not have conserved transcript levels in the stage 2 hybrids, most genes show dominance to one parental species or the other (Figure 6). Few genes show additive patterns of inheritance, in keeping with findings from other studies (Gibson *et al.* 2004; McManus *et al.* 2010). We also report a fraction of misexpressed genes as being over- or under-dominant in the hybrids in relation to expression levels in the parental lines. Of those genes that are dominant at stage 2, a higher proportion 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. In the *D. simulans*/*D. mauritiana* stage 2 comparison, between 39.0% and 39.7% of all genes had transcript levels that were more similar to the *D. simulans* stage 2 samples than the *D. mauritiana* stage 2 samples. In contrast, between 11.2% and 12.4% of all genes in the hybrid have transcript levels more like the *D. mauritiana* stage 2 samples in this comparison. We find that *D. mauritiana* has the least dominance in any cross involving this species. In the *D. sechellia*/*D.mauritiana* comparisions at stage 2, we find that 14.3-14.8% of all genes in the hybrid look more like the *D. mauritiana* samples while 24.0-26.4% of all genes in the hybrid look more like the *D. sechellia* samples. In the comparison with *D. simulans* and *D. sechellia*, more genes in the hybrid had transcript levels similar to the *D. simulans* parental species (26.3%) than the *D. sechellia* parental species (12.58%). Taken together, our findings indicate that *D. simulans* has the most dominant effect on gene expression at the maternal stage in development in this comparison, while *D. mauritiana* has the least dominant effect, with dominance in *D. sechellia* falling between the other two species.

We find similar patterns of inheritance in comparisons of stage 5 samples with a few notable exceptions (Figure 6). Of the mostly zygotic and zygotic-only genes that are not conserved in the hybrids at stage 5, many show a similar pattern of dominant expression where the hybrid expression is more similar to one parental species or the other. As in the maternal stage (stage 2), crosses with *D. simulans* have a higher percentage of genes with expression in the hybrid more similar to *D. simulans* than to the other parental species in the cross. For example, in crosses between *D. simulans* and *D. mauritiana*, between 22.3% and 26.8% of all

genes have an expression level similar to *D. simulans* while only between 7.9% and 12.1% of all genes have an expression level more similar to *D. mauritiana*. In the *D.* simulans/*D. sechellia* comparison, more genes have an expression level in the hybrid more similar to *D. simulans* parental samples (24.2%) than to *D. sechellia* (17.2%). Similar to the stage 2 comparisons, *D. sechellia* exerts the second most dominant effect in the hybrids, with *D. simulans* having the most dominant effect. In crosses with *D. sechellia* and *D. mauritiana*, between 23.7% and 26.1% of genes show *D. sechellia* dominance while only between 11.5% and 13.2% of genes show *D. mauritiana* dominance. Interestingly, in crosses where *D. mauritiana* is the maternal species, there is a slightly higher percentage of genes showing *D. mauritiana* dominance (12.1% versus 7.9% in the crosses involving *D. simulans* and although less striking, 13.2% versus 11.5% in the crosses involving *D. sechellia*). This is likely due to the interaction of residual maternal factors that have not yet degraded by stage 5 with gene products from alleles of the paternal species expressed after zygotic genome activation. In contrast to stage 2, we find more genes that have an additive mode of inheritance or that are misexpressed in the hybrids at stage 5 (between 27.9% and 45.7% of all genes in stage 2 crosses vs. between 50.8% and 61.3% in stage 5 crosses). Previous studies indicate that additive inheritance is associated with *cis* regulatory divergence (Lemos *et al.* 2008; McManus *et al.* 2010). This is in keeping 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.

Discussion

When looking at both regulatory changes and modes of inheritance, we find 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 between species (Atallah and Lott 2018; Heyn et al. 2014) and indicates that the maternal stage is highly conserved. The post-ZGA stage examined here is more complex. While widespread ZGA has occurred, there is still a large proportion of maternal transcripts remaining (while prior studies have varying estimates, roughly 50% of total transcript pool at late stage 5 is maternally derived (Tadros et al. 2007; Lott et al. 2014; Thomsen et al. 2010; De Renzis et al. 2007)). In order to examine how zygotic genes change in regulation, it was necessary to focus on genes without a large maternal contribution to transcript levels at the zygotic stage in development. However, sampling a subset of zygotic genes may have eliminated genes with conserved transcript levels between species. Many genes that are expressed at the zygotic stage are involved in regulating conserved processes and housekeeping functions. Transcript levels for genes involved in these processes may be less susceptible to change over the course of evolution but as many are maternally deposited, they were removed from this analysis. Further, 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 (Barckmann and Simonelig 2013; Rouget et al. 2010; Tadros et al. 2007), so it is not clear that a

high degree of conservation at the transcript level should be necessary. Further study is needed to disentangle conservation at the transcript and protein levels of maternal factors across species.

In addition to finding many genes with conserved transcript abundance between species, we find a large proportion of genes with dominant patterns of inheritance at the maternal and the zygotic stages. At both developmental timepoints, *D. simulans* has the most dominance, where the highest proportion of genes with dominant inheritance in the hybrid is most similar in expression to the *D. simulans* parental line, in crosses involving *D. simulans*. The second largest category of genes with dominant patterns of inheritance in the hybrids have expression levels most similar to the *D. sechellia* parental line. *D. mauritiana* exerts the least amount of dominance in terms of inheritance patterns in the hybrids. The dominant patterns of inheritance are in keeping with the high number of *trans* regulatory changes we see, as previous work identified *trans* changes to more likely to be associated with dominance (Lemos *et al.* 2008). Other work with these species also identified many genes with dominant patterns of expression in hybrids relative to parental lines (Moehring, Teeter, and Noor 2007; Coolon et al. 2015; McManus et al. 2010). We find more genes in hybrids with transcript levels more closely matching those in *D. simulans* in comparison to *D. mauritiana*, similar to previous work (Moehring, Teeter, and Noor 2007). When comparing *D. simulans* and *D. sechellia*, we see a higher proportion of genes in hybrids with total expression levels similar to *D. simulans*. This is opposite of the pattern observed in a previous study, which found equal levels of dominance from both *D. simulans* and *D. sechellia* in the hybrid (Moehring, Teeter, and Noor 2007). Our findings and those from previous studies may differ in the *D. simulans* and *D. sechellia* comparison because different developmental stages were examined in each study. Differences in the species compared and developmental stages or tissue types examined across studies make it difficult to draw conclusions from direct comparisons, but may indicate something fundamental about how gene expression is regulated or evolves at different developmental stages or in different tissues.

Patterns of gene regulatory changes differ greatly between the developmental stages examined, even though the two stages have similar modes of dominant inheritance. We find an overwhelming number of *trans* regulatory changes that result in differential maternal transcript deposition between the species examined. The biological and regulatory context differs between the two stages and may explain why patterns of gene regulatory evolution are different at the developmental timepoints we examined. Maternal transcripts are produced by support cells called nurse cells during oogenesis, and either transported by actin-dependent mechanisms or dumped into the oocyte along with the cytoplasmic contents of the nurse cells upon apoptosis of these cells (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), post-transcriptional regulation of maternal gene products (Salles *et al.* 1994), and subsequent degradation of maternal transcripts (Tadros et al. 2007; Laver et al. 2015; Bushati et al. 2008). Surprisingly, how transcription is regulated in the nurse cells is not well understood. Nurse cells are polyploid and transcribe at a high level to provide the oocyte with the large stock of transcripts needed (we extract >100ng of total RNA at both stages), which is exceptional considering the oocyte begins as essentially a single cell (though a highly specialized

one). We have found here, and also through another study using computational methods to look for binding motifs in maternal factors across the *Drosophila* genus (Omura CS and Lott SE. The conserved basis of regulation of the maternal and zygotic mRNAs in the early *Drosophila* embryo. *in prep.*), that maternal transcription is associated with factors annotated to be insulators and that interact with topologically associated domain boundaries. These two studies both provide evidence that maternal transcription is being controlled broadly at the level of chromatin state. In this context, a few *trans* regulators are predicted to be responsible for the bulk of maternal transcription, and thus changes in the levels of *trans* regulators at this stage may easily be responsible for changes in transcription level for a number of genes. How changes in *trans* regulators affecting chromatin state can have subtle quantitative effects on transcript level, as observed, seems less clear. Functional investigations of mechanisms of regulatory control at the maternal stage are ongoing.

In contrast to the large proportion of regulatory changes in *trans* at the maternal stage, differences in zygotic gene transcription between these species are characterized by a combination of changes in *cis*, *trans*, *cis* + *trans*, and *cis* x *trans*. Zygotic gene transcription is spatially and temporally regulated across the embryo with enhancer regions playing a large role in where and when genes are expressed (Haines and Eisen 2018). Changes in *trans* regulation, which can affect the expression of many genes, may be detrimental to the developing organism at this stage, whereas changes in *cis* regulation are gene-specific and may only affect gene expression in a subset of the embryo. Fundamental differences in the regulatory landscape of the maternal and zygotic stages likely explain why the evolution of gene expression occurs through different mechanisms for transcripts that are maternally deposited and genes that are zygotically transcribed.

