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

Data availability: The raw sequencing reads and processed data are available at GEO/NCBI at accession number (submitted, will add when available).

Running title: Regulatory evolution in embryogenesis

Keywords: gene expression regulation, maternal mRNA deposition, zygotic transcription, *cis* and *trans* regulation

Corresponding Author Information:

Emily L. Cartwright

Department of Evolution and Ecology

2320 Storer Hall

University of California, Davis

Davis, CA 95616

(530)-752-1112

elcartwright@ucdavis.edu

Susan E. Lott

Department of Evolution and Ecology

2320 Storer Hall

University of California, Davis

Davis, CA 95616

(530) 752-2939

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. The maternal to zygotic transition 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.

INTRODUCTION

Many critical early developmental processes are common across all Metazoans, including axial patterning and rapid cleavage cycles. The earliest of these conserved processes are regulated by maternally provided RNA and proteins, 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. 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), which is beneficial as the zygotic genome is not yet transcriptionally active. 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).

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 *et al.* 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 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) 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 within or 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 *et al.* 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 *et al.* 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 transcript levels between species at the maternal stage as compared to the zygotic stage. The species used in this study are very closely related and thus 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 conserved in the hybrids relative to the parental lines. Additionally, at both stages, 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. Overall, we find distinct patterns of gene regulatory changes at the two embryonic timepoints, before and after ZGA, indicating changes in gene regulation differ based on the developmental context.

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 (Bownes 1975; Campos-Ortega and Hartenstein 2013)). 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 (Bownes 1975; Campos-Ortega and Hartenstein 2013)), 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 (Bownes 1975; Campos-Ortega and Hartenstein 2013).

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 S1) 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.

208 **Library preparation and sequencing**

209 The RNA from single embryos was treated with DNase (TurboDNA-free, Life
210 AM1907) using manufacturer instructions and RNA sequencing libraries were
211 constructed with Illumina TruSeq v2 kits following the manufacturer low sample
212 protocol. The Illumina protocol uses oligo (dT) beads to enrich for polyadenylated
213 transcripts. Because poly(A) tail length is important in determining many post-
214 transcriptional processes during early development, including translational efficiency, it
215 is important to ensure that the method used for mRNA selection does not produce a
216 biased set of poly(A) tail lengths. Previous datasets report poly(A) length distributions
217 for transcripts during oogenesis and early development (Lim *et al.* 2016; Eichhorn *et al.*
218 2016). We could not directly compare our data to previous reports, as these studies
219 were done using *D. melanogaster*, which may have a different poly(A) tail length
220 distribution than the species used in our analysis. However, previous studies comparing
221 distributions of poly(A) tail lengths of all genes to poly(A) tail lengths of transcripts
222 recovered through poly(A) selection in *D. melanogaster* have demonstrated that poly(A)
223 selection with commonly used methods does not bias which transcripts are recovered
224 from the total pool of transcripts present (Eichhorn *et al.* 2016). This includes studies
225 that used the same single embryo approaches utilized here (Crofton *et al.* 2018; Atallah
226 and Lott 2018). Therefore, it seems unlikely that poly(A) selection heavily biases the
227 extracted RNA relative to the RNAs present at these developmental stages. cDNA
228 libraries were quantified using dsDNA BR Assay Kits (Qubit) and the quality of the
229 libraries were assessed on High Sensitivity DNA chips on an Agilent Bioanalyzer. The
230 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 (Yang 2017). 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, (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.

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. 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

277 (v2017) and an updated set of annotations from Torres-Oliva, *et al*. The orthologs table
278 (Table S2) was used to compare genes between each species and in each direction of
279 mapping.

280 Mapping bias due to differing genome quality may occur when using two different
281 reference genomes. In order to alleviate mapping bias that may occur when mapping
282 the parental and hybrid samples to each parental reference genome, Poisson Gamma
283 models (León-Novelo *et al.* 2014) were employed to calculate mapping bias for every
284 set of mappings, in each pairwise comparison of species. We found that between 9.6%
285 and 10.9% of genes expressed at stage 2 (with a count > 5) mapped in a biased way to
286 both parental genomes when compared to the total number of orthologous genes
287 between any pair of species. Each between-species comparison has a different number
288 of orthologous genes so the proportion of biased genes varies based on the pair of
289 species compared in a cross. In contrast to the maternal stage, we found that between
290 5.0% and 6.2% of genes expressed at stage 5 mapped in a biased way to parental
291 genomes when compared to the total number of orthologous genes between a pair of
292 species. Overall, when looking at the total proportion of biased genes, not just those
293 that were called “expressed” in our analysis, we found that between 24.8% and 28.3%
294 of genes at stage 2 and between 20.0% and 21.2% of genes at stage 5 mapped in a
295 biased manner when compared to the total number of orthologous genes in any
296 comparison between species in a cross.
297
298
299

300 **Genes used for stage 5 analysis**

301 To focus on the gene regulation from the zygotic genome after ZGA, we removed
302 genes with high levels of maternal transcript deposition from our analysis. We limited
303 the pool of genes analyzed to those that are mostly-zygotic because roughly half of
304 maternal transcripts are not entirely degraded by stage 5 (although studies are
305 somewhat variable in the percent reported) (Tadros *et al.* 2007; De Renzis *et al.* 2007;
306 Thomsen *et al.* 2010; Lott *et al.* 2014) and we wanted to examine only those genes that
307 have a larger contribution to expression from the zygotic genome. For this analysis, we
308 included genes with "zygotic-only" expression (those that are not maternally deposited)
309 and genes that are "mostly zygotic" (those with 8-fold higher expression at stage 5
310 relative to stage 2, a $\log_2$ difference greater than three). We tested several cutoffs but
311 chose the 8-fold threshold because at this conservative cutoff, most genes with high
312 maternal transcript deposition are removed from the analysis. Additionally, for this
313 analysis we used confidence intervals and averages generated from only female
314 samples for genes on the X chromosome because dosage compensation is not
315 complete at stage 5 (Lott *et al*. 2011).
316
317 **Correlation analysis and PCA**

318 We performed correlation analysis (Figure 2, Table S3) between single embryos
319 across replicates, stages, and genotypes in R (R Core Team 2017) using the Spearman
320 option within the *corr* function. PCA analysis was also performed in R using the *prcomp*
321 function (Figure S1).
322

323 ***Cis/Trans* analysis**

324 To address both mapping bias and allelic imbalance, we used 95% confidence
325 intervals (CIs) from Poisson Gamma (PG) models (León-Novelo *et al.* 2014). We used
326 the PG model (with fixed bias parameter, $q = 0.5$) to define differentially expressed
327 genes as genes with CIs that did not overlap the range of 0.49 - 0.51 when comparing
328 the expression levels of parental alleles in each replicate. We set a slightly more
329 conservative standard for classifying allelic imbalance where genes with CIs below 0.49
330 or above 0.51 were called differentially expressed. Genes with CIs close to 0.50 did not
331 appear differentially expressed when looking at the count data, so we used a more
332 conservative cutoff. Genes that appeared differentially expressed in one direction of
333 mapping but not in the other direction of mapping were removed from the analysis, as
334 this was determined to be a result of mapping bias between the two genomes. We also
335 removed genes that had disparate confidence intervals in the two mapping directions
336 (i.e. one mapping direction yielded a CI that fell above 0.5 and the other direction of
337 mapping yielded a CI that fell below 0.5).

338 The genes retained for analysis were categorized using *cis*, *trans*, *cis* + *trans*, *cis*
339 *x trans*, *compensatory*, and *conserved* categories as described in Coolon, *et al*. 2014;
340 McManus, *et al*. 2010; and Landry, *et al*. 2004 (Figures 3, S2 and S3). We assigned the
341 following categories for regulatory change based on the CIs generated from PG models
342 for individual genes (see Figure 4 for individual examples):

343

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

346 the hybrids. (CIs for parental species and hybrids overlap each other for changes
347 purely in *cis*. To determine this, we used the CIs generated from mapping to the
348 *D. simulans* genome for *D. simulans/D. mauritiana* and *D. simulans/ D. sechellia*
349 comparisons and CIs generated from mapping to the *D. sechellia* genome for *D.*
350 *sechellia/D. simulans* comparisons.)
351
352 *trans:* Genes that are differentially expressed between the parental species (CI
353 does not overlap 49% - 51%) but are not differentially expressed in the hybrid (CI
354 overlaps 49% - 51%).
355
356 *cis* + *trans*: Genes that are differentially expressed in the hybrids and between
357 the parental species (CI does not overlap 0.49% - 0.51%) and the CI is in the
358 same direction for both the parents and the hybrid (i.e. both are greater than 51%
359 but the CIs for the parents and hybrid do not overlap. For this comparison, we
360 used the CIs generated from mapping to the *D. simulans* genome for *D.*
361 *simulans/D. mauritiana* and *D. simulans/ D. sechellia* comparisons and CIs
362 generated from mapping to the *D. sechellia* genome for *D. sechellia/D. simulans*
363 comparisons.)
364
365 *cis* x *trans*: Genes that are differentially expressed in the hybrids and between
366 the parental species (CI does not overlap 49% - 51%) and the CI is in opposite
367 directions for the parents and the hybrid (i.e. one is greater than 51%, the other is
368 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.

392 **Candidate transcription factor identification**

393 We took a computational approach to identify potential transcription factors that
394 may change in trans regulation between the species in our analysis. We used motif
395 enrichment programs to find potential binding sites in the upstream regions of genes
396 changing in regulation in *D. sechellia* and *D. simulans*. We omitted *D. mauritiana* from
397 this analysis because the *D. mauritiana* genome is not as well annotated as the
398 genomes for *D. simulans* and *D. sechellia*. We used the Differential Enrichment mode in
399 MEME (Bailey and Elkan 1994) as well as the findMotifs.pl script in HOMER (Heinz *et*
400 *al*. 2010) to identify overrepresented motifs in the regions 500bp upstream of genes
401 changing in regulation or with conserved regulation between species in every set of
402 comparisons at stage 2. In MEME, we used options to find motifs with any number of
403 repetitions and a motif width of 8-12. We used default options for HOMER and supplied
404 a background fasta file for enrichment analysis. The background lists supplied were
405 500bp upstream regions from all annotated genes in the species except for those that
406 were in the target set (either those genes with conserved or changing regulation in any
407 set of comparisons). The 500bp regions were extracted from fasta files (versions were
408 the same as ones used for mapping) for each species using BEDTools (Quinlan and
409 Hall 2010). Significantly represented motifs in the target lists relative to the background
410 supplied were then compared against databases of known transcription factor binding
411 sites using Tomtom (MEME suite) and HOMER. All enriched motifs that appeared in
412 both HOMER and MEME analyses are included in Table S4. All potential targets of
413 discovered motifs with significant E-values (MEME) or high Match Rank scores in

HOMER (>0.8) are also listed in Table S4 (see Figure S4 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, 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 and GO categories unique to a cross are shown in Figure S5. All enriched categories are listed in Table S5.

Data Availability

All sequencing data and processed data files from this study are available at NCBI/GEO at accession number: *submitted, awaiting accession number*.

