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 GSE136646.

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

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

INTRODUCTION

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

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

47 primary mechanism of change in transcript abundance within or between species
48 (Graze *et al.* 2009; Mack *et al.* 2016), while other studies indicate that *trans* changes
49 are more widespread (McManus *et al.* 2010; Glaser-Schmitt *et al.* 2018). Previous
50 research has proposed that the difference in the abundance of *cis* versus *trans* changes
51 affecting gene expression can be explained by the divergence times of the strains or
52 species being compared (Coolon *et al.* 2014), or the particular tissue type examined
53 (Buchberger *et al.* 2019). Given this body of previous work, it is surprising that, to our
54 knowledge, no previous study has compared *cis* and *trans* contributions to gene
55 expression evolution across developmental stages in a model organism.
56 In this study, we ask whether contributions of *cis* and *trans* changes to gene
57 regulatory evolution differ across developmental stages. We chose stages during
58 embryogenesis that are close together in developmental time, yet are likely to broadly
59 differ in the mechanistic basis of regulation. The general regulatory architecture of a
60 particular stage likely affects how regulation at this stage can evolve. The early stages
61 of development utilized here are under regulatory control of entirely different genomes:
62 that of the mother, and that of the zygote. The earliest developmental processes in
63 embryogenesis are regulated by maternally provided RNA and protein, which lay the
64 foundation for the rest of development (Tadros and Lipshitz 2009; Vastenhouw *et al.*
65 2019). These maternally derived gene products carry out all initial developmental events
66 because at the time of fertilization, the zygotic genome is transcriptionally silent.
67 Because the zygotic genome is not yet transcriptionally active, post-transcriptional
68 mechanisms also play an important role in regulating the amount of maternal gene
69 products present (Tadros *et al.* 2007; Rouget *et al.* 2010; Barckmann and Simonelig

70 2013). As the zygotic genome is activated, control of developmental processes is
71 handed off from the maternally deposited factors to those derived from the zygotic
72 genome in a process known as the maternal-to-zygotic transition (MZT). The MZT is a
73 highly conserved and regulated process during early development that occurs in all
74 animals and in some species of flowering plants (Baroux *et al.* 2008; Tadros and
75 Lipshitz 2009; Vastenhouw *et al.* 2019).

76 Regulation of early zygotic gene expression has been extensively studied over
77 many years (Mannervik 2014; Schulz and Harrison 2019), while regulatory control of
78 maternal genes is not well understood. Studies of maternal transcripts have largely
79 focused on their transport into the oocyte (Mische *et al.* 2007; Kugler and Lasko 2009),
80 their localization and movement within the oocyte and embryo (Theurkauf and Hazelrigg
81 1998; Kugler and Lasko 2009), their post-transcriptional regulation (Salles *et al.* 1994),
82 and their degradation (Tadros *et al.* 2007; Bushati *et al.* 2008; Laver *et al.* 2015).
83 Zygotic gene expression is carefully controlled, with classic examples such as the *even-*
84 *skipped* gene having multiple enhancer elements along with multiple transcription
85 factors responsible for producing complex expression patterns in developmental time
86 and embryonic space (Small *et al.* 1992; Perry *et al.* 2011; Mannervik 2014). In contrast,
87 maternal transcripts are produced by support cells called nurse cells during oogenesis,
88 which are polyploid and highly transcriptionally active (Kugler and Lasko 2009; Lasko
89 2012), rapidly producing large amounts of transcripts. Roughly 50-75% of the genome is
90 maternally deposited (Tadros *et al.* 2007; De Renzis *et al.* 2007; Thomsen *et al.* 2010;
91 Lott *et al.* 2011; Vastenhouw *et al.* 2019) and there is considerable post-transcriptional
92 regulation of maternal factors (Tadros *et al.* 2007; Bushati *et al.* 2008; Laver *et al.*

93 2015). For these reasons, regulation at the transcript level may not need to be as
94 precise. While the mechanisms behind maternal transcription are not well understood,
95 the regulatory environments driving the production of the maternal and zygotic
96 transcriptomes are quite different.

97 To study the regulatory basis of differences in transcript levels at the maternal
98 and zygotic stages of early development between species, we focused on three closely
99 related species of *Drosophila* (*D. simulans*, *D. sechellia*, and *D. mauritiana*). Despite
100 having a relatively close divergence time of 250,000 years (McDermott and Kliman
101 2008), these sister species have differences in the pools of transcripts present in the
102 developing embryo both at a stage where only maternal transcripts are present, and at
103 a stage after zygotic genome activation (ZGA; Atallah and Lott 2018). By comparing
104 hybrids and parental lines of the species *D. simulans*, *D. sechellia*, and *D. mauritiana*,
105 we asked whether, at each of these two developmental stages, changes in gene
106 expression between species occurred due to changes in *cis*, in *trans*, or in a
107 combination of the two.