In this study, we find that differences between species in transcript levels of maternally deposited transcripts and zygotically transcribed genes evolve via different patterns of regulatory change. We find that maternal transcript abundance is more conserved and also changes more readily through *trans* regulation in comparison to zygotic complements. Regulatory organization and constraints that are specific to each developmental stage is likely to 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 basis, and more generally what is controlling the regulatory landscape at each stage in development.

Materials and Methods

Crosses and sample acquisition

Hybrid crosses were set up using virgin females from one species and males from another of each of the following three species: *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). Two types of hybrid crosses were established from which embryos were collected: 1) to determine regulatory basis of changes in zygotic

expression, hybrid F1 embryos were collected; and 2) to determine the regulatory basis of changes in maternal expression, embryos produced by hybrid F1 mothers were collected. To sample a developmental timepoint after zygotic genome activation, we chose the very end of blastoderm stage, stage 5 (Bownes' stages (Campos-Ortega and Hartenstein 2013; Bownes 1975)). We define late stage 5 by morphology; it is the point when cellularization is complete but gastrulation has not yet begun. 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. This was done to establish consistency amongst crosses, although there are no known mRNA contributions from the sperm to the zygote, thus the male genotype is unlikely to affect our data. As is conventional in *Drosophila* genetics, we denote these crosses by listing the female genotype first, e.g. *mau* x *sim*, and the male genotype second, in this case *sim*. We then describe the hybrid genotype as (*mau* x *sim*) x *mau* for this cross (also see Figure 1 for cross diagram). This second set of crosses was used to collect stage 2 embryos (Bownes' stages (Campos-Ortega and Hartenstein 2013; Bownes 1975)), during which time only maternal gene products are present. At this point in development, the vitelline membrane has retracted from the anterior and posterior poles of the embryo but pole cells have not yet migrated to the posterior.

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 quickly collected at stage 2 or at the end of stage 5 (Bownes' stages) of embryonic development (Campos-Ortega and Hartenstein 2013; Bownes 1975).

Once staged, 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 S3) designed for use with all three species, two sets for the Y chromosome (ORY and kl2) and one control set (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.

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 low sample protocol. The Illumina protocol uses oligo (dT) beads to enrich for polyadenylated transcripts. Because poly(A) tail length is important in determining 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 (Eichhorn et al. 2016; Lim 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 (Atallah and Lott 2018; Crofton et al. 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 dsDNA BR Assay Kits (Qubit) 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/)(Yang, n.d.). 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. For 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, (H. Li et al. 2009)). Picard tools (version 2.7.1) and GATK tools (Van der Auwera et al., n.d.) 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 then organized and merged using a custom R script (R version 3.4.1, (Team 2013)) (Files S1 and S2). 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). Read assignments to parental genomes were subsequently determined with custom R scripts using the variant sites that exist between the parental genomes. Normalization to the upper quantile was performed across all samples 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.

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 and were removed from the analysis. 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. The orthologs table (Table S4) 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 order to alleviate mapping bias that may occur when mapping the parental and hybrid samples to each parental reference genome, Poisson Gamma models (León-Novelo *et al.* 2014) were employed to calculate mapping bias for every set of mappings, in each pairwise comparison of species.

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) (De Renzis et al. 2007; Thomsen et al. 2010; Tadros et al. 2007; 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 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.

Correlation analysis and PCA

We performed correlation analysis between single embryos across replicates, stages, and genotypes in R (Team 2013) using the Spearman option within the *corr* function. PCA analysis was also performed in R using the *prcomp* function.

Cis/Trans analysis

To address both mapping bias and allelic imbalance, we used 95% confidence intervals (CIs) from Poisson Gamma (PG) models (León-Novelo *et al.* 2014). We used the PG model (with fixed bias parameter, $q = 0.5$) to define differentially expressed genes as genes with CIs that did not overlap 0.50 when comparing the expression levels of parental alleles in each replicate. Genes that appeared differentially expressed in one direction of mapping 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. The genes retained for analysis were categorized using *cis*, *trans*, *cis+trans*, *cisxtrans*, *compensatory*, and *conserved* categories as described in Coolon, et al. 2014, McManus, et al. 2010, and Landry, et al. 2004. We assigned the following categories for regulatory change based on the CIs generated from PG models for individual genes:

cis: Genes categorized as having changes in *cis* are those that are differentially expressed (CIs do not overlap 50%) between the parental species and in the hybrids. (CIs for parental species and hybrids overlap each other for changes purely in *cis*.)

trans: Genes that are differentially expressed between the parental species (CI does not overlap 50%) but are not differentially expressed in the hybrid (CI overlaps 50%).

cis + *trans*: Genes that are differentially expressed in the hybrids and between the parental species (CI does not overlap 50%) and the CI is in the same direction for both the parents and the hybrid (i.e. both are greater than 50%)

cis x *trans*: Genes that are differentially expressed in the hybrids and between the parental species (CI does not overlap 50%) and the CI is in opposite directions for the parents and the hybrid (i.e. one is greater than 50%, the other is less than 50%)

compensatory: Genes that are not differentially expressed between the parental species (CI overlaps 50%) but are differentially expressed in the hybrids (CI does not overlap 50%).

conserved: Genes are not differentially expressed between the parental species or within the hybrids (CIs overlap 50%).

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 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. 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.

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, at each stage, compared to the background of genes that are expressed (having a count >5) in each cross, at each stage. 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 5. All enriched categories are listed in Figure S5.

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 genes changing in regulation or with conserved regulation between species in every set of comparisons at stage 2. 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 represented

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 S2. 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 S2.

Literature Cited

Ali, Tamer, Rainer Renkawitz, and Marek Bartkuhn. 2016. "Insulators and Domains of Gene Expression." *Current Opinion in Genetics & Development* 37 (April): 17–26. doi:10.1016/j.gde.2015.11.009.

Atallah, Joel, and Susan E Lott. 2018. "Evolution of Maternal and Zygotic mRNA Complements in the Early Drosophila Embryo." Edited by Kelly A Dyer. *PLoS Genetics* 14 (12). Public Library of Science: e1007838. doi:10.1371/journal.pgen.1007838.

Bailey, T L, and C Elkan. 1994. "Fitting a Mixture Model by Expectation Maximization to Discover Motifs in Bipolymers."

Barckmann, Bridlin, and Martine Simonelig. 2013. "Control of Maternal mRNA Stability in Germ Cells and Early Embryos." *Biochimica Et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* 1829 (6-7): 714–24. doi: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 (0). Cold Spring Harbor Laboratory Press: 89–100. doi:10.1101/sqb.2008.73.053.

Bownes, Mary. 1975. "A Photographic Study of Development in the Living Embryo of Drosophila Melanogaster." *Development* 33 (3). The Company of Biologists Ltd: 789–801.

Buchberger, Reis, Lu, and Posnien. 2019. "Cloudy with a Chance of Insights: Context Dependent Gene Regulation and Implications for Evolutionary Studies." *Genes* 10 (7). Multidisciplinary Digital Publishing Institute: 492–21. doi:10.3390/genes10070492.

Bushati, Natascha, Alexander Stark, Julius Brennecke, and Stephen M Cohen. 2008. "Temporal Reciprocity of miRNAs and Their Targets During the Maternal-to-Zygotic Transition in Drosophila." *Current Biology* 18 (7): 501–6. doi:10.1016/j.cub.2008.02.081.

Campos-Ortega, Jose A, and Volker Hartenstein. 2013. *The Embryonic Development of Drosophila Melanogaster*. Berlin, Heidelberg: Springer Science & Business Media. doi:10.1007/978-3-662-02454-6.

Clark, Andrew G, Michael B Eisen, Douglas R Smith, Casey M Bergman, Brian Oliver, Therese A Markow, Thomas C Kaufman, et al. 2007. "Evolution of Genes and Genomes on the Drosophila Phylogeny." *Nature* 450 (7167): 203–18. doi:10.1038/nature06341.

Coolon, Joseph D, C Joel McManus, Kraig R Stevenson, Brenton R Graveley, and Patricia J Wittkopp. 2014. "Tempo and Mode of Regulatory Evolution in Drosophila." *Genome Research* 24 (5). Cold Spring Harbor Lab: 797–808. doi:10.1101/gr.163014.113.

Coolon, Joseph D, Kraig R Stevenson, C Joel McManus, Bing Yang, Brenton R Graveley, and Patricia J Wittkopp. 2015. “Molecular Mechanisms and Evolutionary Processes Contributing to Accelerated Divergence of Gene Expression on the Drosophila X Chromosome.” *Molecular Biology and Evolution* 32 (10). Oxford University Press: 2605–15. doi:10.1093/molbev/msv135.

Crofton, Amanda E, Emily L Cartwright, Anna A Feitzinger, and Susan E Lott. 2018. “Effect of Larval Nutrition on Maternal mRNA Contribution to the Drosophila Egg..” *G3: Genes|Genomes|Genetics* 8 (6). G3: Genes, Genomes, Genetics: 1933–41. doi:10.1534/g3.118.200283.

De Renzis, Stefano, Olivier Elemento, Saeed Tavazoie, and Eric F Wieschaus. 2007. “Unmasking Activation of the Zygotic Genome Using Chromosomal Deletions in the Drosophila Embryo.” *PLOS Biology* 5 (5). Public Library of Science: e117. doi:10.1371/journal.pbio.0050117.