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; Bownes 1975; Campos-Ortega and Hartenstein 2013). 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 hybrid 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.* 2011; 2014; 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 S3). 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 S3). Stage 5 hybrids from the *mau* x *sim* cross have equally high correlation coefficients, ranging from 0.980 to 0.996 (Table S3). 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 S3), 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 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 samples of different stages yields strikingly lower correlation coefficients (Figure 2, C,F, Table S3), emphasizing the turnover of transcripts between these stages. 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 S3). 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

506 highly correlated by their Spearman's rank correlation coefficients but stage-matched
507 embryos separate out by genotype in principal component analysis (PCA; Figure S1) by
508 the second principal component. The second principal component of this PCA accounts
509 for between 6.94-8.44% of the variance in the three sets of comparisons (the first
510 principal component corresponds to developmental stage and explains between 80.65%
511 and 81.86% of the variance). Although samples of the same stage have similar
512 transcript abundances, as evidenced by correlation coefficients, they are most similar to
513 samples of the same genotype, as seen through PCA.
514
515 **Regulatory changes at the maternal stage of development**
516 Changes in gene expression can occur through many mechanisms: alterations in
517 chromatin state, differences in *cis* or in *trans* regulation, or through post-transcriptional
518 modifications. Here, we examine how *cis* and *trans* regulation evolves to differentially
519 affect transcript levels during the maternal to zygotic transition, across species of
520 *Drosophila*. Changes in *cis* regulation occur through changes in the DNA of regulatory
521 regions proximal to the gene that they regulate. These types of regulatory changes have
522 an allele-specific effect on gene expression. In contrast, changes in *trans* regulation
523 occur via changes in factors that bind to the DNA, such as transcription factor proteins.
524 Changes in *trans* regulation affect gene expression independent of allele-specificity.
525 In order to determine regulatory changes in *cis* and in *trans* that lead to
526 differences in maternal transcript deposition between *D. simulans*, *D. sechellia*, and *D.*
527 *mauritiana*, we used Poisson Gamma (PG) models (León-Novelo *et al.* 2014) to
528 determine allelic imbalance between parental lines and within hybrid embryos. Reads

529 from each sample were aligned to both genomes of the parental species used in the
530 cross. In these alignments, we identified genomic sites with fixed differences between
531 lines of species used to determine the parental species of origin for each read (see
532 Methods). We used the results from both directions of mapping, meaning mapping to
533 both parental genomes, to determine mapping bias and allelic imbalance. When reads
534 differentially mapped to the two parental genomes, we identified these genes as biased
535 in their mappings and removed them from the analysis (see Methods). Genes that were
536 not biased in their mapping and had a read count of >5 reads were retained for analysis.
537 Transcript levels between each parental species were compared using PG
538 models to identify differential maternal transcript deposition between the parental lines.
539 Similarly, transcript levels of species-specific alleles in the stage 2 embryos produced
540 by hybrid mothers were compared to determine differential deposition of maternal
541 transcripts into the embryo. We compared the two sets of analyses to determine the
542 regulatory basis of differential maternal transcript deposition between species. We used
543 the logic of Landry, *et al*. 2005 to classify genes as having changed in *cis* or in *trans*
544 regulation, based on comparisons made with confidence intervals of the bias
545 parameters generated through the PG models (see Methods).
546 We find that most regulatory changes underlying differentially maternally
547 deposited transcripts occurred in *trans* between each pair of species examined (Figure
548 3A and C, Figure S2), where a change in a transcription factor or other *trans*-acting
549 regulatory element affects both alleles equally (shown in Figure 4A and B). In all pairs of
550 comparisons, the proportion of *trans* changes was higher than any other category of
551 changes. Interestingly, comparisons between *D. simulans* and *D. mauritiana* had the

552 highest percentage of *trans*-only regulatory changes (between 49.4% and 50.8%) while
553 comparisons between *D. simulans* and *D. sechellia* had a lower percentage of
554 regulatory changes solely in *trans* (32.9%). The second highest proportion (between
555 15.0% and 26.7%) of regulatory changes between species at the maternal stage
556 occurred only in *cis* regulation. Slightly fewer regulatory changes occur due to a
557 combination of *cis* and *trans* acting factors (between 13.4% and 15.8% in all
558 comparisons). Most genes that change in *cis* and *trans* are assigned to the *cis* + *trans*
559 category, which indicates that the allele with higher expression in the parental lines is
560 also preserved as the allele with higher expression in the hybrid. We also find a
561 percentage of genes with conserved levels of maternal transcript deposition between
562 species, between 16.4% and 25.1% in all crosses. *D. simulans* and *D. sechellia* have
563 the highest percentage of conserved genes while *D. simulans* and *D. mauritiana* have
564 the lowest percentage of conserved genes. We also find a small proportion of genes
565 (between 4.2% and 4.7% in all comparisons) that have evolved compensatory
566 mechanisms of regulation, where the genes are not differentially expressed between the
567 parental samples but are differentially expressed in hybrids. This implies that while
568 transcript levels are the same between species, regulatory changes have occurred,
569 which then become visible in the environment of the hybrid. Genes with representative
570 *trans* changes include regulators with critical functions in important processes governed
571 by maternal gene products, such as *Cdk1*, a cell-cycle regulator necessary for the rapid
572 cleavage cycles in early development(Farrell and O'Farrell 2014). Examples of
573 individual genes with changes attributed to *trans* regulatory differences at the maternal
574 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 (Bailey and Elkan 1994; Heinz *et al.* 2010) 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), Table S4. The Dref/BEAF-32 binding site (BEAF-32 and Dref bind overlapping DNA sequences, Hart *et al.* 1999) is the most significantly enriched (Table S4), and these factors are annotated as insulators (Matzat and Lei 2014; Ali *et al.* 2016) 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 S4). M1BP is involved in transcriptional regulation and RNA polymerase II pausing at the promoter of genes (Li

598 and Gilmour 2013), which may also be associated with chromatin state (Ramírez *et al.*
599 2018). Transcript abundance data from our study indicates that *Dref*, *BEAF-32*, and
600 *M1BP* are differentially maternally deposited in several between-species comparisons
601 (Figure S4), although in certain crosses, hybrid reads mapped in a biased way to *Dref*,
602 *BEAF-32* and *M1BP*, and thus they were excluded from our regulatory analysis. As the
603 motifs for these *trans*-acting factors are enriched in all maternal genes, they are likely
604 important regulators of transcription during oogenesis, and therefore also likely targets
605 of regulatory evolution between species.
606
607 **Evolution of regulation for zygotically expressed genes**
608 To determine the regulatory basis of changes in zygotic transcript abundance
609 between species, we compared expression levels in late stage 5 parental species
610 samples to late stage 5 hybrid samples and used PG models to identify *cis* and *trans*
611 regulatory changes, similar to our maternal analysis (see Methods). We limited our
612 analysis at the zygotic stage to those genes that are mostly-zygotic: zygotically
613 expressed but not maternally deposited (zygotic-only) or expressed at the zygotic stage
614 at an 8-fold higher level when compared to the maternal stage (we will refer to these as
615 mostly zygotic genes, see Methods).
616 While we find that most genes change in *trans* regulation between *D. simulans*
617 and its sister species at the maternal stage of development, we see different patterns of
618 regulatory changes after ZGA. At the zygotic stage, differences in gene regulation
619 between the three species examined occur mostly in both *cis* and *trans*, where changes
620 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 (Landry *et al.* 2005; McManus *et al.* 2010; Coolon *et al.* 2014), 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 46.7% 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 35.0% of all genes) as opposed to a *cis* x *trans* pattern (between 9.1% and 11.7% 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 21.0% of genes changing only in *trans* regulation. There are between 16.2% 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 7.4% and 9.9% of all genes) in gene regulation and a smaller proportion of genes that are conserved (between 6.3% and 8.3% 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

644 maternal transcripts to be more highly conserved between species than zygotic
645 transcripts (Atallah and Lott 2018). Examples of evolved changes include regulators
646 critical to important early zygotic processes, such as gap gene *Kruppel* and pair-rule
647 gene *Sloppy paired 1*, which are required for segmentation along the anterior-posterior
648 axis (Nüsslein-Volhard and Wieschaus 1980; Grossniklaus *et al.* 1992) (Figure 4C and
649 D).

650 In contrast to what we have found for regulation at the maternal stage, where
651 transcription may be broadly determined by chromatin boundaries, regulation at the
652 zygotic stage can be gene or pathway specific and involve transcription only in a
653 spatially localized subset of cells (Jäckle *et al.* 1986; St Johnston and Nüsslein-Volhard
654 1992). As such, if a *trans* regulator changed at the zygotic stage, it may affect genes
655 involved in a specific developmental process. For these reasons, we wanted to ask if
656 changes in zygotic *trans* regulation affected particular types or categories of genes. We
657 used PANTHER (Mi *et al.* 2019) to perform gene ontology (GO) analysis on genes
658 changing only in *trans* regulation in each pairwise comparison of species at stage 5 (see
659 Methods). The significantly enriched GO categories for molecular function and biological
660 process categories that are shared amongst two or more crosses are shown in Figure 5
661 and Table S5. Identifying GO categories over multiple crosses identifies the types of
662 genes that evolve changes repeatedly over evolution. Shared categories include those
663 related to DNA binding, positive regulation of transcription by RNA polymerase II, cell
664 fate determination, and several developmental categories. We also find biological
665 process categories unique to a specific cross (Figure S5 and Table 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 (Michalak 2003; Moehring *et al.* 2007; Mack *et al.* 2016). 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.* 2014), 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

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

691 Strikingly, while many genes show conservation of expression levels between
692 parental species and in the hybrids at both developmental stages, we find a much
693 higher percentage of conserved transcript levels between parents and hybrids for genes
694 that are maternally deposited (Figure 6). We find a high proportion of genes with
695 conserved transcript levels at stage 2 in all crosses, between 15.4% and 31.4% of all
696 genes. In contrast, in stage 5 crosses we find conserved transcript abundance in
697 between 4.3% and 7.8% of all genes that are either zygotic-only or are mostly zygotic
698 (see Methods for definitions). While there is a large difference in the percentage of
699 conserved genes between the two stages, our stage 5 analysis is limited to those genes
700 with much higher expression at the zygotic stage in comparison to the maternal stage of
701 development. There may be more genes that are mostly zygotic or zygotic-only that are
702 misregulated at this stage in development relative to all of the genes that are expressed
703 at stage 5.