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

116 necessary for the intricate control of gene expression in time and space at this stage of
117 embryogenesis. The large proportion of *trans* regulatory signal found at the maternal
118 stage may reveal fundamental properties of the regulatory architecture during
119 oogenesis. *Trans* regulators can affect a large number of genes at the same time, as
120 might be necessary to maximize mRNA production to load sufficient numbers of
121 transcripts into the egg. We also identified motifs associated with *trans* regulation at the
122 level of chromatin at the maternal stage, which lends evidence to an emerging picture of
123 how gene expression might be regulated during oogenesis. Overall, we find distinct
124 patterns of gene regulatory changes at the two embryonic timepoints, before and after
125 ZGA, indicating evolved changes in gene regulation differ based on the developmental
126 context.
127

128 **MATERIALS AND METHODS**

129

130 **Crossing scheme and sample acquisition**

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

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

162 species in the initial cross in a similar way, e.g. (*mau* x *sim*) x *mau* (also see Figure 1 for
163 cross diagram).
164 All flies were raised in vials on standard cornmeal media at 25°C. Flies were
165 allowed to lay eggs for ~2 hours (for collecting stage 2 embryos) and ~3 hours (for
166 collecting stage 5 embryos) before they were transferred to a new vial so that the eggs
167 could be harvested. Eggs were collected from 4-14 day old females, dechorionated
168 using 50% bleach and moved into halocarbon oil on a microscope slide for staging.
169 Embryos were staged at the appropriate developmental time point under a microscope
170 (Zeiss AxioImager M2), imaged, and promptly collected at stage 2 or at the end of stage
171 5 (Bownes' stages; Bownes 1975; Campos-Ortega and Hartenstein 2013), of embryonic
172 development.
173 To collect the samples after staging, the embryos were quickly transferred with a
174 paintbrush to Parafilm (Bemis) and rolled (to remove excess halocarbon oil) into a drop
175 of TRIzol (Ambion). The embryos were ruptured with a needle so that the contents
176 dissolved in the TRIzol and were transferred to a tube to be frozen at -80°C until
177 extraction. RNA was extracted using glycogen as a carrier (as per manufacturer
178 instructions) in a total volume of 1mL TRIzol. Approximately 80-120ng total RNA was
179 extracted from individual embryos, measured using a Qubit 2.0 fluorimeter (Invitrogen).
180 The quality of the RNA was validated on RNA Pico Chips using an Agilent Bioanalyzer.
181 Genotyping was performed to determine embryo sex for stage 5 samples, as
182 dosage compensation is not complete and transcript levels for genes on the X
183 chromosome may differ for males and females at this time in development (Lott *et al.*
184 2014). DNA was extracted from each sample along with the RNA as per manufacturer

185 instructions and amplified using a whole genome amplification kit (illustra GenomePhi
186 v2, GE Healthcare). Sex-specific primers (Table S1) designed for use with all three
187 species, two sets for the Y chromosome (to the *ORY* and *kl2* genes) and one control set
188 (to *ftz*), were used to genotype the single embryos after genome amplification. For the
189 stage 5 samples, a total of three male and three female embryos from each cross were
190 used for sequencing. One noted exception is that in the *sim* x *mau* cross, a total of four
191 female and two male embryos were collected, as determined from the transcriptomic
192 data. A total of three stage 2 embryos were collected. We did not perform genotyping
193 for embryo sex on the stage 2 embryos because the zygotic genome is not yet active at
194 this stage in development.
195
196 **Library preparation and sequencing**
197 The RNA from single embryos was treated with DNase (TurboDNA-free, Life
198 AM1907) using manufacturer instructions and RNA sequencing libraries were
199 constructed with Illumina TruSeq v2 kits following the manufacturer's low sample
200 protocol. The Illumina protocol uses oligo (dT) beads to enrich for polyadenylated
201 transcripts. Because poly(A) tail length is important in determining many post-
202 transcriptional processes during early development, including translational efficiency, it
203 is important to ensure that the method used for mRNA selection does not produce a
204 biased set of poly(A) tail lengths. Previous datasets report poly(A) length distributions
205 for transcripts during oogenesis and early development (Lim *et al.* 2016; Eichhorn *et al.*
206 2016). We could not directly compare our data to previous reports, as these studies
207 were done using *D. melanogaster*, which may have a different poly(A) tail length

208 distribution than the species used in our analysis. However, previous studies comparing
209 distributions of poly(A) tail lengths of all genes to poly(A) tail lengths of transcripts
210 recovered through poly(A) selection in *D. melanogaster* have demonstrated that poly(A)
211 selection with commonly used methods does not bias which transcripts are recovered
212 from the total pool of transcripts present (Eichhorn *et al.* 2016). This includes studies
213 that used the same single embryo approaches utilized here (Crofton *et al.* 2018; Atallah
214 and Lott 2018). Therefore, it seems unlikely that poly(A) selection heavily biases the
215 extracted RNA relative to the RNAs present at these developmental stages. cDNA
216 libraries were quantified using a Qubit 2.0 fluorimeter (dsDNA BR Assay Kits) and the
217 quality of the libraries were assessed on High Sensitivity DNA chips on an Agilent
218 Bioanalyzer. The libraries were pooled (11-12 samples per lane) and sequenced
219 (100bp, paired-end) in four lanes on an Illumina HiSeq2500 sequencer. Sequencing
220 was done at the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley.
221
222 **Data Processing**
223 Raw reads were processed to remove adapter sequences and gently trimmed
224 (PHRED Q<5; Macmanes 2014) using Cutadapt (version 1.7.1; Martin 2011). TopHat
225 (version 2.0.13; Trapnell *et al.* 2012) was used to align reads to the *D. simulans* (version
226 r2.02) and *D. sechellia* (version r1.3) genome assemblies (from the twelve species
227 project, downloaded from Flybase) and to the *D. mauritiana* MS17 assembly (Nolte *et*
228 *al.* 2013). Because the *D. mauritiana* line used for sequencing and the line used to
229 construct the genome assembly differed, variant sites from the lab line, called using
230 Genome Analysis Toolkit's (GATK) Haplotypecaller, were incorporated into the MS17