Eichhorn, S W, A O Subtelny, I Kronja, and JC Kwasnieski Elife. 2016. “mRNA Poly (a)-Tail Changes Specified by Deadenylation Broadly Reshape Translation in Drosophila Oocytes and Early Embryos.” *Cdn.Elifesciences.org*

, July.

Gibson, Greg, Rebecca Riley-Berger, Larry Harshman, Artyom Kopp, Scott Vacha, Sergey Nuzhdin, and Marta Wayne. 2004. “Extensive Sex-Specific Nonadditivity of Gene Expression in Drosophila Melanogaster.” *Genetics* 167 (4). Genetics: 1791–99. doi:10.1534/genetics.104.026583.

Glaser-Schmitt, Amanda, Aleksandra Zečić, and John Parsch. 2018. “Gene Regulatory Variation in Drosophilamelanogaster Renal Tissue.” *Genetics*, July, genetics.301073.2018–46. doi:10.1534/genetics.118.301073.

Graze, Rita M, Lauren M McIntyre, Bradley J Main, Marta L Wayne, and Sergey V Nuzhdin. 2009. “Regulatory Divergence in Drosophila Melanogaster and D. Simulans, a Genomewide Analysis of Allele-Specific Expression.” *Genetics* 183 (2). Genetics: 547–61. doi:10.1534/genetics.109.105957.

Haines, Jenna E, and Michael B Eisen. 2018. “Patterns of Chromatin Accessibility Along the Anterior-Posterior Axis in the Early Drosophila Embryo.” Edited by Juan M Vaquerizas. *PLoS Genetics* 14 (5): e1007367–25. doi:10.1371/journal.pgen.1007367.

Hart, Craig M, Olivier Cuvier, and Ulrich K Laemmli. 1999. “Evidence for an Antagonistic Relationship Between the Boundary Element-Associated Factor BEAF and the Transcription Factor DREF.” *Chromosoma* 108 (6). Springer-Verlag: 375–83. doi:10.1007/s004120050389.

Heinz, Sven, Christopher Benner, Nathanael Spann, Eric Bertolino, Yin C Lin, Peter Laslo, Jason X Cheng, Cornelis Murre, Harinder Singh, and Christopher K Glass. 2010. “Simple Combinations of Lineage-Determining Transcription Factors Prime Cis-Regulatory Elements Required for Macrophage and B Cell Identities.” *Molecular Cell* 38 (4). Cell Press: 576–89. doi:10.1016/j.molcel.2010.05.004.

Heyn, Patricia, Martin Kircher, Andreas Dahl, Janet Kelso, Pavel Tomancak, Alex T Kalinka, and Karla M Neugebauer. 2014. “The Earliest Transcribed Zygotic Genes Are Short, Newly Evolved, and Different Across Species.” *CellReports* 6 (2). The Authors: 285–92. doi:10.1016/j.celrep.2013.12.030.

Jäckle, Herbert, Diethard Tautz, Reinhard Schuh, Eveline Seifert, and Ruth Lehmann. 1986. "Cross-Regulatory Interactions Among the Gap Genes of Drosophila." *Nature* 324 (6098). Nature Publishing Group: 668–70. doi:10.1038/324668a0.

Johnston, D St, C Nüsslein-Volhard Cell, 1992. n.d. "The Origin of Pattern and Polarity in the Drosophila Embryo." *Elsevier*. doi:10.1016/0092-8674(92)90466-P.

Kugler, Jan-Michael, and Paul Lasko. 2009. "Localization, Anchoring and Translational Control of Oskar, Gurken, Bicoid and Nanos mRNA During Drosophila Oogenesis." *Fly* 3 (1). Taylor & Francis: 15–28. doi:10.4161/fly.3.1.7751.

Lachaise, Daniel, Jean R David, Francoise Lemeunier, Leonidas Tsacas, and Michael Ashburner. 1986. "The Reproductive Relationships of Drosophila Sechellia with D. Mauritiana, D. Simulans, and D. Melanogaster From the Afrotropical Region." *Evolution* 40 (2). John Wiley & Sons, Ltd (10.1111): 262–71. doi:10.1111/j.1558-5646.1986.tb00468.x.

Landry, C R, P J Wittkopp, C H Taubes, J M Ranz, A G Clark, and D L Hartl. 2005. "Compensatory Cis-Trans Evolution and the Dysregulation of Gene Expression in Interspecific Hybrids of Drosophila." *Genetics* 171 (4). Genetics: 1813–22. doi:10.1534/genetics.105.047449.

Laver, John D, Xiao Li, Debashish Ray, Kate B Cook, Noah A Hahn, Syed Nabeel-Shah, Mariana Kekis, 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 (1). BioMed Central: 94. doi: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 (38). National Academy of Sciences: 14471–76. doi:10.1073/pnas.0805160105.

León-Novelo, Luis G, Lauren M McIntyre, Justin M Fear, and Rita M Graze. 2014. "A Flexible Bayesian Method for Detecting Allelic Imbalance in RNA-Seq Data." *BMC Genomics* 15 (1). BioMed Central: 920. doi:10.1186/1471-2164-15-920.

Li, H, B Handsaker, A Wysoker, T Fennell, J Ruan, N Homer, G Marth, G Abecasis, R Durbin, 1000 Genome Project Data Processing Subgroup. 2009. "The Sequence Alignment/Map Format and SAMtools." *Bioinformatics* 25 (16): 2078–79. doi:10.1093/bioinformatics/btp352.

Li, Jian, and David S Gilmour. 2013. "Distinct Mechanisms of Transcriptional Pausing Orchestrated by GAGA Factor and M1BP, a Novel Transcription Factor." *The EMBO Journal* 32 (13). John Wiley & Sons, Ltd: 1829–41. doi:10.1038/emboj.2013.111.

Liang, Jun, Laurent Lacroix, Adrien Gamot, Suresh Cuddapah, Sophie Queille, Priscillia Lhoumaud, Pierre Lepetit, et al. 2014. "Chromatin Immunoprecipitation Indirect Peaks Highlight Long-Range Interactions of Insulator Proteins and Pol II Pausing." *Molecular Cell* 53 (4). Cell Press: 672–81. doi:10.1016/j.molcel.2013.12.029.

Lim, Jaechul, Mihye Lee, Ahyeon Son, Hyeshik Chang, and V Narry Kim. 2016. "mTAIL-Seq Reveals Dynamic Poly(a) Tail Regulation in Oocyte-to-Embryo Development." *Genes & Development* 30 (14): 1671–82. doi:10.1101/gad.284802.116.

Lott, Susan E, Jacqueline E Villalta, Gary P Schroth, Shujun Luo, Leath A Tonkin, and Michael B Eisen. 2011. "Noncanonical Compensation of Zygotic X Transcription in Early Drosophila Melanogaster Development Revealed Through Single-Embryo RNA-Seq." *PLOS Biology* 9 (2). Public Library of Science: e1000590. doi:10.1371/journal.pbio.1000590.

Lott, Susan E, Jacqueline E Villalta, Qi Zhou, Doris Bachtrog, and Michael B Eisen. 2014. “Sex-Specific Embryonic Gene Expression in Species with Newly Evolved Sex Chromosomes..” Edited by Harmit S Malik. *PLoS Genetics* 10 (2). Public Library of Science: e1004159. doi:10.1371/journal.pgen.1004159.

Mack, Katya L, Polly Campbell, and Michael W Nachman. 2016. “Gene Regulation and Speciation in House Mice.” *Genome Research*, February. Cold Spring Harbor Lab, gr.195743.115. doi:10.1101/gr.195743.115.

Macmanes, Matthew D. 2014. “On the Optimal Trimming of High-Throughput mRNA Sequence Data.” *Frontiers in Genetics* 5: 13. doi:10.3389/fgene.2014.00013.

Martin, Marcel. 2011. “Cutadapt Removes Adapter Sequences From High-Throughput Sequencing Reads.” *EMBnet.Journal* 17 (1): pp.10–pp.12. doi:10.14806/ej.17.1.200.

Matzat, Leah H, and Elissa 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 (3). Elsevier: 203–14. doi:10.1016/j.bbagrm.2013.10.007.

McDermott, Shannon R, and Richard M Kliman. 2008. “Estimation of Isolation Times of the Island Species in the Drosophila Simulans Complex From Multilocus DNA Sequence Data.” *PLoS ONE* 3 (6). Public Library of Science: e2442. doi:10.1371/journal.pone.0002442.

McManus, C J, J D Coolon, M O Duff, J Eipper-Mains, B R Graveley, and P J Wittkopp. 2010. “Regulatory Divergence in Drosophila Revealed by mRNA-Seq.” *Genome Research* 20 (6): 816–25. doi:10.1101/gr.102491.109.

Mi, Huaiyu, Anushya Muruganujan, Xiaosong Huang, Dustin Ebert, Caitlin Mills, Xinyu Guo, and Paul D Thomas. 2019. “Protocol Update for Large-Scale Genome and Gene Function Analysis with the PANTHER Classification System (v.14.0).” *Nature Protocols*, February. Springer US, 1–21. doi: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 (7): 1070–76. doi:10.1093/molbev/msg119.