704 Of the genes that do not have conserved transcript levels in the stage 2 hybrids,
705 most genes show dominance to one parental species or the other (Figure 6). Few
706 genes show additive patterns of inheritance, in keeping with findings from other studies
707 (Gibson *et al.* 2004; McManus *et al.* 2010). We also report a fraction of misexpressed
708 genes as being over- or under- dominant in the hybrids in relation to expression levels
709 in the parental lines. Of those genes that are dominant at stage 2, a higher proportion
710 have *D. simulans*-like expression (in any cross involving *D. simulans*) in comparison to
711 the proportion that have expression more like the other parental line in the cross. In the

712 *D. simulans*/*D. mauritiana* stage 2 comparison, between 39.0% and 39.7% of all genes
713 had transcript levels that were more similar to the *D. simulans* stage 2 samples than the
714 *D. mauritiana* stage 2 samples. In contrast, between 11.2% and 12.4% of all genes in
715 the hybrid have transcript levels more like the *D. mauritiana* stage 2 samples in this
716 comparison. We find that *D. mauritiana* has the least dominance in any cross involving
717 this species. In the *D. sechellia*/*D.mauritiana* comparisions at stage 2, we find that 14.3-
718 14.8% of transcript levels in the hybrid look more like the *D. mauritiana* samples while
719 24.0-26.4% of transcript levels in the hybrid look more like the *D. sechellia* samples. In
720 the comparison with *D. simulans* and *D. sechellia*, more genes in the hybrid had
721 transcript levels similar to the *D. simulans* parental species (26.3%) than the *D.*
722 *sechellia* parental species (12.6%). Taken together, our findings indicate that *D.*
723 *simulans* has the most dominant effect on gene expression at the maternal stage in
724 development in this comparison, while *D. mauritiana* has the least dominant effect, with
725 dominance in *D. sechellia* falling between the other two species.
726 We find similar patterns of inheritance in comparisons of stage 5 samples with a
727 few notable exceptions (Figure 6). Of the mostly zygotic and zygotic-only genes that are
728 not conserved in the hybrids at stage 5, many show a similar pattern of dominant
729 expression where the hybrid expression is more similar to one parental species or the
730 other. As in the maternal stage (stage 2), crosses with *D. simulans* have a higher
731 percentage of genes with expression in the hybrid more similar to *D. simulans* than to
732 the other parental species in the cross. For example, in crosses between *D. simulans*
733 and *D. mauritiana*, between 22.3% and 26.8% of all genes have an expression level
734 similar to *D. simulans* while only between 7.9% and 12.1% of all genes have an

735 expression level more similar to *D. mauritiana*. In the *D.* simulans/*D. sechellia*
736 comparison, more genes have an expression level in the hybrid more similar to *D.*
737 *simulans* parental samples (24.2%) than to *D. sechellia* (17.2%). Similar to the stage 2
738 comparisons, *D. sechellia* exerts the second most dominant effect in the hybrids, with *D.*
739 *simulans* having the most dominant effect. In crosses with *D. sechellia* and *D.*
740 *mauritiana*, between 23.7% and 26.1% of genes show *D. sechellia* dominance while
741 only between 11.5% and 13.2% of genes show *D. mauritiana* dominance. Interestingly,
742 in crosses where *D. mauritiana* is the maternal species, there is a slightly higher
743 percentage of genes showing *D. mauritiana* dominance (12.1% versus 7.9% in the
744 crosses involving *D. simulans* and although less striking, 13.2% versus 11.5% in the
745 crosses involving *D. sechellia*). This is likely due to the interaction of residual maternal
746 factors from *D. mauritiana* that have not yet degraded by stage 5 with gene products
747 from alleles of the paternal species expressed after zygotic genome activation. In
748 contrast to stage 2, we find more genes that have an additive mode of inheritance or
749 that are misexpressed in the hybrids at stage 5 (between 27.9% and 45.7% of all genes
750 in stage 2 crosses vs. between 50.8% and 61.3% in stage 5 crosses). Previous studies
751 indicate that additive inheritance is associated with *cis* regulatory divergence (Lemos *et*
752 *al.* 2008; McManus *et al.* 2010). This is in keeping with our findings that a larger
753 proportion of genes at the zygotic stage have expression divergence due, in part or
754 wholly, to *cis* regulatory changes and that more zygotic genes show an additive pattern
755 of inheritance.
756
757

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 (Heyn *et al.* 2014; Atallah and Lott 2018) 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; De Renzis *et al.* 2007; Thomsen *et al.* 2010; Lott *et al.* 2014). 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 (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

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

783 In addition to finding many genes with conserved transcript abundance between
784 species, we find a large proportion of genes with dominant patterns of inheritance at the
785 maternal and the zygotic stages. At both developmental timepoints, *D. simulans* has the
786 most dominance, where the highest proportion of genes with dominant inheritance in
787 the hybrid is most similar in expression to the *D. simulans* parental line, in crosses
788 involving *D. simulans*. The second largest category of genes with dominant patterns of
789 inheritance in the hybrids has expression levels most similar to the *D. sechellia* parental
790 line. *D. mauritiana* exerts the least amount of dominance in terms of inheritance
791 patterns in the hybrids. The dominant patterns of inheritance are in keeping with the
792 high number of *trans* regulatory changes we see, as previous work identified *trans*
793 changes to more likely to be associated with dominance (Lemos *et al.* 2008). Other
794 work with these species also identified many genes with dominant patterns of
795 expression in hybrids relative to parental lines (Moehring *et al.* 2007; McManus *et al.*
796 2010; Coolon *et al.* 2014). We find more genes in hybrids with transcript levels more
797 closely matching those in *D. simulans* in comparison to *D. mauritiana*, similar to
798 previous work (Moehring *et al.* 2007). When comparing *D. simulans* and *D. sechellia*,
799 we see a higher proportion of genes in hybrids with total expression levels similar to *D.*
800 *simulans*. This is opposite of the pattern observed in a previous study, which found
801 equal levels of dominance from both *D. simulans* and *D. sechellia* in the hybrid
802 (Moehring *et al.* 2007). Our findings and those from previous studies may differ in the *D.*
803 *simulans* and *D. sechellia* comparison because different developmental stages were

804 examined in each study. Differences in the species compared and developmental
805 stages or tissue types examined across studies make it difficult to draw conclusions
806 from direct comparisons, but may indicate something fundamental about how gene
807 expression is regulated or evolves at different developmental stages or in different
808 tissues.

809 Patterns of gene regulatory changes differ greatly between the developmental
810 stages examined, even though the two stages have similar modes of dominant
811 inheritance. We find an overwhelming number of *trans* regulatory changes that result in
812 differential maternal transcript deposition between the species examined. The biological
813 and regulatory context differs between the two stages and may explain why patterns of
814 gene regulatory evolution are different at the developmental timepoints we examined.
815 Maternal transcripts are produced by support cells called nurse cells during oogenesis,
816 and either transported by actin-dependent mechanisms or dumped into the oocyte
817 along with the cytoplasmic contents of the nurse cells upon apoptosis of these cells
818 (Kugler and Lasko 2009). Many aspects of maternal provisioning have been well-
819 studied in *D. melanogaster*, including transport of transcripts into the oocyte (Mische *et*
820 *al.* 2007), localization of transcripts within the oocyte (Theurkauf and Hazelrigg 1998),
821 post-transcriptional regulation of maternal gene products (Salles *et al.* 1994), and
822 subsequent degradation of maternal transcripts (Tadros *et al.* 2007; Bushati *et al.* 2008;
823 Laver *et al.* 2015). Surprisingly, how transcription is regulated in the nurse cells is not
824 well understood. Nurse cells are polyploid and transcribe at a high level to provide the
825 oocyte with the large stock of transcripts needed (we extract >100ng of total RNA at
826 both stages), which is exceptional considering the oocyte begins as essentially a single

827 cell (though a highly specialized one). We have found here, and also through another
828 study using computational methods to look for binding motifs in maternal factors across
829 the *Drosophila* genus (Omura CS and Lott SE. The conserved regulatory basis of
830 mRNA contributions to the early embryo differs between the maternal and zygotic
831 genomes. *in prep.*; submitted as a supplemental file for initial submission), that maternal
832 transcription is associated with factors annotated to be insulators and that interact with
833 topologically associated domain boundaries. These two studies both provide evidence
834 that maternal transcription is being controlled broadly at the level of chromatin state. In
835 this context, a few *trans* regulators are predicted to be responsible for the bulk of
836 maternal transcription, and thus changes in the levels of *trans* regulators at this stage
837 may easily be responsible for changes in transcription level for a number of genes. How
838 changes in *trans* regulators affecting chromatin state can have subtle quantitative
839 effects on transcript level, as observed, seems less clear. Functional investigations of
840 mechanisms of regulatory control at the maternal stage are ongoing.
841 In contrast to the large proportion of regulatory changes in *trans* at the maternal
842 stage, differences in zygotic gene transcription between these species are
843 characterized by a combination of changes in *cis*, *trans*, *cis* + *trans*, and *cis* x *trans*.
844 Zygotic gene transcription is spatially and temporally regulated across the embryo with
845 enhancer regions playing a large role in where and when genes are expressed (Haines
846 and Eisen 2018). Changes in *trans* regulation, which can affect the expression of many
847 genes, may be detrimental to the developing organism at this stage, whereas changes
848 in *cis* regulation are gene-specific and may only affect gene expression in a subset of
849 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.

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.

872 **REFERENCES**

873 Ali T., Renkawitz R., Bartkuhn M., 2016 Insulators and domains of gene expression.
874 Current Opinion in Genetics & Development **37**: 17–26.

875 Atallah J., Lott S. E., 2018 Evolution of maternal and zygotic mRNA complements in the
876 early Drosophila embryo (KA Dyer, Ed.). PLoS Genet. **14**: e1007838.

877 Bailey T. L., Elkan C., 1994 Fitting a mixture model by expectation maximization to
878 discover motifs in bipolymers.

879 Barckmann B., Simonelig M., 2013 Control of maternal mRNA stability in germ cells and
880 early embryos. Biochimica et Biophysica Acta (BBA) - Gene Regulatory
881 Mechanisms **1829**: 714–724.

882 Baroux C., Autran D., Gillmor C. S., Grimanelli D., Grossniklaus U., 2008 The Maternal
883 to Zygotic Transition in Animals and Plants. Cold Spring Harb Symp Quant Biol **73**:
884 89–100.

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

887 Buchberger, Reis, Lu, Posnien, 2019 Cloudy with a Chance of Insights: Context
888 Dependent Gene Regulation and Implications for Evolutionary Studies. Genes 2018,
889 Vol. 9, Page 82 **10**: 492–21.

890 Bushati N., Stark A., Brennecke J., Cohen S. M., 2008 Temporal reciprocity of miRNAs
891 and their targets during the maternal-to-zygotic transition in Drosophila. Current
892 Biology **18**: 501–506.

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

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

897 Coolon J. D., McManus C. J., Stevenson K. R., Graveley B. R., Wittkopp P. J., 2014
898 Tempo and mode of regulatory evolution in Drosophila. Genome Research **24**: 797–
899 808.

900 Crofton A. E., Cartwright E. L., Feitzinger A. A., Lott S. E., 2018 Effect of Larval Nutrition
901 on Maternal mRNA Contribution to the Drosophila Egg. G3 **8**: 1933–1941.

902 De Renzis S., Elemento O., Tavazoie S., Wieschaus E. F., 2007 Unmasking Activation
903 of the Zygotic Genome Using Chromosomal Deletions in the Drosophila Embryo.
904 PLOS Biology **5**: e117.

905 Eichhorn S. W., Subtelny A. O., Kronja I., Elife J. K., 2016 mRNA poly (A)-tail changes
906 specified by deadenylation broadly reshape translation in Drosophila oocytes and
907 early embryos. cdn.elifesciences.org.

908 Farrell J. A., O'Farrell P. H., 2014 From Egg to Gastrula: How the Cell Cycle Is
909 Remodeled During the Drosophila Mid-Blastula Transition. Annual Review of
910 Genetics, Vol 48 **48**: 269–294.