231 assembly using Pseudoref (http://yangjl.com/pseudoRef/; Xu *et al.* 2020). Additionally,
232 an updated annotation file for the MS17 assembly (Torres-Oliva *et al.* 2016) was used
233 during alignment and in subsequent processing steps. Annotation files for *D. simulans*
234 and *D. sechellia* were obtained from the same versions of the genome release of each
235 species. Read alignment, mismatches, edit distance, and gap length were all set to
236 three when using TopHat (version 2.0.13; Trapnell *et al.* 2012) to allow for a higher rate
237 of read alignment.

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

254 Normalization to the upper quantile was performed across all samples in each set of
255 pairwise species comparisons. This was used to account for differences in the number
256 of reads for each sample as a result of sequencing.

257 Read counts represent the number of reads mapping to variant sites within a
258 gene. A cutoff of 5 or more reads mapped to any given gene was set to determine if a
259 gene was expressed. Genes with read counts <5 in both species in any pairwise
260 comparison were not considered to be expressed in either species so were removed
261 from the analysis. This cutoff was tested empirically and was set to exclude genes with
262 low count numbers that had a higher frequency of mapping in a biased manner to both
263 parental genomes. Genes analyzed in this analysis were also limited to those with
264 annotated orthologs in both species in any pairwise comparison. An orthologs table
265 from Atallah and Lott, 2018, was updated using the annotations available on FlyBase
266 (v2017) and an updated set of annotations from Torres-Oliva, *et al*. 2016. This revised
267 orthologs table (Table S2) was used to compare genes between each species and in
268 each direction of mapping.

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

277 direction of mapping. We set a slightly conservative standard for classifying allelic
278 imbalance where genes with CIs below 0.49 or above 0.51 were called differentially
279 expressed. Genes with CIs close to 0.50 did not appear differentially expressed when
280 looking at the count data, so we used a more conservative cutoff. Genes that appeared
281 differentially expressed in one direction of mapping (CIs fell outside of the range of 0.49
282 - 0.51 when comparing the expression levels of parental alleles in each replicate) but
283 not in the other direction of mapping were removed from the analysis, as this was
284 determined to be a result of mapping bias between the two genomes. We also removed
285 genes that had disparate confidence intervals in the two mapping directions (i.e. one
286 mapping direction yielded a CI that fell above the 0.49 - 0.51 range and the other
287 direction of mapping yielded a CI that fell below the 0.49 - 0.51 range).
288 We found that between 9.6% and 10.9% of genes expressed at stage 2 (with a
289 count >5) mapped in a biased way to parental genomes when compared to the total
290 number of orthologous genes between any pair of species. Each between-species
291 comparison has a different number of orthologous genes so the proportion of biased
292 genes varies based on the pair of species compared in a cross. In contrast to the
293 maternal stage, we found that between 5.0% and 6.2% of genes expressed at stage 5
294 mapped in a biased way to parental genomes when compared to the total number of
295 orthologous genes between a pair of species. Overall, when looking at the total
296 proportion of biased genes, not just those that were called “expressed” in our analysis,
297 we found that between 24.8% and 28.3% of genes at stage 2 and between 20.0% and
298 21.2% of genes at stage 5 mapped in a biased manner when compared to the total
299 number of orthologous genes in any comparison between species in a cross. All the

300 genes that mapped in a biased manner were removed from our analysis. Genes that
301 were not biased in their mapping and had a read count of >5 reads were retained for
302 analysis.
303
304 **Genes used for stage 5 analysis**
305 To focus on the gene regulation from the zygotic genome after ZGA, we removed
306 genes with high levels of maternal transcript deposition from our analysis. We limited
307 the pool of genes analyzed to those that are mostly zygotic because roughly half of
308 maternal transcripts are not entirely degraded by stage 5 (although studies are
309 somewhat variable in the percent reported; Tadros *et al.* 2007; De Renzis *et al.* 2007;
310 Thomsen *et al.* 2010; Lott *et al.* 2014) and we wanted to examine only those genes that
311 have a larger contribution to expression from the zygotic genome. For this analysis, we
312 included genes with "zygotic-only" expression (those that are not maternally deposited)
313 and genes that are "mostly zygotic" (those with 8-fold higher expression at stage 5
314 relative to stage 2, a $\log_2$ difference greater than three). We added a count of 1 to the
315 transcript count for genes at the maternal and zygotic stages when calculating the $\log_2$
316 difference to avoid errors associated with taking the $\log_2$ of zero. We tested several
317 cutoffs but chose the 8-fold threshold because at this conservative cutoff, most genes
318 with high maternal transcript deposition are removed from the analysis. Additionally, for
319 this analysis we used confidence intervals and averages generated from only female
320 samples for genes on the X chromosome because dosage compensation is not
321 complete at stage 5 (Lott *et al.* 2011).
322

323

324 **Correlation analysis and PCA**

325 We performed correlation analysis (Figure 2, Table S3) between single embryos
326 across replicates, stages, and genotypes in R (R Core Team 2017) using the Spearman
327 option within the *corr* function. Principal component analysis (PCA) was also performed
328 in R using the *prcomp* function (Figure S1).