Mische, Sarah, Mingang Li, Madeline Serr, and Thomas S Hays. 2007. “Direct Observation of Regulated Ribonucleoprotein Transport Across the Nurse Cell/Oocyte Boundary.” Edited by Susan Wente. *Molecular Biology of the Cell* 18 (6): 2254–63. doi:10.1091/mbc.e06-10-0959.

Moehring, Amanda J, Katherine C Teeter, and Mohamed 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 (1). Oxford University Press: 137–45. doi:10.1093/molbev/msl142.

Nolte, Viola, Ram Vinay Pandey, Robert Kofler, and Christian Schlötterer. 2013. “Genome-Wide Patterns of Natural Variation Reveal Strong Selective Sweeps and Ongoing Genomic Conflict in Drosophila Mauritiana.” *Genome Research* 23 (1). Cold Spring Harbor Lab: 99–110. doi:10.1101/gr.139873.112.

(Omura CS and Lott SE. The conserved basis of regulation of the maternal and zygotic mRNAs in the early *Drosophila* embryo. *in prep.*)

Paris, Mathilde, Jacqueline E Villalta, Michael B Eisen, and Susan E Lott. 2015. “Sex Bias and Maternal Contribution to Gene Expression Divergence in Drosophila Blastoderm Embryos..” Edited by Andrew G Clark. *PLoS Genetics* 11 (10). Public Library of Science: e1005592. doi:10.1371/journal.pgen.1005592.

Quinlan, Aaron R, and Ira M Hall. 2010. “BEDTools: a Flexible Suite of Utilities for Comparing Genomic Features.” *Bioinformatics* 26 (6): 841–42. doi:10.1093/bioinformatics/btq033.

R Core Team. 2013. “R: A Language and Environment for Statistical Computing.”

Ramírez, Fidel, Vivek Bhardwaj, Laura Arrigoni, Kin Chung Lam, Björn A Grüning, José Villaveces, Bianca Habermann, Asifa Akhtar, and Thomas Manke. 2018. “High-Resolution TADs Reveal DNA Sequences Underlying Genome Organization in Flies.” *Nature Communications* 9 (1). Nature Publishing Group: 1–15. doi:10.1038/s41467-017-02525-w.

Rouget, Christel, Catherine Papin, Anthony Boureux, Anne-Cécile Meunier, Bénédicte Franco, Nicolas Robine, Eric C Lai, Alain Pelisson, and Martine Simonelig. 2010. “Maternal mRNA Deadenylation and Decay by the piRNA Pathway in the Early Drosophila Embryo.” *Nature* 467 (7319). Nature Publishing Group: 1128–44. doi: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 (5193). American Association for the Advancement of Science: 1996–99. doi:10.1126/science.7801127.

Supek, Fran, Matko Bošnjak, Nives Škunca, and Tomislav Šmuc. 2011. “REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms.” Edited by Cynthia Gibas. *PLoS ONE* 6 (7): e21800–e21809. doi:10.1371/journal.pone.0021800.

Tadros, W, and H D Lipshitz. 2009. “The Maternal-to-Zygotic Transition: a Play in Two Acts.” *Development* 136 (18): 3033–42. doi:10.1242/dev.033183.

Tadros, Wael, Aaron L Goldman, Tomas Babak, Fiona Menzies, Leah Vardy, Terry Orr-Weaver, Timothy R Hughes, J Timothy Westwood, Craig A Smibert, and Howard D Lipshitz. 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 (1): 143–55. doi:10.1016/j.devcel.2006.10.005.

Tadros, Wael, and Howard D Lipshitz. 2005. “Setting the Stage for Development: mRNA Translation and Stability During Oocyte Maturation and Egg Activation in Drosophila.” *Developmental Dynamics* 232 (3). John Wiley & Sons, Ltd: 593–608. doi:10.1002/dvdy.20297.

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..” *Development* 125 (18). The Company of Biologists Ltd: 3655–66.

Thomsen, Stefan, Simon Anders, Sarath Chandra Janga, Wolfgang Huber, and Claudio R Alonso. 2010. “Genome-Wide Analysis of mRNA Decay Patterns During Early Drosophila Development.” *Genome Biology* 11 (9). BioMed Central: R93. doi:10.1186/gb-2010-11-9-r93.

Torres-Oliva, Montserrat, Isabel Almudi, Alistair P McGregor, and Nico 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 (1). BioMed Central: 392. doi:10.1186/s12864-016-2646-x.

Trapnell, Cole, Adam Roberts, Loyal Goff, Geo Pertea, Daehwan Kim, David R Kelley, Harold Pimentel, Steven L Salzberg, John L Rinn, and Lior Pachter. 2012. “Differential Gene and

Transcript Expression Analysis of RNA-Seq Experiments with TopHat and Cufflinks." *Nature Protocols* 7 (3). Nature Publishing Group: 562. doi:10.1038/nprot.2012.016.

Van der Auwera, Geraldine A, Mauricio O Carneiro, Christopher Hartl, Ryan Poplin, Guillermo del Angel, Ami Levy-Moonshine, Tadeusz Jordan, et al. n.d. *From FastQ Data to High-Confidence Variant Calls: the Genome Analysis Toolkit Best Practices Pipeline*. *Wiley Online Library*. John Wiley & Sons, Inc. doi:10.1002/0471250953.bi1110s43.

Vastenhouw, Nadine L, Wen Xi Cao, and Howard D Lipshitz. 2019. "The Maternal-to-Zygotic Transition Revisited." *Development* 146 (11): dev161471–20. doi: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 (6995). Nature Publishing Group: 85–88. doi:10.1038/nature02698.

Yang, Jinliang. n.d. "Substitute Reference Genome with Sample Specific SNP Variants." http://yangjl.com/pseudoRef/.

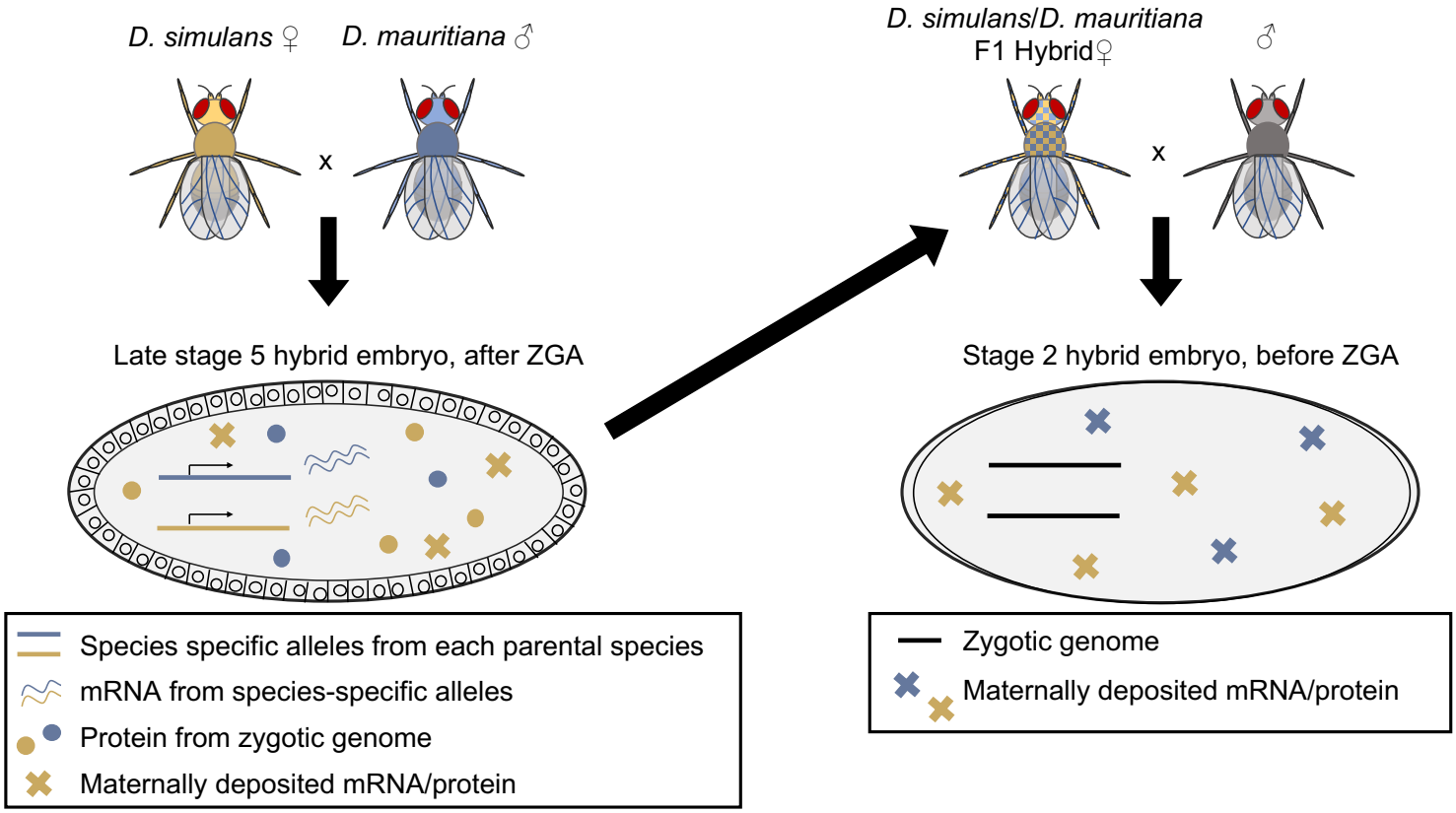


Figure 1: Crosses to produce hybrid embryos for the zygotic and maternal stages. To look at changes in regulation for zygotic genes, hybrid stage 5 embryos (left) were produced by crossing two parental species and collecting their eggs at the appropriate stage (late stage 5). To look at regulatory changes in maternal transcript deposition, F1 hybrid mothers were 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).