911 Gibson G., Riley-Berger R., Harshman L., Kopp A., Vacha S., *et al*., 2004 Extensive
912 Sex-Specific Nonadditivity of Gene Expression in Drosophila melanogaster.
913 Genetics **167**: 1791–1799.

914 Glaser-Schmitt A., Zečić A., Parsch J., 2018 Gene Regulatory Variation in *Drosophila*
915 *melanogaster* Renal Tissue. Genetics: genetics.301073.2018–46.

916 Graze R. M., McIntyre L. M., Main B. J., Wayne M. L., Nuzhdin S. V., 2009 Regulatory
917 Divergence in Drosophila melanogaster and D. simulans, a Genomewide Analysis
918 of Allele-Specific Expression. Genetics **183**: 547–561.

919 Grossniklaus U., Pearson R. K., Gehring W. J., 1992 The Drosophila sloppy paired
920 locus encodes two proteins involved in segmentation that show homology to
921 mammalian transcription factors. Genes Dev. **6**: 1030–1051.

922 Haines J. E., Eisen M. B., 2018 Patterns of chromatin accessibility along the anterior-
923 posterior axis in the early Drosophila embryo (JM Vaquerizas, Ed.). PLoS Genet.
924 **14**: e1007367–25.

925 Hart C. M., Cuvier O., Laemmli U. K., 1999 Evidence for an antagonistic relationship
926 between the boundary element-associated factor BEAF and the transcription factor
927 DREF. Chromosoma **108**: 375–383.

928 Heinz S., Benner C., Spann N., Bertolino E., Lin Y. C., *et al*., 2010 Simple Combinations
929 of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements
930 Required for Macrophage and B Cell Identities. Molecular Cell **38**: 576–589.

931 Heyn P., Kircher M., Dahl A., Kelso J., Tomancak P., *et al*., 2014 The Earliest
932 Transcribed Zygotic Genes Are Short, Newly Evolved, and Different across Species.
933 CellReports **6**: 285–292.

934 Jäckle H., Tautz D., Schuh R., Seifert E., Lehmann R., 1986 Cross-regulatory
935 interactions among the gap genes of Drosophila. Nature **324**: 668–670.

936 Kugler J.-M., Lasko P., 2009 Localization, anchoring and translational control of oskar,
937 gurken, bicoid and nanos mRNA during Drosophila oogenesis. Fly **3**: 15–28.

938 Lachaise D., David J. R., Lemeunier F., Tsacas L., Ashburner M., 1986 The
939 Reproductive Relationships of *Drosophila sechellia* with *D. mauritiana*, *D. simulans*,
940 and *D. melanogaster* from the Afrotropical Region. Evolution **40**: 262–271.

941 Landry C. R., Wittkopp P. J., Taubes C. H., Ranz J. M., Clark A. G., Hartl D. L., 2005
942 Compensatory cis-trans evolution and the dysregulation of gene expression in
943 interspecific hybrids of Drosophila. Genetics **171**: 1813–1822.

944 Laver J. D., Li X., Ray D., Cook K. B., Hahn N. A., *et al*., 2015 Brain tumor is a
945 sequence-specific RNA-binding protein that directs maternal mRNA clearance
946 during the Drosophila maternal-to-zygotic transition. Genome Biol. **16**: 94.

947 Lemos B., Araripe L. O., Fontanillas P., Hartl D. L., 2008 Dominance and the
948 evolutionary accumulation of cis- and trans-effects on gene expression.
949 Proceedings of the National Academy of Sciences **105**: 14471–14476.

950 León-Novelo L. G., McIntyre L. M., Fear J. M., Graze R. M., 2014 A flexible Bayesian
951 method for detecting allelic imbalance in RNA-seq data. BMC Genomics **15**: 920.

952 Li H., Handsaker B., Wysoker A., Fennell T., Ruan J., *et al*., 1000 Genome Project Data
953 Processing Subgroup, 2009 The Sequence Alignment/Map format and SAMtools.
954 Bioinformatics **25**: 2078–2079.

955 Li J., Gilmour D. S., 2013 Distinct mechanisms of transcriptional pausing orchestrated
956 by GAGA factor and M1BP, a novel transcription factor. The EMBO Journal **32**:
957 1829–1841.

958 Liang J., Lacroix L., Gamot A., Cuddapah S., Queille S., *et al*., 2014 Chromatin
959 Immunoprecipitation Indirect Peaks Highlight Long-Range Interactions of Insulator
960 Proteins and Pol II Pausing. Molecular Cell **53**: 672–681.

961 Lim J., Lee M., Son A., Chang H., Kim V. N., 2016 mTAIL-seq reveals dynamic poly(A)
962 tail regulation in oocyte-to-embryo development. Genes Dev. **30**: 1671–1682.

963 Lott S. E., Villalta J. E., Schroth G. P., Luo S., Tonkin L. A., Eisen M. B., 2011
964 Noncanonical Compensation of Zygotic X Transcription in Early Drosophila
965 melanogaster Development Revealed through Single-Embryo RNA-Seq. PLOS
966 Biology **9**: e1000590.

967 Lott S. E., Villalta J. E., Zhou Q., Bachtrog D., Eisen M. B., 2014 Sex-specific embryonic
968 gene expression in species with newly evolved sex chromosomes. (HS Malik, Ed.).
969 PLoS Genet. **10**: e1004159.

970 Mack K. L., Campbell P., Nachman M. W., 2016 Gene regulation and speciation in
971 house mice. Genome Research: gr.195743.115.

972 Macmanes M. D., 2014 On the optimal trimming of high-throughput mRNA sequence
973 data. Front Genet **5**: 13.

974 Martin M., 2011 Cutadapt removes adapter sequences from high-throughput
975 sequencing reads. EMBnet.journal **17**: pp. 10–12.

976 Matzat L. H., Lei E. P., 2014 Surviving an identity crisis: A revised view of chromatin
977 insulators in the genomics era. Biochimica et Biophysica Acta (BBA) - Gene
978 Regulatory Mechanisms **1839**: 203–214.

979 McDermott S. R., Kliman R. M., 2008 Estimation of Isolation Times of the Island
980 Species in the Drosophila simulans Complex from Multilocus DNA Sequence Data.
981 PLoS ONE **3**: e2442.

982 McManus C. J., Coolon J. D., Duff M. O., Eipper-Mains J., Graveley B. R., Wittkopp P.
983 J., 2010 Regulatory divergence in Drosophila revealed by mRNA-seq. Genome
984 Research **20**: 816–825.

985 Mi H., Muruganujan A., Huang X., Ebert D., Mills C., *et al*., 2019 Protocol Update for
986 large-scale genome and gene function analysis with the PANTHER classification
987 system (v.14.0). Nature Protocols 2012 7:3: 1–21.

988 Michalak P., 2003 Genome-Wide Patterns of Expression in Drosophila Pure Species
989 and Hybrid Males. Mol Biol Evol **20**: 1070–1076.

990 Mische S., Li M., Serr M., Hays T. S., 2007 Direct Observation of Regulated
991 Ribonucleoprotein Transport Across the Nurse Cell/Oocyte Boundary (S Wente,
992 Ed.). MBoC **18**: 2254–2263.

993 Moehring A. J., Teeter K. C., Noor M. A. F., 2007 Genome-Wide Patterns of Expression
994 in Drosophila Pure Species and Hybrid Males. II. Examination of Multiple-Species
995 Hybridizations, Platforms, and Life Cycle Stages. Mol Biol Evol **24**: 137–145.

996 Nolte V., Pandey R. V., Kofler R., Schlötterer C., 2013 Genome-wide patterns of natural
997 variation reveal strong selective sweeps and ongoing genomic conflict in Drosophila
998 mauritiana. Genome Research **23**: 99–110.

999 Nüsslein-Volhard C., Wieschaus E., 1980 Mutations affecting segment number and
1000 polarity in Drosophila. Nature **287**: 795–801.

1001 Omura CS and Lott SE. The conserved regulatory basis of mRNA contributions to the
1002 early embryo differs between the maternal and zygotic genomes. *in prep.*

1003 Paris M., Villalta J. E., Eisen M. B., Lott S. E., 2015 Sex Bias and Maternal Contribution
1004 to Gene Expression Divergence in Drosophila Blastoderm Embryos. (AG Clark,
1005 Ed.). PLoS Genet. **11**: e1005592.

1006 Quinlan A. R., Hall I. M., 2010 BEDTools: a flexible suite of utilities for comparing
1007 genomic features. Bioinformatics **26**: 841–842.

1008 R Core Team, 2017 *R: A language and environment for statistical computing*. R
1009 Foundation for Statistical Computing, Vienna, Austria.

1010 Ramírez F., Bhardwaj V., Arrigoni L., Lam K. C., Grüning B. A., *et al*., 2018 High-
1011 resolution TADs reveal DNA sequences underlying genome organization in flies.
1012 Nature Communications 2018 9:1 **9**: 1–15.

1013 Rouget C., Papin C., Boureux A., Meunier A.-C., Franco B., *et al*., 2010 Maternal mRNA
1014 deadenylation and decay by the piRNA pathway in the early Drosophila embryo.
1015 Nature **467**: 1128–U144.

1016 Salles F. J., Lieberfarb M. E., Wreden C., Gergen J. P., Strickland S., 1994 Coordinate
1017 initiation of Drosophila development by regulated polyadenylation of maternal
1018 messenger RNAs. Science **266**: 1996–1999.

1019 St Johnston D., Nüsslein-Volhard C., 1992 The Origin of Pattern and Polarity in the
1020 Drosophila Embryo. Cell **68**: 201–219.

1021 Supek F., Bošnjak M., Škunca N., Šmuc T., 2011 REVIGO Summarizes and Visualizes
1022 Long Lists of Gene Ontology Terms (C Gibas, Ed.). PLoS ONE **6**: e21800–9.

1023 Tadros W., Lipshitz H. D., 2009 The maternal-to-zygotic transition: a play in two acts.
1024 Development **136**: 3033–3042.

1025 Tadros W., Goldman A. L., Babak T., Menzies F., Vardy L., *et al*., 2007 SMAUG Is a
1026 Major Regulator of Maternal mRNA Destabilization in Drosophila and Its Translation
1027 Is Activated by the PAN GU Kinase. Developmental Cell **12**: 143–155.

1028 Theurkauf W. E., Hazelrigg T. I., 1998 In vivo analyses of cytoplasmic transport and
1029 cytoskeletal organization during Drosophila oogenesis: characterization of a multi-
1030 step anterior localization pathway. Development **125**: 3655–3666.

1031 Thomsen S., Anders S., Janga S. C., Huber W., Alonso C. R., 2010 Genome-wide
1032 analysis of mRNA decay patterns during early Drosophila development. Genome
1033 Biol. **11**: R93.

1034 Torres-Oliva M., Almudi I., McGregor A. P., Posnien N., 2016 A robust (re-)annotation
1035 approach to generate unbiased mapping references for RNA-seq-based analyses of
1036 differential expression across closely related species. BMC Genomics **17**: 392.

1037 Trapnell C., Roberts A., Goff L., Pertea G., Kim D., *et al*., 2012 Differential gene and
1038 transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks.
1039 Nature Protocols 2012 7:3 **7**: 562.