329

330 ***Cis/Trans* analysis**

331 To identify evolved regulatory changes between species, we first determined
332 which genes showed differential expression between alleles using the 95% CIs from PG
333 models (León-Novelo *et al.* 2014) that were also used to interpret mapping bias. We
334 compared expression levels of alleles for individual genes within the hybrids and also
335 between parental samples using this model. Genes with CIs that fell outside of the 49%
336 - 51% range were defined as differentially expressed, either within the hybrid or
337 between parental samples, while those falling within the 49% - 51% range were
338 identified as having the same level of expression. Genes were then categorized using
339 *cis*, *trans*, *cis* + *trans*, *cis x trans*, *compensatory*, and *conserved* categories as described
340 in Landry, *et al*. 2005; McManus, *et al*. 2010; and Coolon, *et al*. 2014 (Figures 3, S2 and
341 S3). We assigned the following categories for regulatory change based on the CIs
342 generated from PG models 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. mauritiana* 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.*
363 *mauritiana* 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%)

369

370 *compensatory*: Genes that are not differentially expressed between the parental

371 species (CI overlaps 49% - 51%) but are differentially expressed in the hybrids

372 (CI does not overlap 49% - 51%).

373

374 *conserved*: Genes are not differentially expressed between the parental species

375 or within the hybrids (CIs overlap 49% - 51%).

376

377 **Inheritance Patterns**

378 Previous studies from Gibson, *et al*. 2004 and McManus, *et al*. 2010 identified

379 and outlined ways to classify inheritance patterns of transcript abundance in hybrids in

380 relation to parental samples. We used these methods in our study to compare the

381 averages of total expression levels in the hybrids relative to those of parental samples.

382 Gene expression was considered conserved if the expression level between parental

383 samples and the total expression in the hybrid (sum of the expression of the two

384 species-specific alleles in the hybrid) were within 1.25-fold of one another, a $\log_2$-fold

385 change of 0.32. Overdominant genes were expressed at least 1.25-fold more in the

386 hybrid than in either parent while underdominant genes were expressed at least 1.25-

387 fold lower in the hybrid than in either parent. Genes that were expressed at an

388 intermediate level in the hybrid in comparison to the parental species samples involved

389 in the cross were defined as additive. Dominance was determined when the hybrid had

390 expression within 1.25-fold of one of the parental species such that total transcript levels

391 in the hybrid was more similar to transcript levels in one parental species than in the
392 other parental species.
393
394 **Candidate transcription factor identification**
395 We took a computational approach to identify potential transcription factors that
396 may change in *trans* regulation between the species in our analysis. We used motif
397 enrichment programs to find potential binding sites in the upstream regions of genes
398 changing in regulation in *D. sechellia* and *D. simulans*. We omitted *D. mauritiana* from
399 this analysis because the *D. mauritiana* genome is not as well annotated as the
400 genomes for *D. simulans* and *D. sechellia*. We used the Differential Enrichment mode in
401 MEME (Bailey and Elkan 1994) as well as the findMotifs.pl script in HOMER (Heinz *et*
402 *al.* 2010) to identify overrepresented motifs in the regions 500bp upstream of the
403 annotated starting location for genes changing in regulation or with conserved
404 regulation between species in every set of comparisons at stage 2. We utilized a 500bp
405 region as this was empirically determined to give the highest enrichment of signal for
406 motifs in target genes relative to background (see Omura and Lott 2020 for more
407 information). In MEME, we used options to find motifs with any number of repetitions
408 and a motif width of 8-12. We used default options for HOMER and supplied a
409 background FASTA file for enrichment analysis. The background lists supplied were
410 500bp upstream regions from all annotated genes in the species except for those that
411 were in the target set (either those genes with conserved or changing regulation in any
412 set of comparisons). The 500bp regions were extracted from FASTA files (versions
413 were the same as ones used for mapping) for each species using BEDTools (Quinlan

414 and Hall 2010). Significantly overrepresented motifs in the target lists relative to the
415 background supplied were then compared against databases of known transcription
416 factor binding sites using Tomtom (MEME suite) and HOMER. All enriched motifs that
417 appeared in both HOMER and MEME analyses are included in Table S4. All potential
418 targets of discovered motifs with significant E-values (MEME) or high Match Rank
419 scores in HOMER (>0.8) are also listed in Table S4 (see Figure S4 for transcript levels
420 of differentially maternally deposited targets in embryos of parental species).
421
422 **Gene Ontology**
423 Gene ontology (GO) analysis was done with the statistical overrepresentation
424 test in PANTHER (Mi *et al.* 2019) using the default settings. We looked at the GO
425 complete annotations for biological processes and molecular function but did not find
426 any significant terms represented in the cellular component categories. For this
427 analysis, we set a cutoff of Bonferroni adjusted p-value < 0.05. We searched for
428 enrichment of GO categories amongst genes that change in *trans* in each cross,
429 compared to the background of genes that are expressed (having a count >5) in each
430 cross. We used REVIGO (Supek *et al.* 2011) to reduce the number of redundant GO
431 categories and used the small (0.5) level of similarity as a cutoff for redundant GO
432 terms. GO categories shared between two or more crosses at stage 5 are represented
433 in Figure 5 and GO categories unique to a cross are shown in Figure S5. All enriched
434 categories are listed in Table S5.
435
436 **Data Availability**

437 All sequencing data and processed data files from this study are available at NCBI/GEO
438 at accession number: GSE136646.