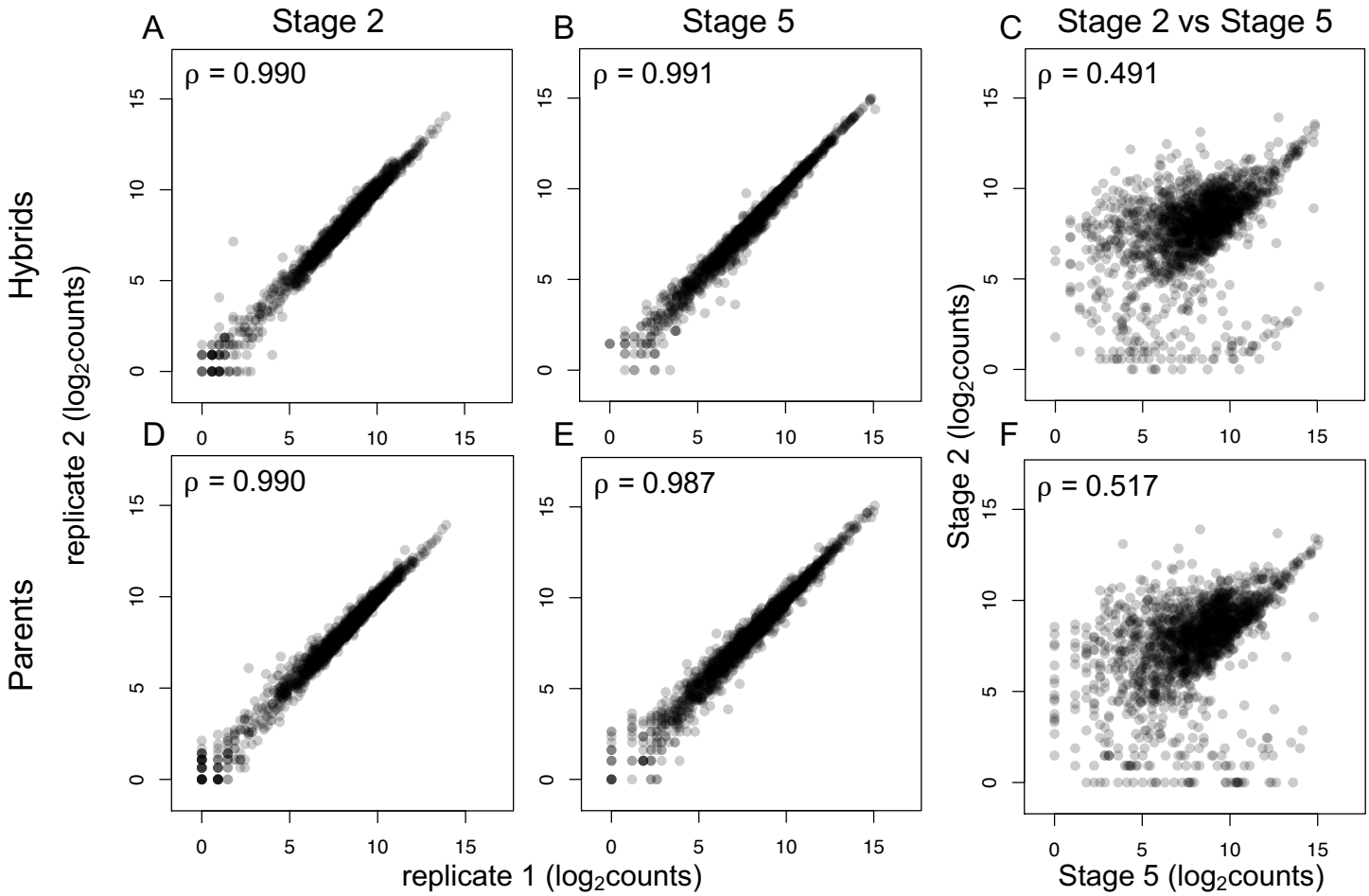


Figure 2: Hybrid and parental species single embryo transcript levels are highly reproducible. (A,B,D,E) Spearman's rank correlation coefficients are high when counts from replicate transcriptomes of the same stage and genotype are compared. Correlation coefficients are similarly high in parental species (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.

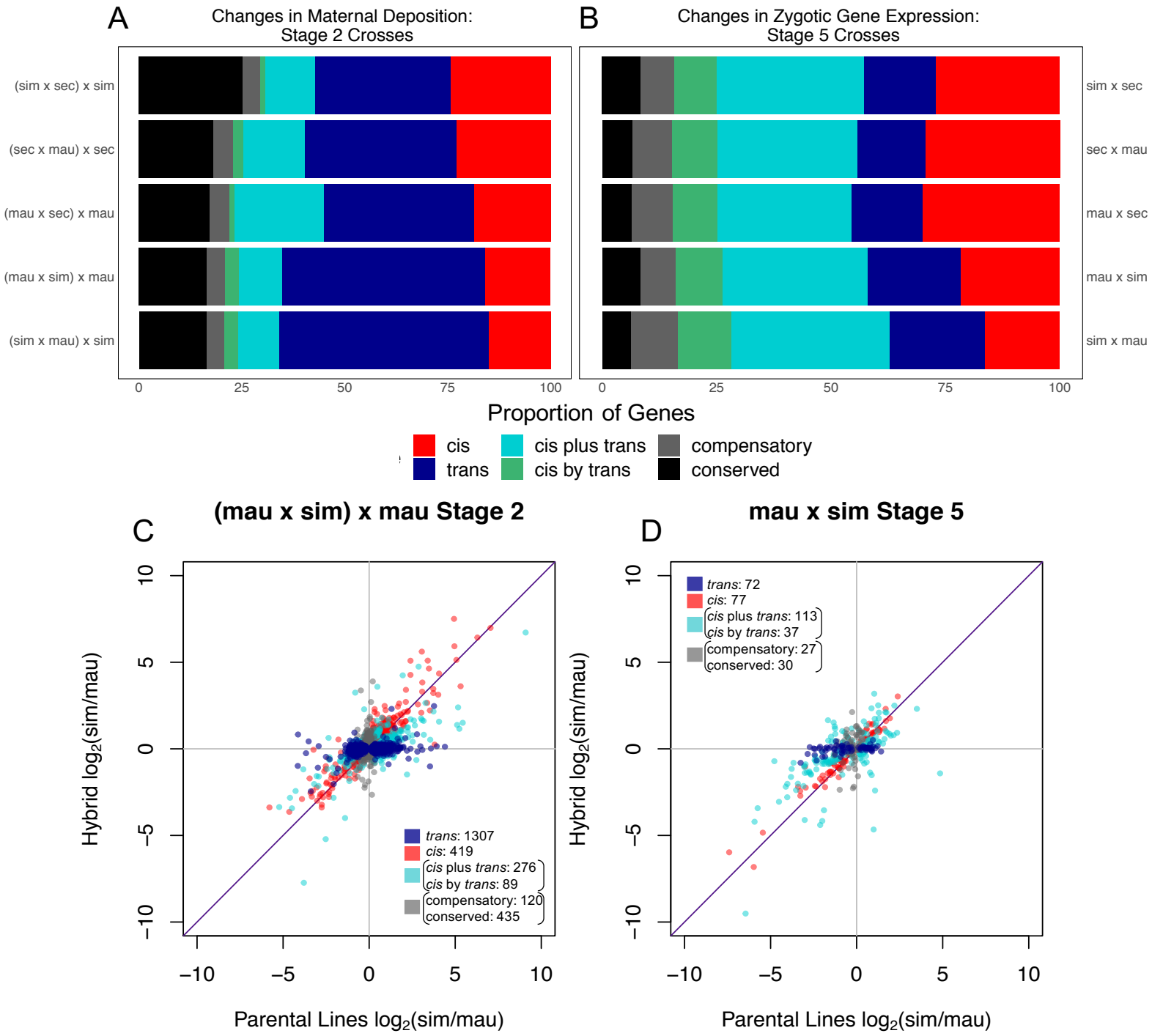


Figure 3: Different types of evolved regulatory changes dominate in maternal transcript deposition vs. zygotic transcription. Proportion of genes that fall into categories of regulatory change for each cross are shown for both the maternal transcript deposition (A) and zygotic gene transcription (B), 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 S2 and S3).