1040 Van der Auwera G. A., Carneiro M. O., Hartl C., Poplin R., del Angel G., *et al*., 2013
1041 From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit
1042 Best Practices Pipeline. Current Protocols in Bioinformatics **43**: 11.10.1–11.10.33.

1043 Vastenhouw N. L., Cao W. X., Lipshitz H. D., 2019 The maternal-to-zygotic transition
1044 revisited. Development **146**: dev161471–20.

1045 Wittkopp P. J., Haerum B. K., Clark A. G., 2004 Evolutionary changes in cis and trans
1046 gene regulation. Nature **430**: 85–88.

1047 Yang J., 2017 Substitute Reference Genome with Sample Specific SNP variants.
1048 Github Repository, https://github.com/yangjl/pseudoRef.

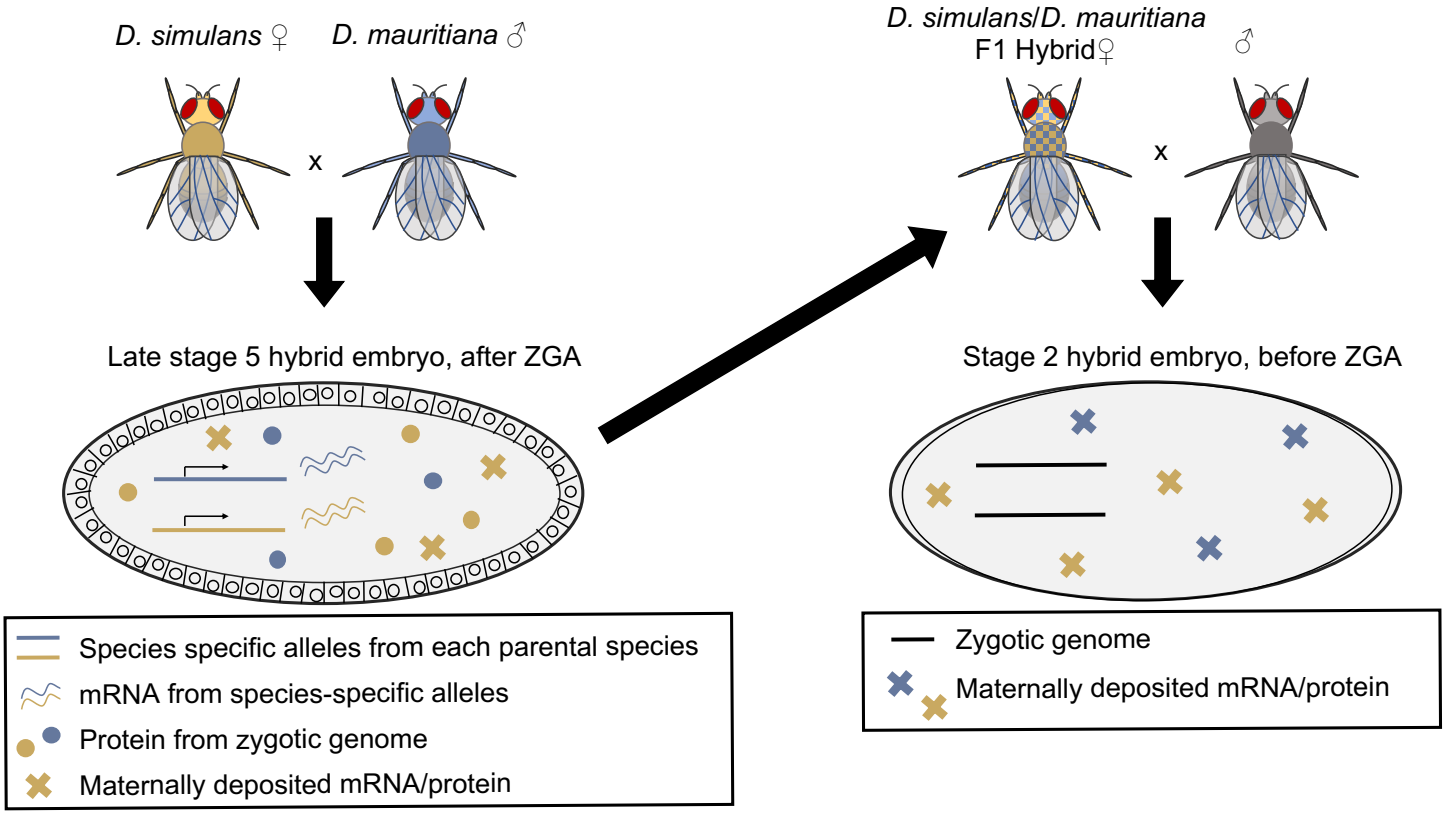


Figure 1: Crosses to produce hybrid embryos for the zygotic and maternal stages. To look at changes in regulation for zygotic genes, hybrid stage 5 embryos (left) were produced by crossing two parental species and collecting their eggs at the appropriate stage (late stage 5). To look at regulatory changes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

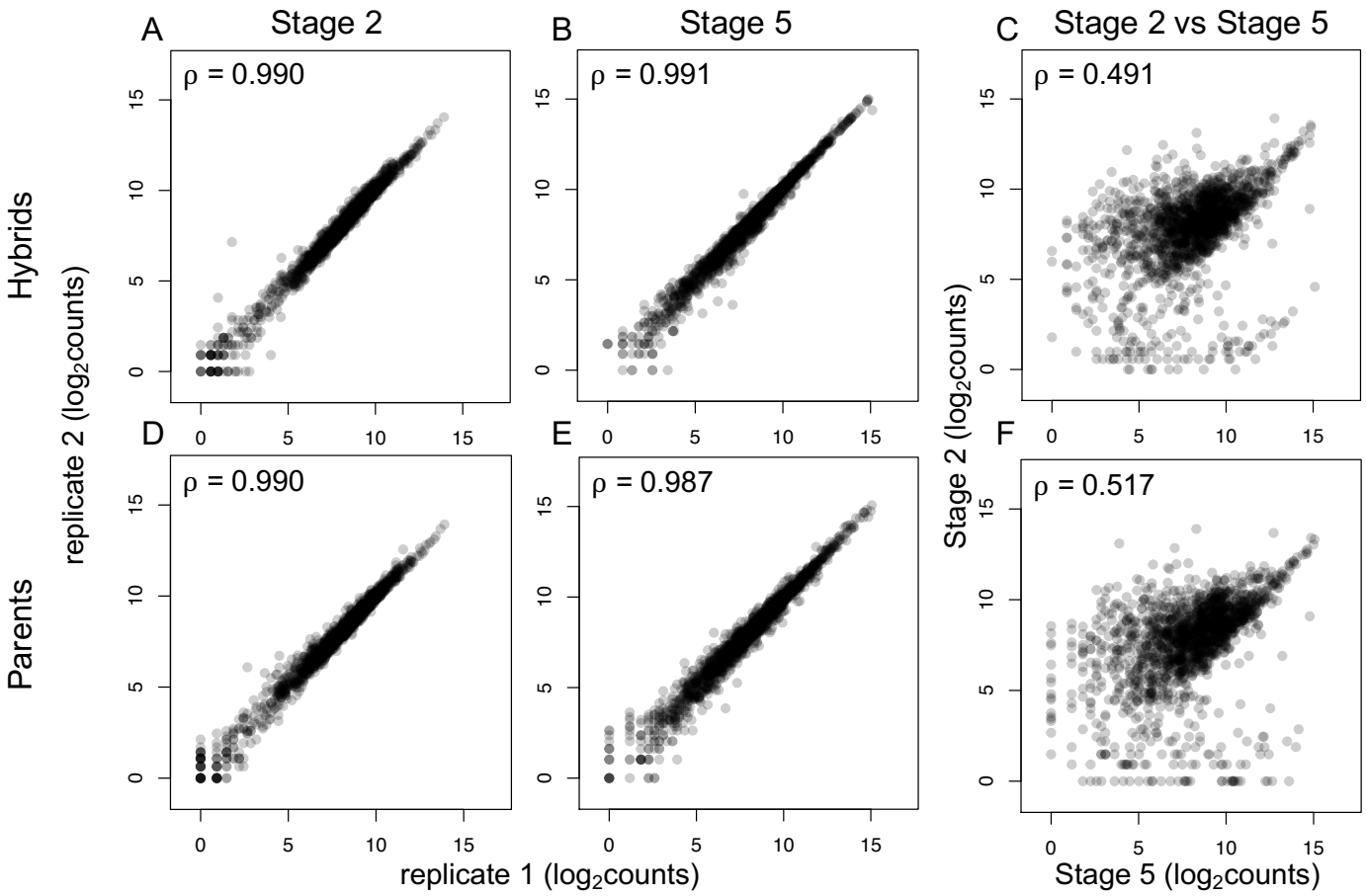


Figure 2: Hybrid and parental species single embryo transcript levels are highly reproducible. (A,B,D,E) Spearman's rank correlation coefficients are high when counts from replicate transcriptomes of the same stage and genotype are compared. Correlation coefficients are similarly high in parental species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