439 **RESULTS**

440

441 In order to compare the regulatory basis of evolved changes in gene expression
442 at different stages of early embryogenesis, one stage where all the transcripts are
443 maternally provided and the other after zygotic genome activation (ZGA), we performed
444 a series of crosses between closely related species followed by RNA-seq on resulting
445 embryos (Figure 1). We used the sister species *D. simulans*, *D. sechellia*, and *D.*
446 *mauritiana*, all of which may be crossed reciprocally (with the exception of *D. sechellia*
447 females to *D. simulans* males; Lachaise *et al.* 1986). As transcripts in the two early
448 embryonic stages of interest are produced by different genomes, that of the mother and
449 that of the zygote, we performed crosses to produce a hybrid genome in the appropriate
450 generation (that of the mother or that of the zygote; Figure 1). To investigate regulatory
451 changes in zygotic gene expression, the three species were crossed pairwise (with the
452 noted exception), to produce F1 hybrid embryos, which were collected at a stage after
453 zygotic genome activation (end of blastoderm stage or the end of stage 5, Bownes'
454 stages; Bownes 1975; Campos-Ortega and Hartenstein 2013). While the zygotic
455 genome is fully activated at this developmental stage, maternal transcripts are not yet
456 entirely degraded so we limited our analysis to those genes that are expressed at a
457 much higher level after ZGA than before the zygotic genome is activated (see Methods).
458 To discover the regulatory basis of changes in maternal transcript deposition, the F1
459 females were crossed to males of the same species as the maternal species in the

460 initial cross. Resulting embryos were collected at stage 2 (Bownes' stages; Bownes
461 1975; Campos-Ortega and Hartenstein 2013), when all the transcripts in the egg are
462 maternal in origin (Figure 1). Three replicate samples were obtained for each cross at
463 stage 2, and since stage 5 features incomplete X chromosomal dosage compensation
464 (Lott *et al.* 2011), 6 replicates were obtained for each cross at late stage 5 (3 female
465 and 3 male embryos with one noted exception, see Methods). mRNA-sequencing
466 libraries were constructed from each embryo sample using poly(A) selection. Libraries
467 were sequenced paired-end, 100bp, on an Illumina HiSeq2500.
468
469 **Reproducibility of Single Embryo RNA Sequencing Data**
470 Previous studies have shown that single-embryo RNA-seq data is highly
471 reproducible, despite replicate samples representing both biological and technical
472 replicates (Lott *et al.* 2011, 2014; Paris *et al.* 2015; Atallah and Lott 2018). Our current
473 study extends this to include replicates of F1 crosses and subsequent backcrosses
474 between closely related species, which are as reproducible as the within-species
475 replicates. Spearman's rank correlation coefficients are high between replicate samples
476 of the same species or cross at the same developmental stage (Figure 2, A,B,D,E,
477 Table S3). For example, when comparing the RNA-seq data from stage 2 samples of
478 the (*mau* x *sim*) x *mau* and (*sim* x *mau*) x *sim* hybrid crosses, correlation coefficients
479 range from 0.965 to 0.995 (Table S3). Stage 5 hybrids from the *mau* x *sim* cross have
480 equally high correlation coefficients, ranging from 0.980 to 0.996 (Table S3). Similarly,
481 correlation coefficients for the RNA-seq data from *D. simulans* stage 5 embryos, when
482 compared with other *D. simulans* stage 5 embryos, range from 0.985 to 0.990. The high

483 correlation coefficients between replicates may be due, in part, to the removal of genes
484 with differential mapping to either parental genome and those genes with very low
485 transcript abundances (see Methods) from this analysis.

486 Transcript levels for embryos of the same stage but different genotypes (parental
487 lines and hybrids) are highly similar, as indicated by their Spearman's rank correlation
488 coefficients (Table S3), with one notable exception. When we compare RNA-seq
489 profiles from stage 5 hybrids to stage 5 embryos of the paternal species in the cross, we
490 see more divergent patterns of gene expression than when we compare stage 5 hybrids
491 to stage 5 embryos of the maternal species in the cross. This is due to the fact that
492 many maternal transcripts are still present at the zygotic stage, and thus the hybrid
493 zygotic embryo has many remaining maternal transcripts from the maternal species, not
494 the paternal species, in the cross. For example, comparisons between *D. simulans*
495 stage 5 embryos and stage 5 embryos of the *sim* x *mau* cross, where *D. simulans* is the
496 maternal species in the cross, yield high correlation coefficients, ranging from 0.955 to
497 0.972. In contrast, correlation coefficients are much lower when comparing *sim* x *mau*
498 stage 5 hybrid embryos to stage 5 embryos of the paternal species in the cross, *D.*
499 *mauritiana*, ranging from 0.863 to 0.887. In this particular comparison, the lower
500 correlation coefficients are likely due to having *D. simulans* as the maternal species in
501 the hybrid cross for the *sim x mau* embryos. Remaining maternal transcripts are from
502 the *D. simulans* alleles and likely explain why these hybrid embryos correlate more
503 highly with *D. simulans* stage 5 embryos.