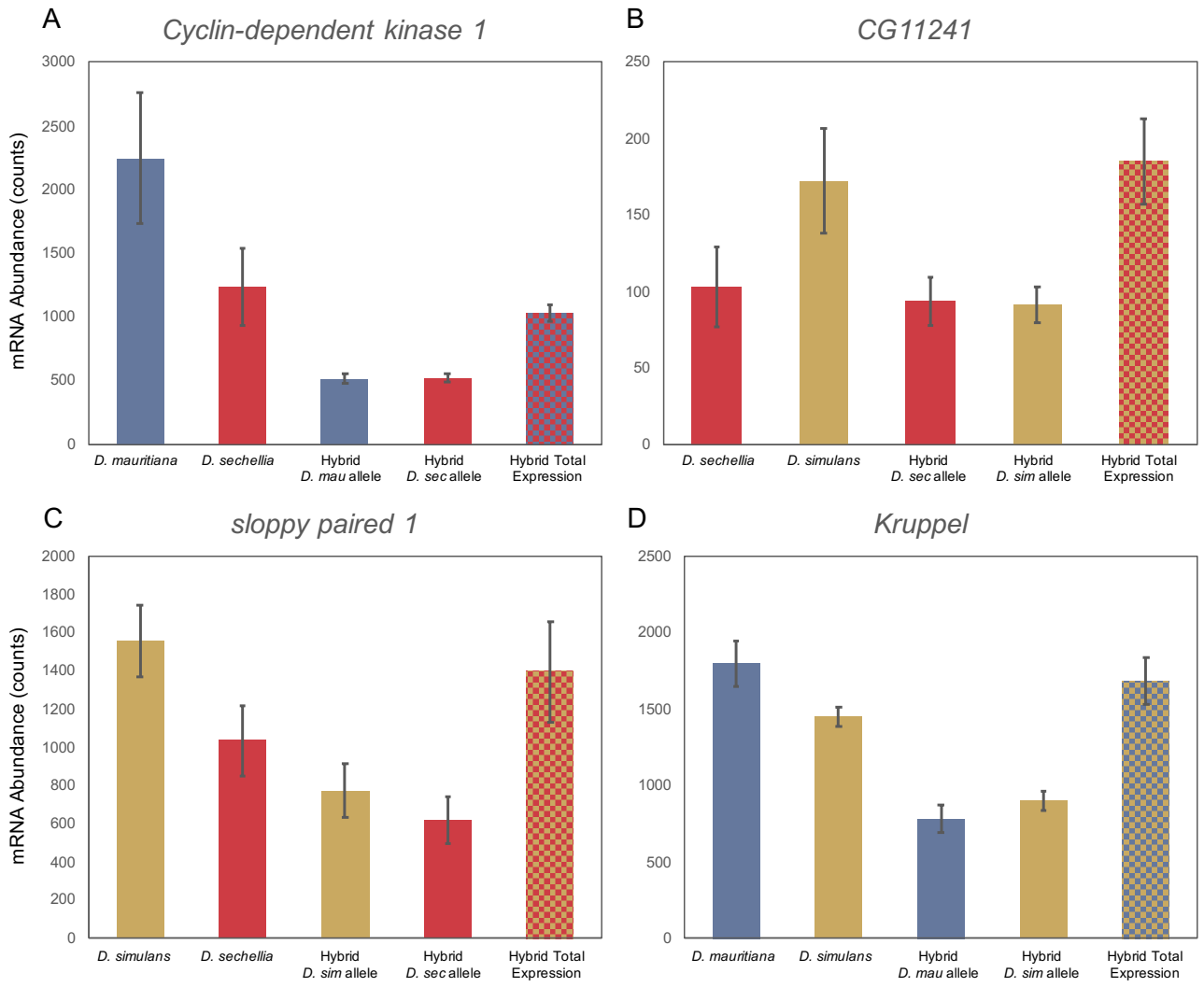


Figure 4: Examples of the type of regulatory changes observed, for individual genes. Transcript abundance, shown in counts, for each gene is plotted for both parental lines and for each parental allele within the hybrid; error bars shown represent the standard deviation. Total transcript abundance in 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* x 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* + *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* + *trans* regulation.

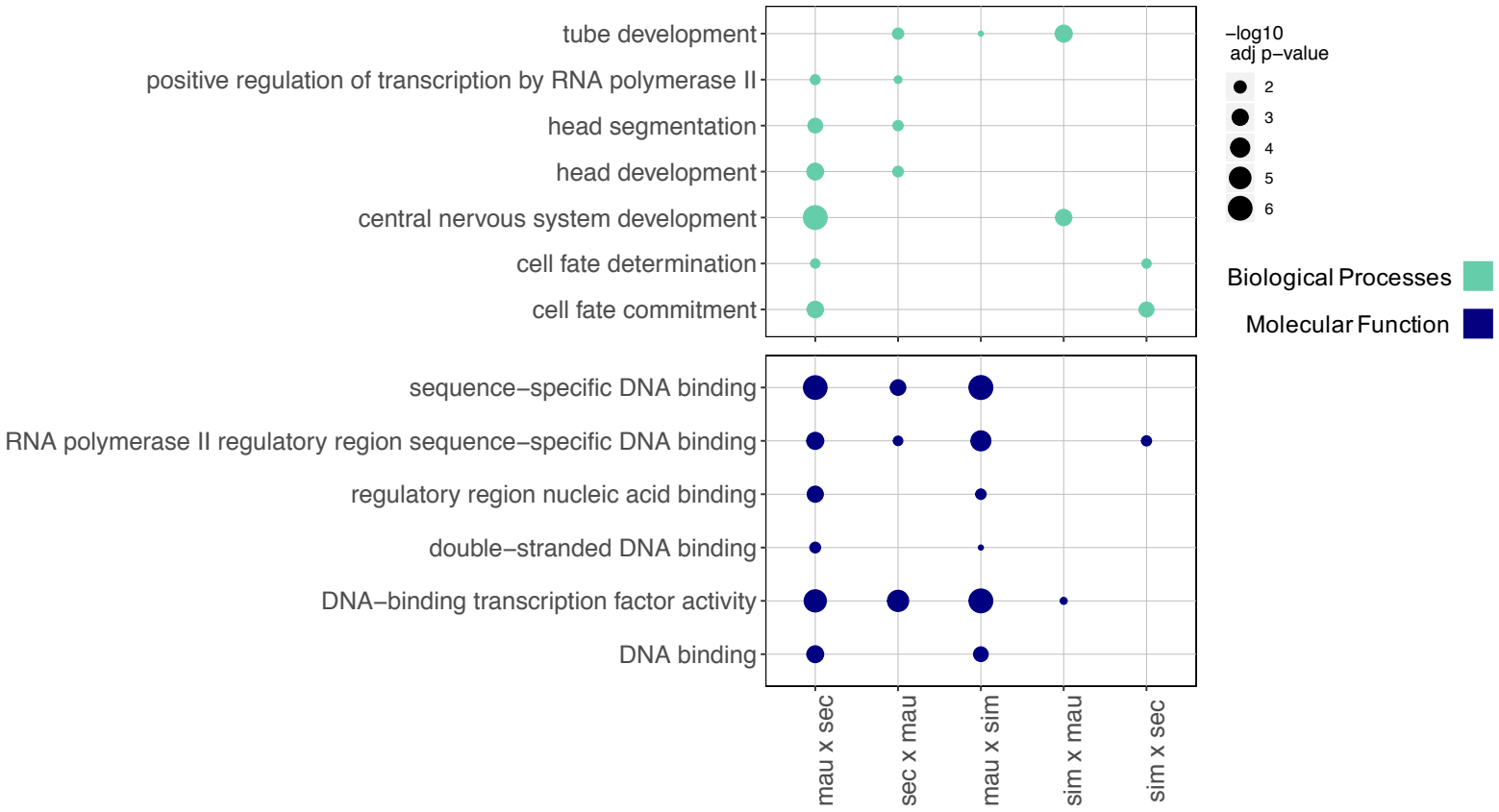


Figure 5: Gene ontology (GO) analysis identifies transcription factors that act in developmental processes as types of genes that change zygotically. Significantly enriched GO terms are listed for zygotically transcribed genes that change in *trans* regulation between each pair of species compared. Genes 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 S5. Biological process categories identified relate to development, molecular function categories identify functions consistent with DNA binding and regulation.