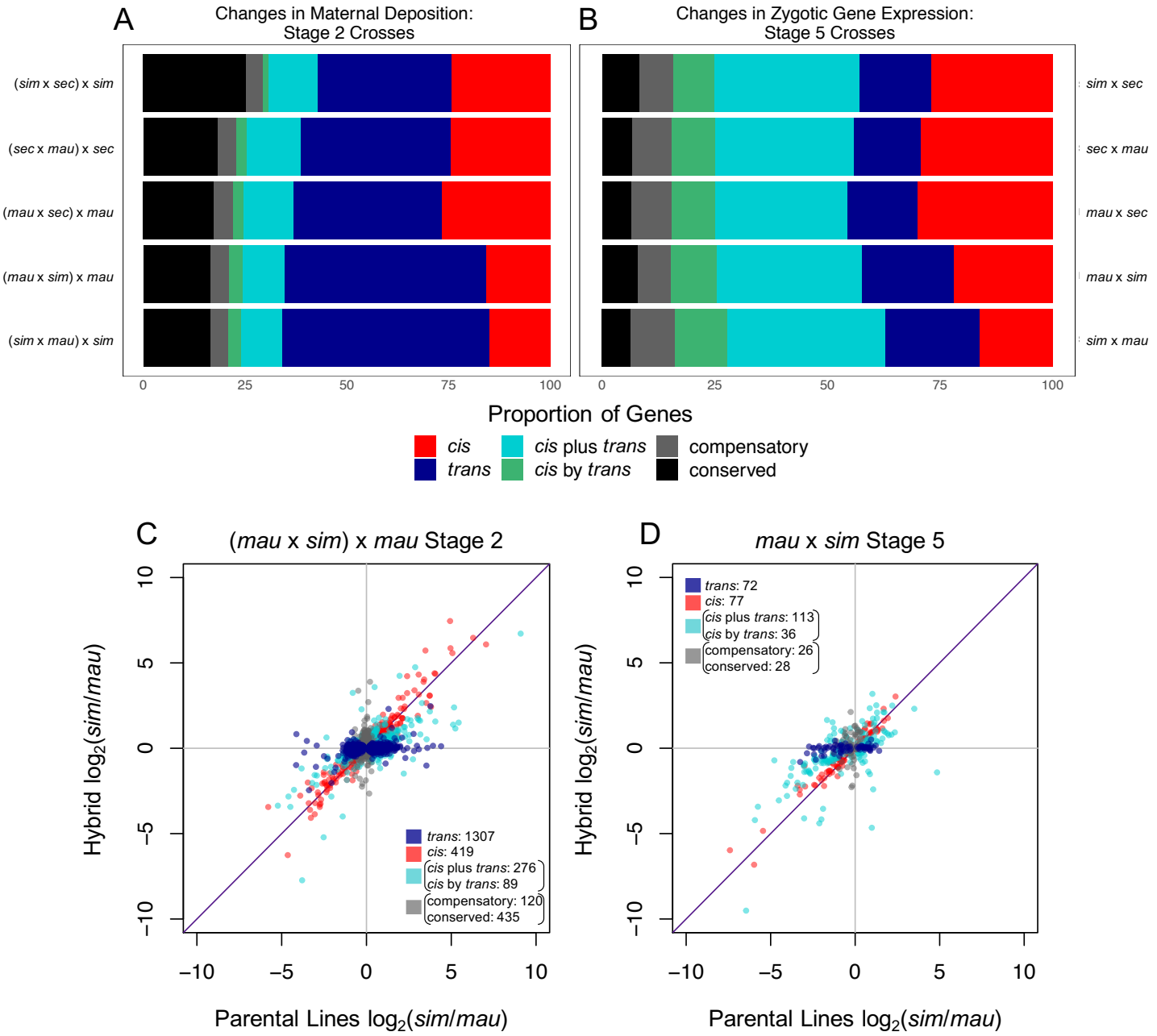


Figure 3: Different types of evolved regulatory changes dominate in maternal transcript deposition vs. zygotic transcription. Proportion of genes that fall into categories of regulatory change for each cross are shown for both the maternal transcript deposition (A) and zygotic gene transcription (B),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

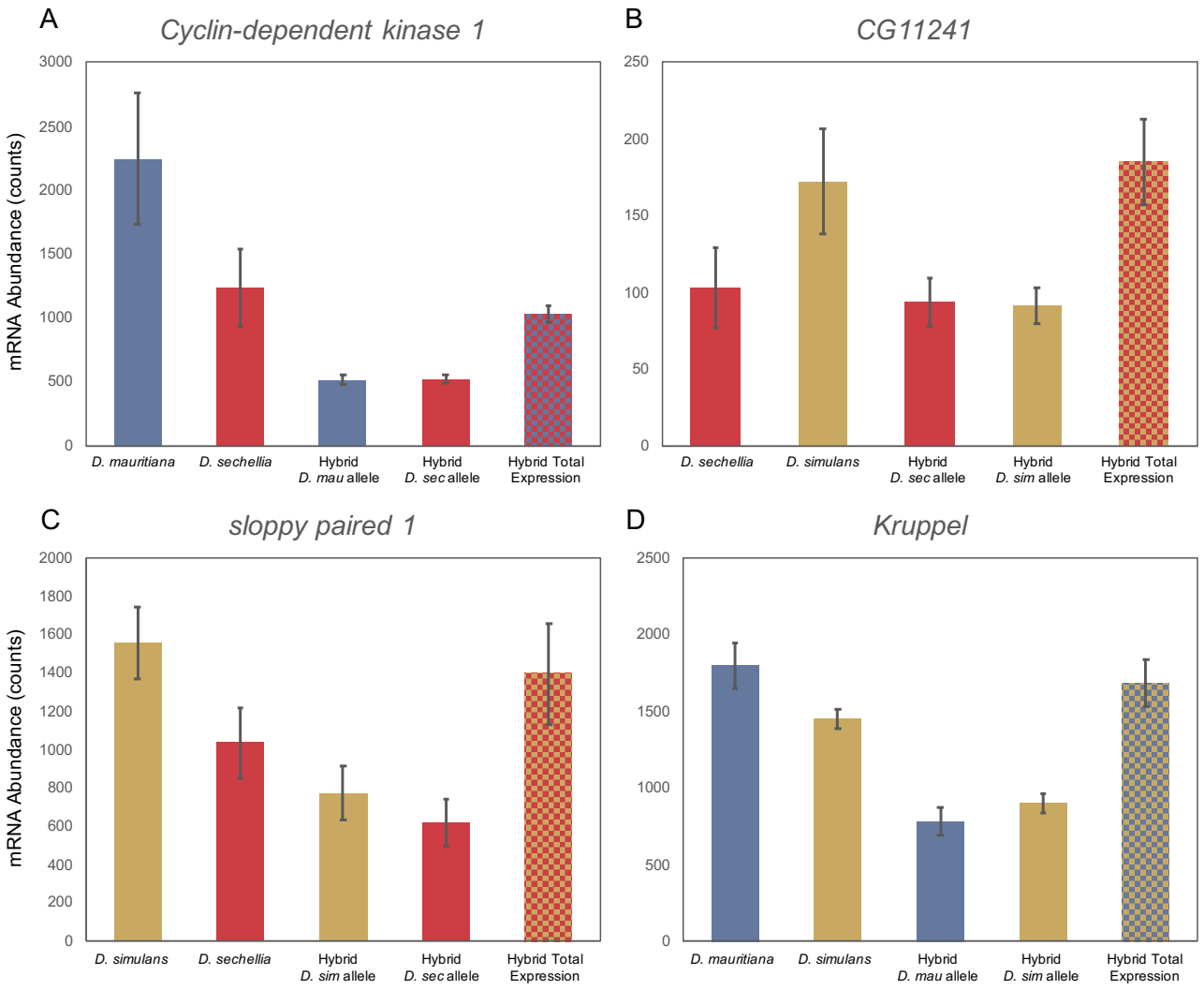


Figure 4: Examples of the type of regulatory changes observed, for individual genes. Transcript abundance, shown in counts, for each gene is plotted for both parental lines and for each parental allele within the hybrid; error bars shown represent the standard deviation. Total transcript abundance in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

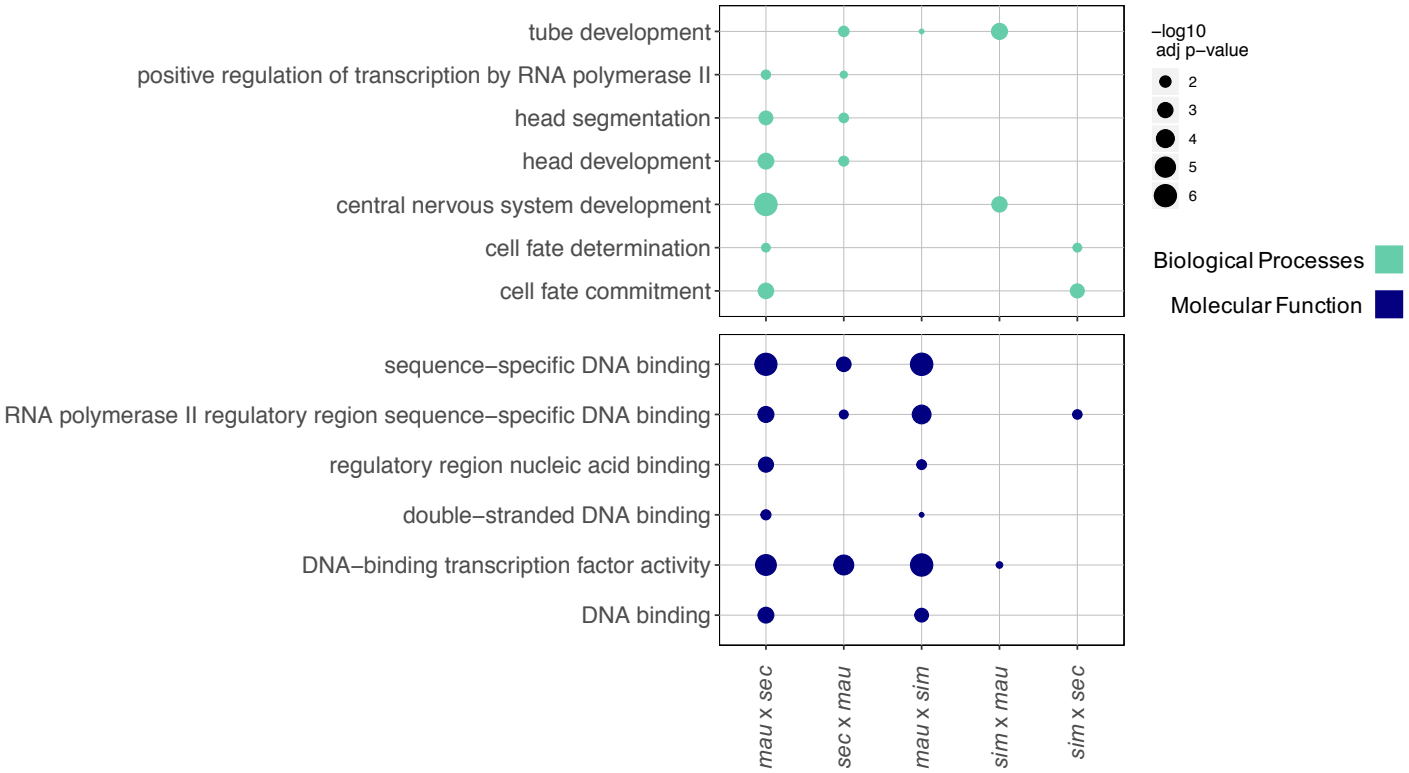


Figure 5: Gene ontology (GO) analysis identifies transcription factors that act in developmental processes as types of genes that change zygotically. Significantly enriched GO terms are listed for zygotically transcribed genes that change in *trans* regulation between each pair of species compared. Genes 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.