504 In contrast to highly correlated samples within a stage, comparing the transcript
505 abundance of different stages yields strikingly lower correlation coefficients (Figure 2C

506 and F, Table S3), emphasizing the turnover of transcripts between these stages. For
507 example, when comparing stage 2 hybrids from crosses with *D. mauritiana* and *D.*
508 *simulans* to stage 5 hybrids from the same cross, correlation coefficients range from
509 0.483 to 0.573 (Table S3). The correlation coefficients are lower when comparing
510 transcript abundances of embryos of different stages than when comparing embryos
511 within a stage, indicating that the pool of transcripts present at the maternal stage is
512 different from that at the zygotic stage of development.

513 The above finding is reinforced by principal component analysis (PCA; Figure
514 S1) for RNA expression profiles of the samples in each set of pairwise comparisons
515 between species. We found that the first principal component corresponds to
516 developmental stage and explains between 80.65% and 81.86% of the variance in the
517 three sets of comparisons. The second principal component of this PCA accounts for
518 between 6.94-8.44% of the variance in the three sets of pairwise comparisons between
519 species and corresponds to genotype. This indicates that there is a more substantial
520 difference between pools of transcripts at the maternal and the zygotic stage of
521 development than there is between the pools of transcripts present in embryos of
522 different genotypes (parental species and hybrids) at the same developmental stage.
523

524 **Regulatory changes at the maternal stage of development**

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

529 changes differs based on developmental stage. Changes in *cis* regulation can occur
530 through changes in the DNA of regulatory regions proximal to the gene that they
531 regulate. These types of regulatory changes have an allele-specific effect on gene
532 expression. In contrast, changes in *trans* regulation typically occur via changes in
533 factors that bind to the DNA, such as transcription factor proteins. Changes in *trans*
534 regulation affect the expression of both alleles.

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

547 We found that most regulatory changes underlying differentially maternally
548 deposited transcripts occurred in *trans* between each pair of species examined (Figure
549 3A and C, Figure S2), where a change in a transcription factor or other *trans*-acting
550 regulatory element affects both alleles equally (shown in Figure 4A and B). In all pairs of
551 comparisons between species, the proportion of *trans* changes was higher than any

552 other category of changes. Comparisons between *D. simulans* and *D. mauritiana* had
553 the highest percentage of *trans*-only regulatory changes (between 49.4% and 50.8%)
554 while comparisons between *D. simulans* and *D. sechellia* had a lower percentage of
555 regulatory changes solely in *trans* (32.9%). The second highest proportion (between
556 15.0% and 26.7%) of regulatory changes between species at the maternal stage
557 occurred only in *cis* regulation. Slightly fewer regulatory changes occurred due to a
558 combination of *cis* and *trans* acting factors (between 13.4% and 15.8% in all
559 comparisons). Most genes that change in *cis* and in *trans* regulation are assigned to the
560 *cis* + *trans* category, which indicates that the allele with higher expression in the
561 parental lines is also preserved as the allele with higher expression in the hybrid (the
562 changes in *cis* and in *trans* affect gene expression in the same direction; Landry *et al.*
563 2005; McManus *et al.* 2010; Coolon *et al.* 2014). We found a smaller proportion of
564 genes changed in regulation through *cis* x *trans* interactions, where changes in *cis* and
565 in *trans* have opposing effects on gene expression and the allele with the lower level of
566 expression in the hybrid is from the parental line with the higher level of expression
567 (Landry *et al.* 2005; McManus *et al.* 2010; Coolon *et al.* 2014). We also found a
568 percentage of genes with conserved levels of maternal transcript deposition between
569 species, between 16.4% and 25.1% in all crosses. *D. simulans* and *D. sechellia* have
570 the highest percentage of conserved genes while *D. simulans* and *D. mauritiana* have
571 the lowest percentage of conserved genes. We also found a small proportion of genes,
572 between 4.2% and 4.7% in all comparisons, that have evolved compensatory
573 mechanisms of regulation, where the genes are not differentially expressed between the
574 parental samples but are differentially expressed in hybrids. This implies that while

575 transcript levels are the same between species, regulatory changes have occurred,
576 which then become visible in the environment of the hybrid.
577 Genes with expression differences due to *trans* regulatory changes include
578 regulators with critical functions in important processes governed by maternal gene
579 products, such as *Cdk1*, a cell-cycle regulator necessary for the rapid cleavage cycles
580 in early development (Farrell and O'Farrell 2014). Examples of individual genes with
581 changes attributed to *trans* regulatory differences at the maternal stage are represented
582 in Figure 4A and B.
583
584 **Binding sites for chromatin modifiers are enriched in the regulatory regions of**
585 **maternally deposited genes**
586 As *trans* regulatory changes can affect numerous genetic loci, we asked whether
587 there are *trans* regulatory factors that may affect the differential deposition of a number
588 of maternal transcripts between the species studied. For this, we identified binding sites
589 in the predicted *cis*-regulatory regions of all differentially expressed genes and
590 compared them to identified binding sites in the *cis*-regulatory regions of genes with
591 conserved expression between species at the maternal stage. In the pool of genes with
592 altered expression between species, we included not only those with differences in
593 *trans* regulation, but also those with differences in other regulatory categories (*trans*,
594 *cis*, *cis* + *trans*, *cis* x *trans*, and compensatory). This is because genes with changes in
595 *cis* regulation may have had changes that affect the binding of the same *trans*
596 regulators. For genes with differential expression, and separately for genes with
597 conserved expression between species, we took a computational approach. We used