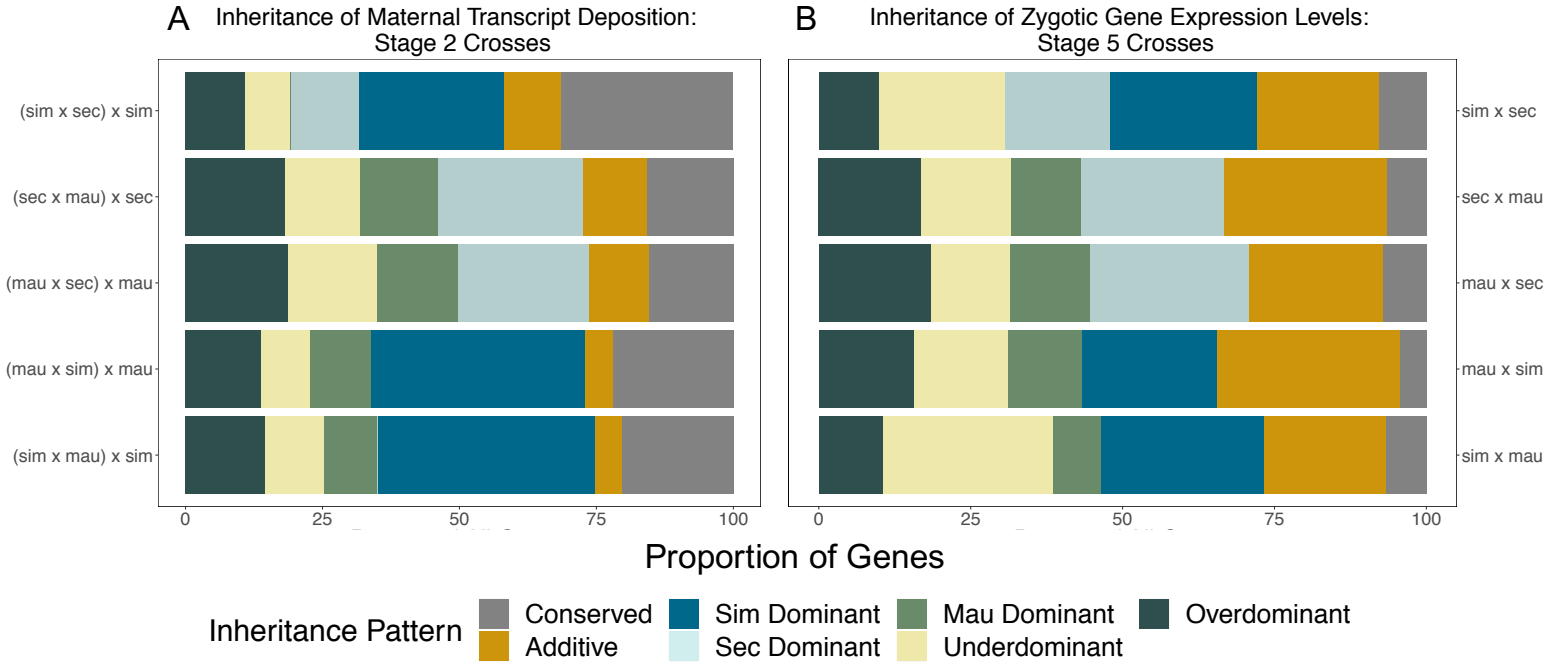


Figure 6: Patterns of inheritance show dominance of particular parental genomes at both stages. A) Shows patterns of inheritance for stage 2, over all genes and all crosses. B) Shows patterns of inheritance for stage 5, for mostly zygotic genes (see Methods) and all crosses. The maternal stage (A) shows 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).

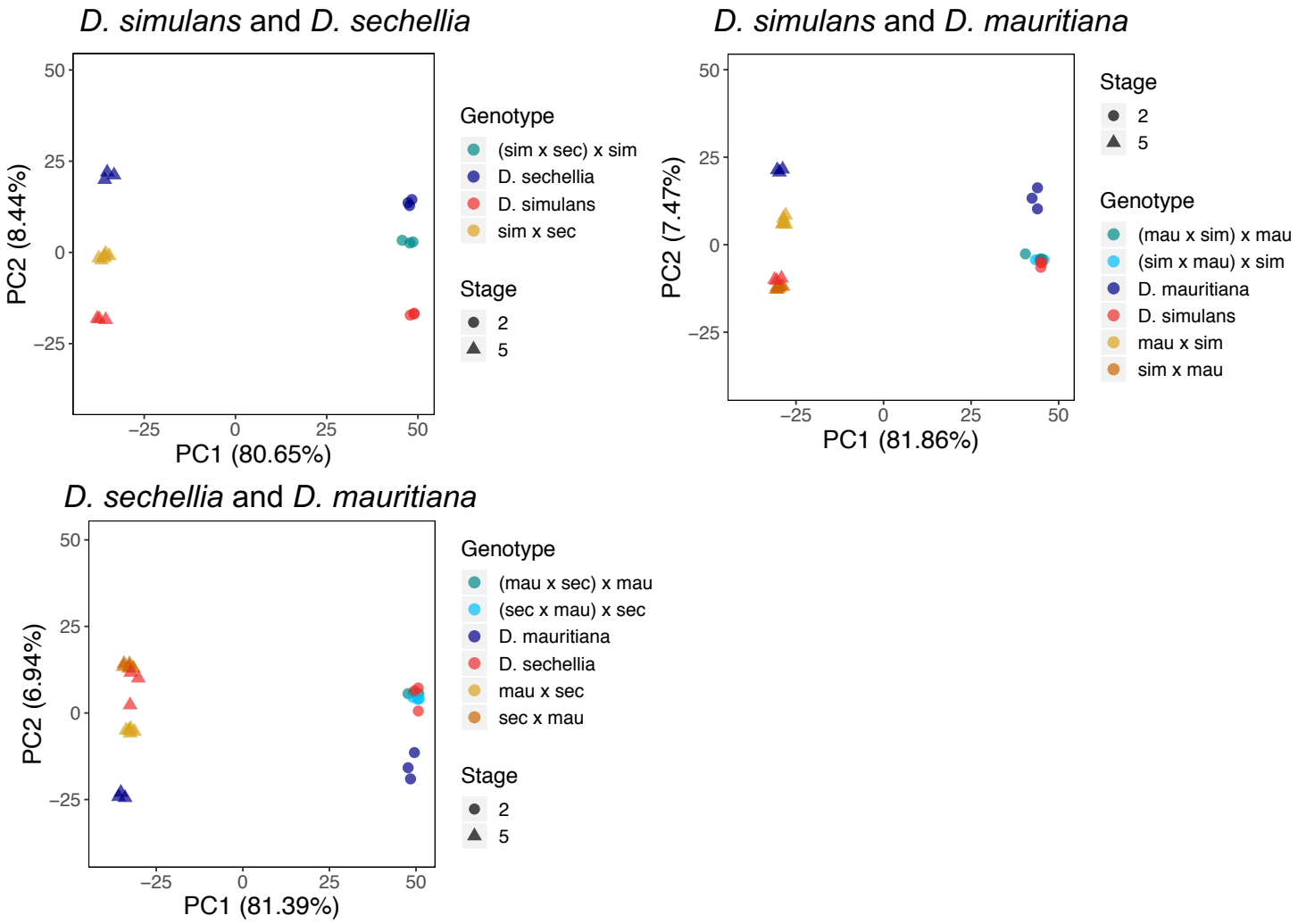


Figure S1: PCA plots for transcript abundance in all crosses. Samples of each stage, 2 or 5, cluster together. Samples of each genotype also cluster together, parental samples and hybrids. Proportion of variance explained by each principal component is listed on each axis.

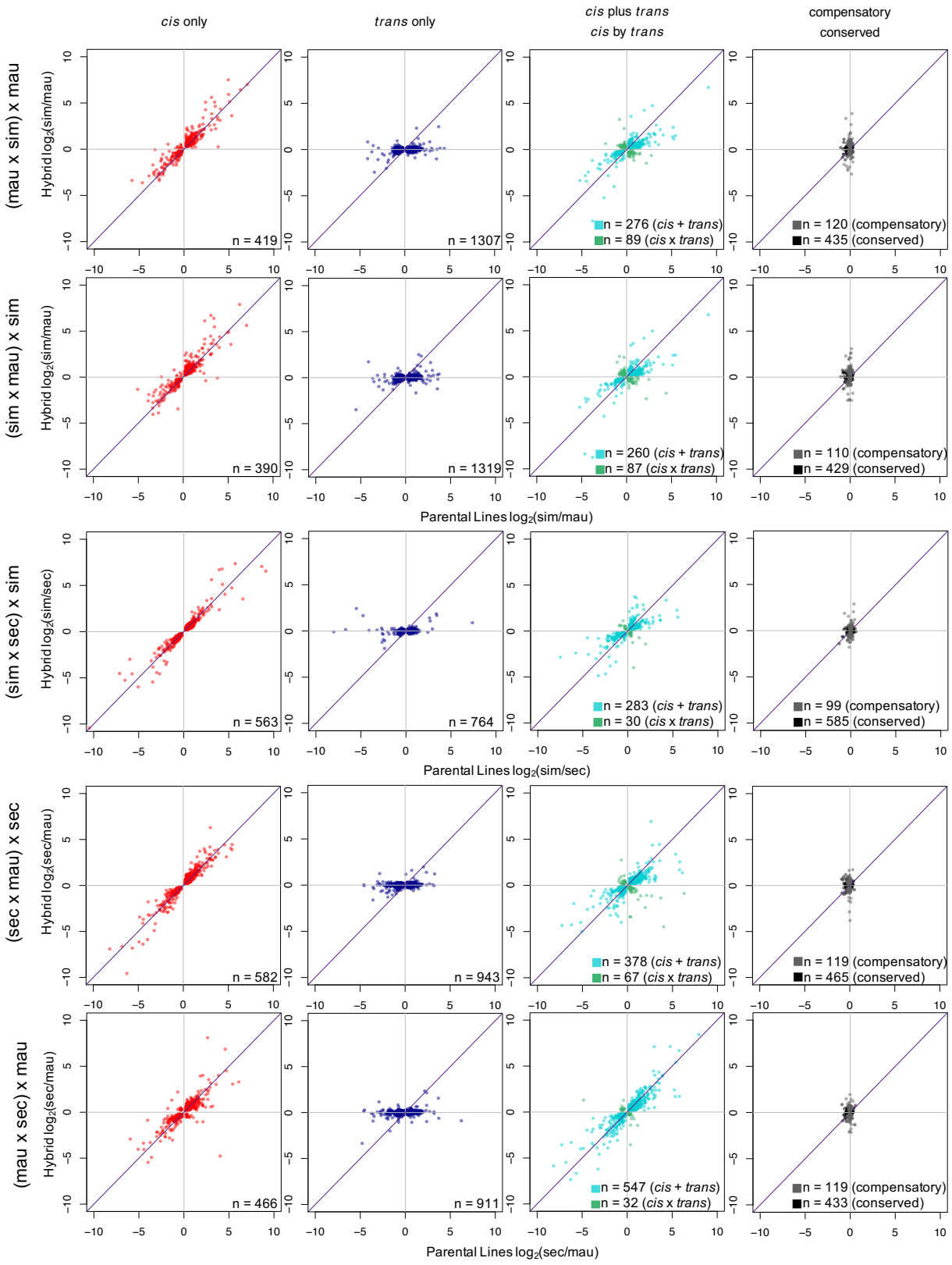


Figure S2: Regulatory changes in all pairwise comparisons for maternally deposited transcripts. Transcript level ratios between parental lines and within hybrids at stage 2 describe regulatory changes between species in each set of crosses. The number of genes in each category of regulatory change (n=) is listed in each plot. For definitions of categories of changes and criteria, see Methods.