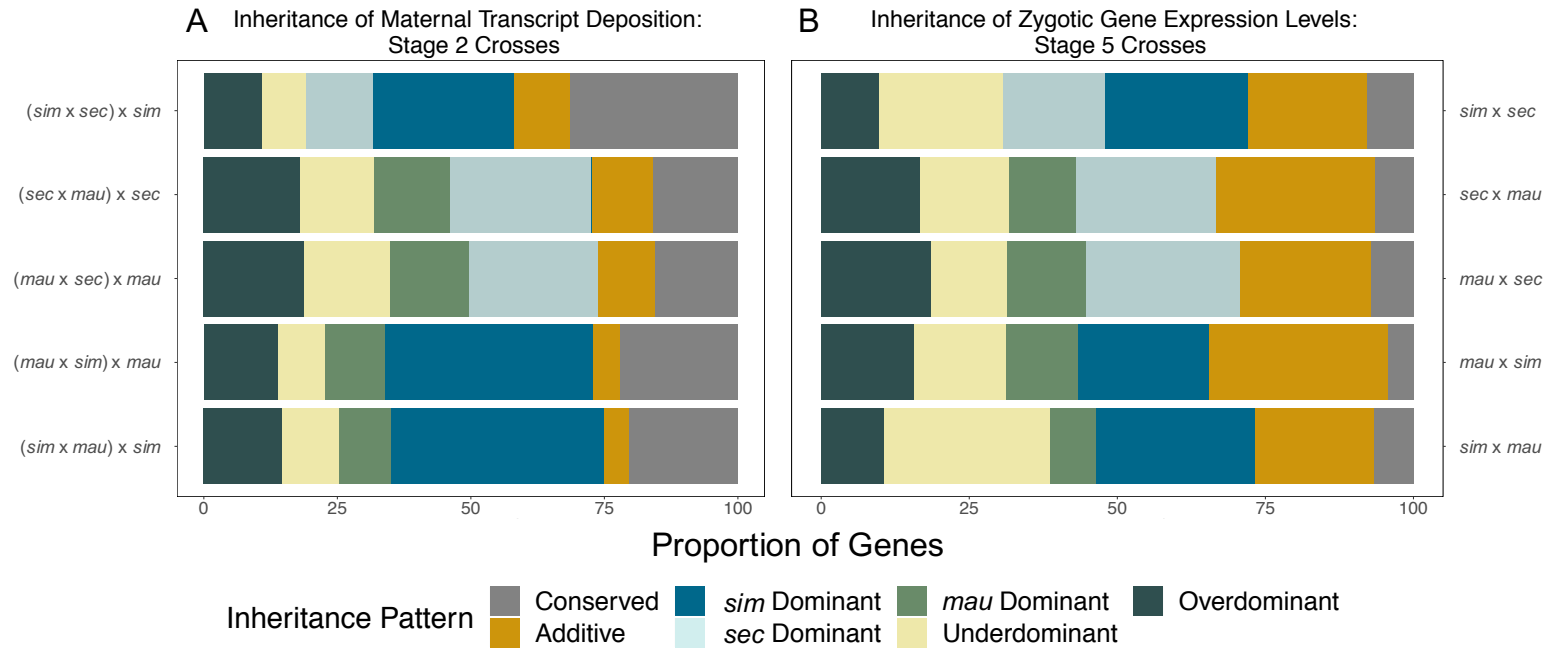


Figure 6: Patterns of inheritance show dominance of particular parental genomes at both stages. A) Shows patterns of inheritance for stage 2, over all genes and all crosses. B) Shows patterns of inheritance for stage 5, for mostly zygotic genes (see Methods) and all crosses. The maternal stage (A) shows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

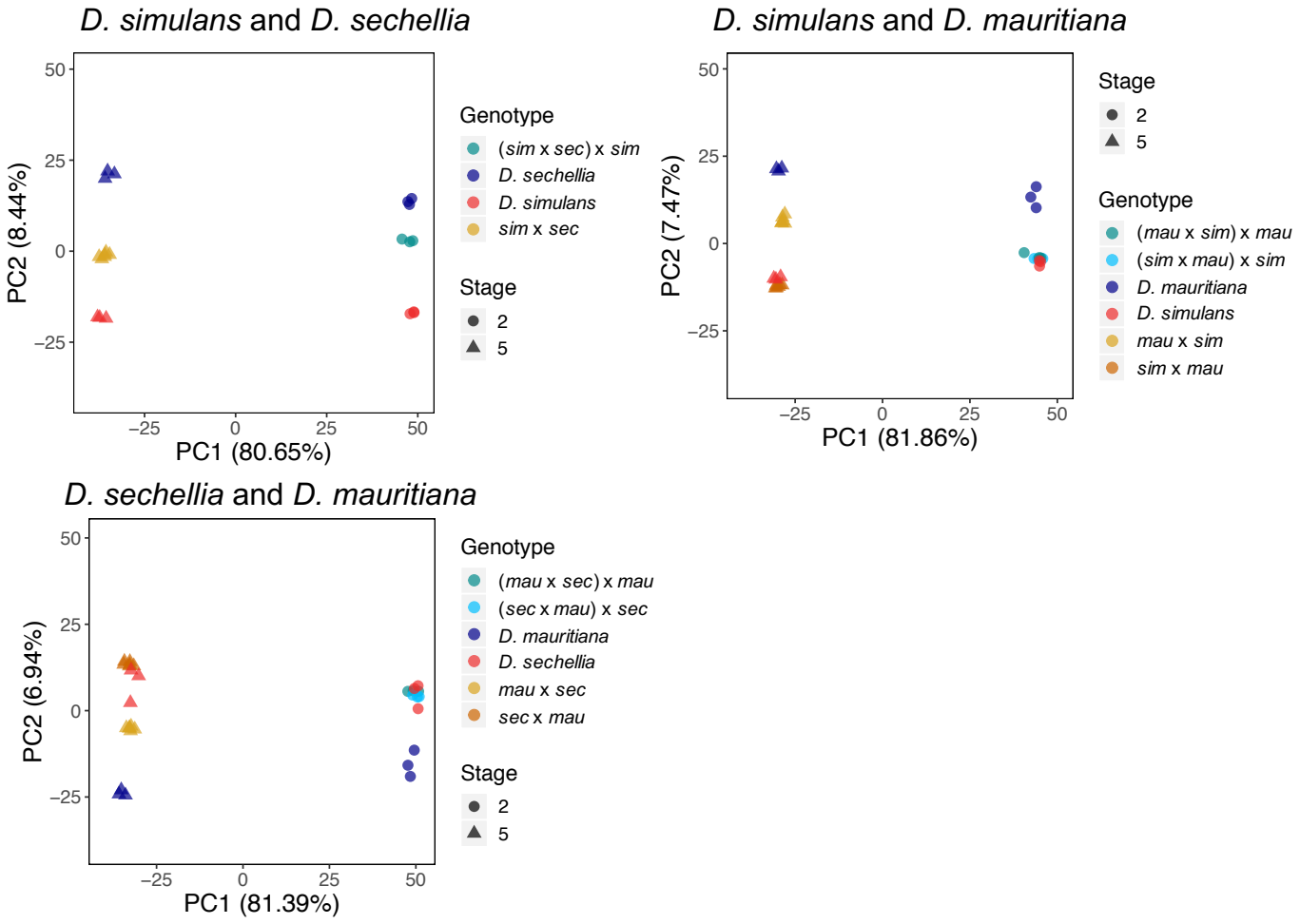


Figure S1: PCA plots for transcript abundance in all crosses. Samples of each stage, 2 or 5, cluster together. Samples of each genotype also cluster together, parental samples and hybrids. Proportion of variance explained by each principal component is listed on each axis.

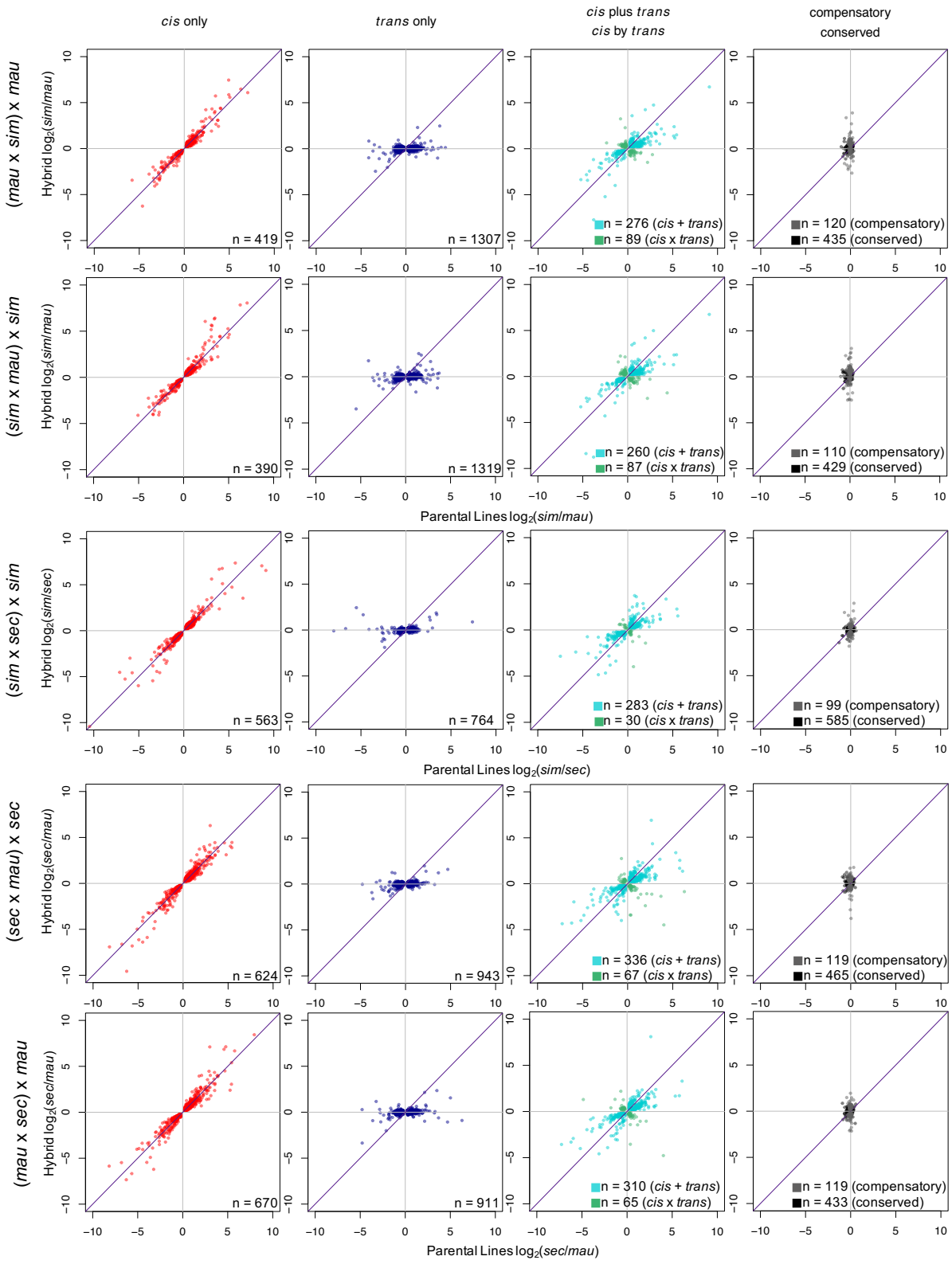


Figure S2: Regulatory changes in all pairwise comparisons for maternally deposited transcripts. Transcript level ratios between parental lines and within hybrids at stage 2 describe regulatory changes between species in each set of crosses. The number of genes in each category of regulatory change (n=)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