598 both HOMER and MEME (Bailey and Elkan 1994; Heinz *et al.* 2010), to search for
599 overrepresented motifs within 500 base pairs upstream of transcription start sites, as
600 compared to the rest of the genome (see Methods). We used the upstream regions of
601 genes in *D. simulans* and *D. sechellia* because the *D. mauritiana* genome is not as well
602 annotated, as compared to the other two species in this study.

603 Interestingly, we found that the *cis*-regulatory regions from both genes with
604 conserved and genes with differential transcript levels between species are enriched in
605 motifs associated with insulator binding (Table S4). Specifically, we found that the
606 Dref/BEAF-32 binding site (BEAF-32 and Dref bind overlapping DNA sequences; Hart
607 *et al.* 1999) is the most significantly enriched (Table S4). These factors are annotated
608 as insulators (Matzat and Lei 2014; Ali *et al.* 2016) and known to be associated with
609 topologically associated domains (TADs) (Liang *et al.* 2014; Ramírez *et al.* 2018). The
610 binding site for M1BP also appeared significantly enriched in both sets of genes that
611 change in regulation and in ones that are conserved in regulation across species (Table
612 S4). M1BP is involved in transcriptional regulation and RNA polymerase II pausing at
613 the promoter of genes (Li and Gilmour 2013), which may also be associated with
614 regulating chromatin state (Ramírez *et al.* 2018). Our findings are consistent with
615 previous studies that identified the enrichment of binding sites for M1BP, BEAF-32 and
616 Dref in the promoter regions of genes that are maternally deposited (Chen *et al.* 2013;
617 Omura and Lott 2020). These binding sites have also been associated with
618 housekeeping genes (Zabidi *et al.* 2015), and the pool of maternal transcripts is
619 enriched with housekeeping genes (Liu *et al.* 2014). However, maternal genes that are
620 not housekeeping genes are even more highly enriched for these binding sites (Omura

621 and Lott 2020), thus the involvement of insulators or other chromatin regulators in
622 maternal regulation is unlikely due only to the inclusion of housekeeping genes among
623 maternal genes. Transcript abundance data from our study indicates that *Dref*, *BEAF-*
624 *32*, and *M1BP* are differentially maternally deposited in several between-species
625 comparisons (Figure S4), although in certain crosses, hybrid reads mapped in a biased
626 way to *Dref*, *BEAF-32* and *M1BP*, and thus they were excluded from our regulatory
627 analysis. As the motifs for these *trans*-acting factors are enriched in the upstream
628 regions of all maternal genes, relative to upstream regions of all annotated genes, they
629 are likely important regulators of transcription during oogenesis, and therefore also likely
630 targets of regulatory evolution between species.
631
632 **Evolution of regulation for zygotically expressed genes**
633 To determine the regulatory basis of changes in zygotic transcript abundance
634 between species, we compared expression levels in late stage 5 parental species
635 samples to late stage 5 hybrid samples and used PG models to identify *cis* and *trans*
636 regulatory changes, similar to our maternal analysis (see Methods). We limited our
637 analysis at the zygotic stage to those genes that are mostly-zygotic: zygotically
638 expressed but not maternally deposited (zygotic-only) or expressed at the zygotic stage
639 at an 8-fold higher level when compared to the maternal stage (we will refer to these as
640 mostly zygotic genes, see Methods).
641 While we found that most gene expression changes at the maternal stage of
642 development are due to changes in *trans* regulation between the three sister species,
643 we see strikingly different patterns of regulatory changes after ZGA. At the zygotic

644 stage, differences in gene expression between the three species examined occur
645 mostly due to regulatory changes in both *cis* and *trans*, either by *cis* + *trans* or *cis* x
646 *trans* interactions (Figure 3B and D, Figure S3). Changes in both *cis* and *trans*
647 regulatory elements (either *cis* + *trans* or *cis* x *trans* interactions) account for expression
648 differences in 39% to 47% of zygotic genes at stage 5 in our between-species
649 comparisons. We also see a higher proportion of these interactions occurring in a *cis* +
650 *trans* pattern (between 29% and 35% of all genes) as opposed to a *cis* x *trans* pattern
651 (between 9% and 12% of all genes) of regulatory interactions. In contrast, *cis*-only and
652 *trans*-only changes account for a smaller number of differences in gene expression
653 levels at this stage in development. In all comparisons, we found between 15% and
654 21% of genes changing only in *trans* regulation. There are between 16% and 30% of
655 genes that change only in *cis* regulation between each pair of species compared at this
656 stage in development. Compared to the maternal stage, we found a larger proportion of
657 genes with compensatory changes (between 7% and 10% of all genes) in gene
658 regulation and a smaller proportion of genes that are conserved (between 6% and 8%
659 of all genes) between each pair of species comparisons. The smaller proportion of
660 genes with conserved transcript levels at the zygotic stage compared to the maternal
661 stage is consistent with earlier findings showing maternal transcripts to be more highly
662 conserved between species than zygotic transcripts (Atallah and Lott 2018). Examples
663 of evolved changes include regulators critical to important early zygotic processes, such
664 as gap gene *Kruppel* and pair-rule gene *sloppy paired 1*, which are required for
665 segmentation along the anterior-posterior axis (Nüsslein-Volhard and Wieschaus 1980;
666 Grossniklaus *et al.* 1992) (Figure 4C and D).