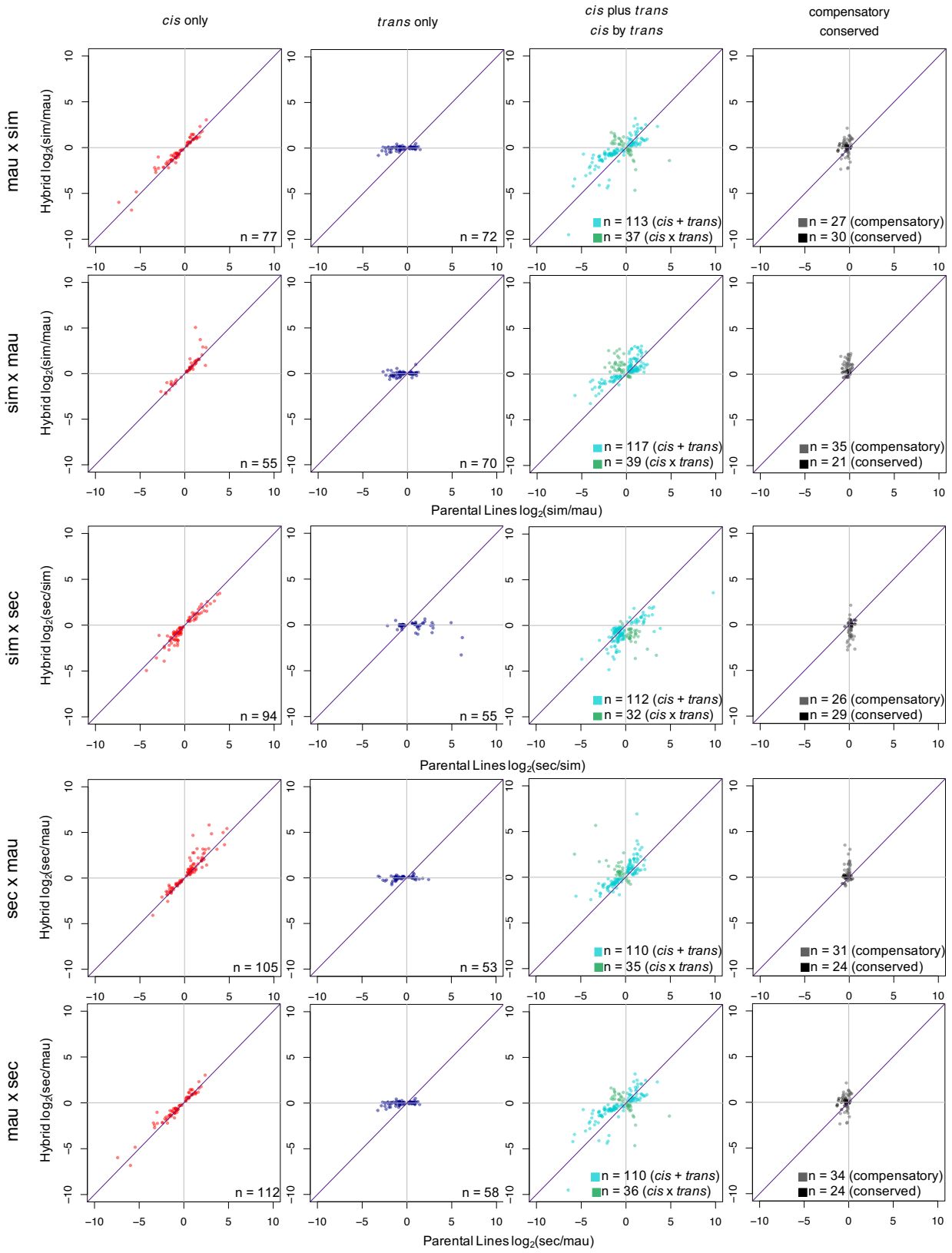


Figure S3: Regulatory changes in all pairwise comparisons for mostly zygotic genes. Transcript level ratios between parental lines and within hybrids at for mostly zygotic genes (see Methods) at stage 5 describe regulatory changes between species in each set of crosses. The number of genes in each category of regulatory change (n=) is listed in each plot. For definitions of categories of changes and criteria, see Methods.

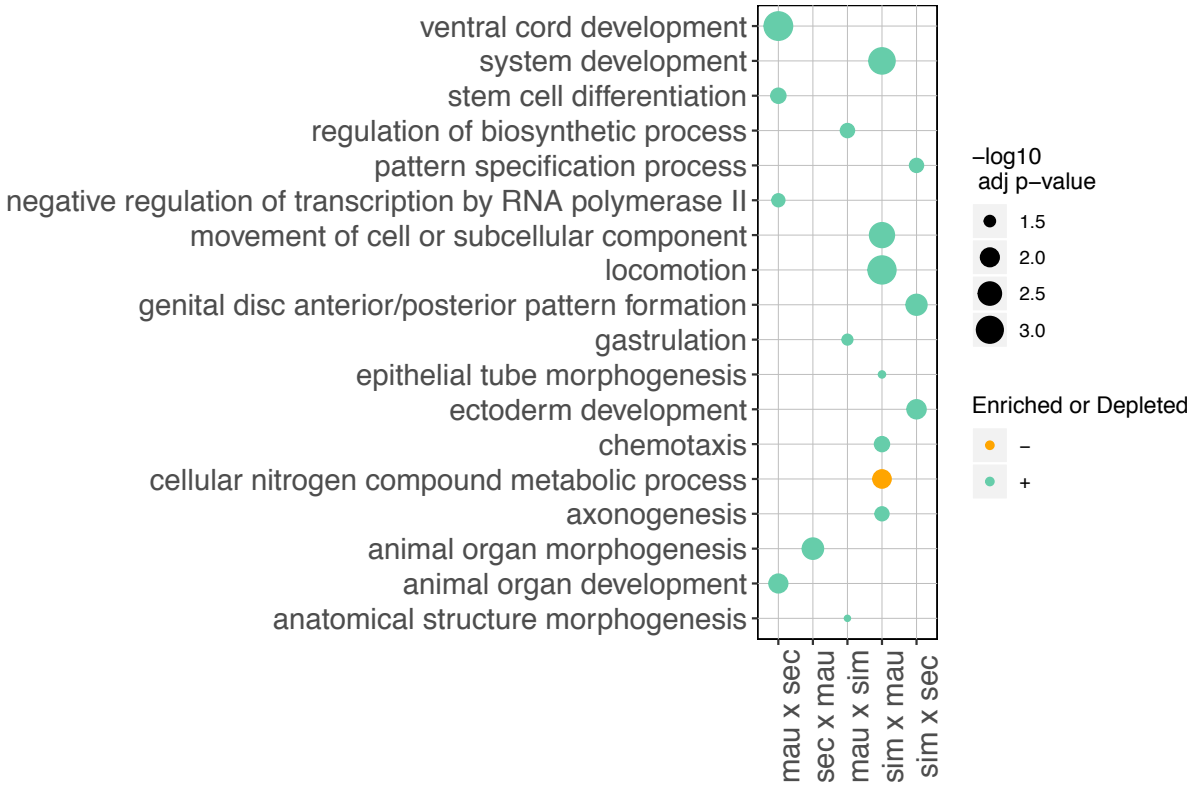


Figure S5: Gene ontology (GO) analysis for categories unique to a specific cross show enrichment for specific developmental processes. Significantly enriched GO terms are listed for zygotically transcribed genes that change in *trans* regulation between each pair of species compared. Again, zygotically 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.

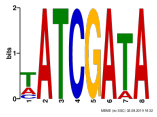
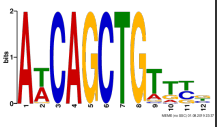
Motif	Predicted Binding Proteins	Enriched in upstream regions of *D. sechellia* genes that are conserved in expression (E-value)	Enriched in upstream regions of *D. sechellia* genes that are changing in expression (E-value)	Enriched in upstream regions of *D. simulans* genes that are conserved in expression (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 secxmau and simxsec 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 S2: 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.

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

Table S3: 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.