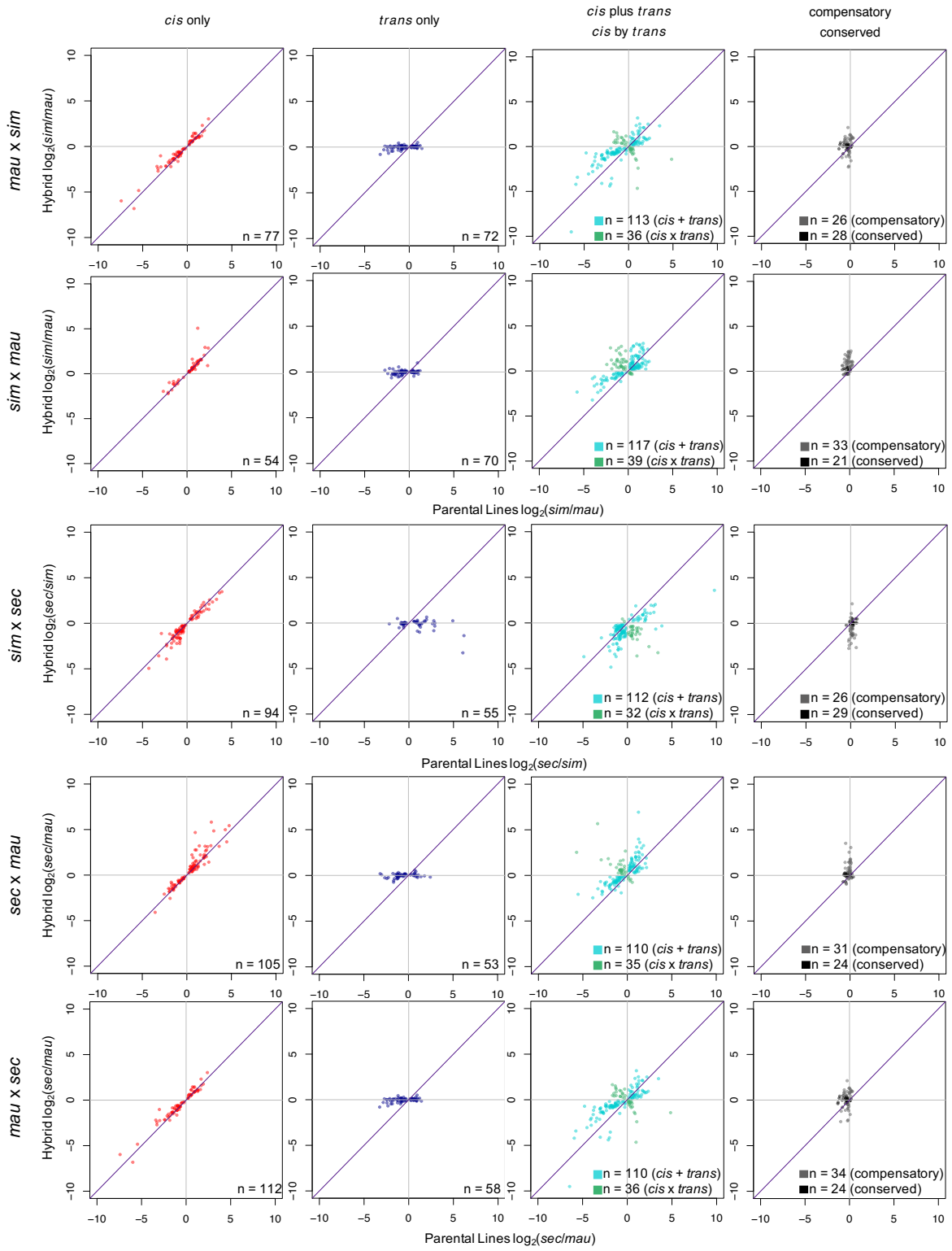


Figure S3: Regulatory changes in all pairwise comparisons for mostly zygotic genes. Transcript level ratios between parental lines and within hybrids at for mostly zygotic genes (see Methods) at stage 5 describe regulatory changes between species in each set of crosses. The number of genes in each category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

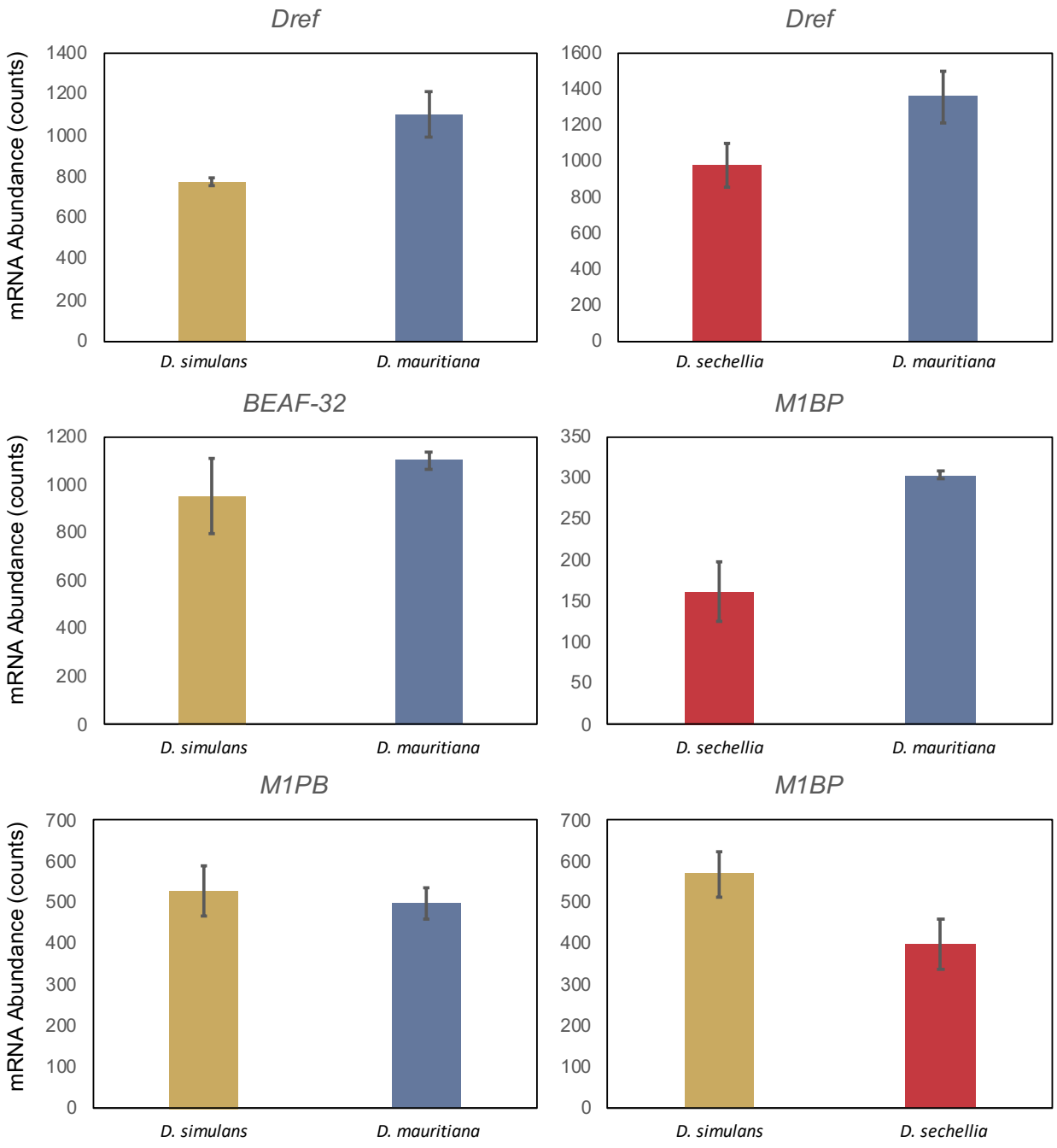


Figure S4: Transcript abundance from parental lines at stage 2 demonstrates differential maternal deposition of *M1BP*, *Dref* and *BEAF-32*. Counts for *D. simulans*/*D. mauritiana* and *D. simulans*/*D. sechellia* comparisons are averages across replicates from alignment to the *D. simulans genome*. 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.

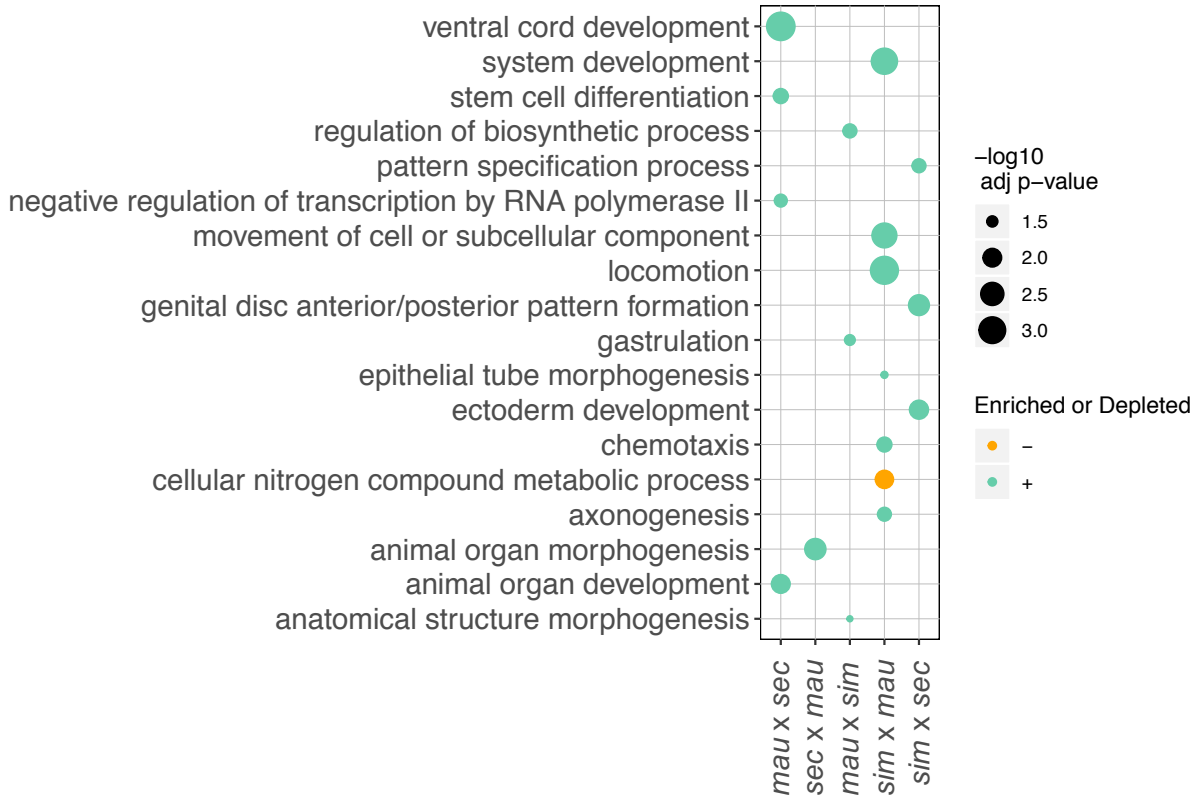


Figure S5: Gene ontology (GO) analysis for categories unique to a specific cross show enrichment for specific developmental processes. Significantly enriched GO terms are listed for zygotically transcribed genes that change in *trans* regulation between each pair of species compared. Again, zygotically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

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

Table S1: 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.

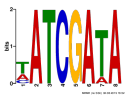
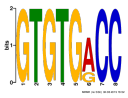
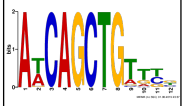
Motif	Predicted Binding Proteins	Enriched in upstream regions of *D. sechellia* genes that are conserved in expression (E-value)	Enriched in upstream regions of *D. sechellia* genes that are changing in expression (E-value)	Enriched in upstream regions of *D. simulans* genes that are conserved in expression (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 S4: 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.