667 Transcriptional regulation at the maternal stage may be broadly determined by
668 regulation at the level of chromatin, as evidenced in this work and by another study
669 (Omura and Lott 2020). In contrast, regulation at the zygotic stage can be gene or
670 pathway specific and involve transcription only in a spatially localized subset of cells
671 (Jäckle *et al.* 1986; Johnston and Nüsslein-Volhard 1992). As such, if a *trans* regulator
672 changed at the zygotic stage, it may affect genes involved in a specific developmental
673 process. For these reasons, we wanted to ask if genes whose zygotic expression
674 differed between species due to changes solely in *trans* regulation had a specific
675 molecular function or were part of a particular biological process. We used PANTHER
676 (Mi *et al.* 2019) to perform gene ontology (GO) analysis on genes changing only in *trans*
677 regulation in each pairwise comparison of species at stage 5 (see Methods). Identifying
678 GO categories over multiple crosses identifies the types of genes that evolve changes
679 repeatedly over evolution. Shared categories were broad, and included those related to
680 DNA binding, positive regulation of transcription by RNA polymerase II, cell fate
681 determination, and several developmental categories (Figure 5). The range in GO
682 categories represented, while broadly important at this developmental timepoint,
683 demonstrate how genes changing in *trans* are distributed across developmental
684 processes. As may be expected for zygotic genes at this stage in development, this
685 finding suggests that changes in *trans* regulators of zygotic genes can affect a broad
686 range of molecular and developmental processes. We also investigated biological
687 process categories unique to each specific cross, these are represented in Figure S5
688 and Table S5.
689

690 **Modes of Inheritance in Hybrids**

691 Misexpression in hybrid offspring has been used to examine regulatory
692 incompatibilities that may contribute to speciation (Michalak 2003; Ortíz-Barrientos *et al.*
693 2006; Moehring *et al.* 2007; Mack *et al.* 2016). As the maternal and zygotic transcripts
694 examined during embryogenesis showed different patterns of gene expression
695 evolution, we also asked whether there were more hybrid incompatibilities present at
696 one stage than the other by looking at whether these two developmental stages showed
697 different levels of transcript misexpression. One way to identify misexpression in hybrids
698 is to compare the inheritance of transcript levels in hybrids to transcript levels in each
699 parental species. Here, we quantify the total transcript abundance for a gene by
700 summing the levels of both species-specific alleles in the hybrid and comparing this
701 level to the transcript abundance in both parental lines. We used methods developed by
702 Gibson, *et al*., 2004 to define modes of inheritance in our hybrids; as in previous studies
703 (Gibson *et al.* 2004; Landry *et al.* 2005; McManus *et al.* 2010; Coolon *et al.* 2014), we
704 used a conservative fold change of 1.25 ($\log_2$-fold change of 0.32) to define those
705 genes that do not change in the total transcript abundance between genotypes
706 (conserved). Genes with transcript levels that are higher (overdominant) or lower
707 (underdominant) in the hybrid relative to either parental species are categorized as
708 misexpressed. The total transcript abundance in the hybrids can also be more similar to
709 one parent versus the other. Here, we categorize the parental line with expression most
710 similar to the hybrid as the dominant parent. Expression in the hybrid can also have a
711 level intermediate to both parental species (additive).

712 Both the maternal and zygotic stages show a significant proportion of genes
713 where one species' allele is dominant (Figure 6, Table S6). In both stages, the species
714 that is dominant in each set of crosses is consistent. A higher proportion of genes that
715 are dominant have *D. simulans*-like expression (in any cross involving *D. simulans*) in
716 comparison to the proportion that have expression more like the other parental line in
717 the cross (see Table S6 for percentages). We found that *D. mauritiana* has the least
718 dominance in any cross involving this species. Taken together, our findings indicate that
719 *D. simulans* has the most dominant effect on gene expression at both developmental
720 stages, while *D. mauritiana* has the least dominant effect, with dominance in *D.*
721 *sechellia* falling between the other two species. While there has been previous work
722 proposing relationships between the proportion of dominance and the physiology of
723 unique species (McManus *et al.* 2010), it is difficult to determine any known factors
724 between these three species that would predict this pattern of relative dominance
725 (species range, effective population size, egg size/maternal investment). It is, however,
726 interesting that while the proportion of changes in *cis* and *trans* vary considerably
727 between stages that both stages have dominance among the largest categories of
728 modes of inheritance, and that the relative patterns of which species are dominant is
729 conserved.
730 Strikingly, while many genes show conservation of expression levels between
731 parental species and in the hybrids at both developmental stages, we found a much
732 higher percentage of conserved transcript levels between parents and hybrids for genes
733 that are maternally deposited (Figure 6, Table S6). We found a high proportion of genes
734 with conserved transcript levels at stage 2 in all crosses, between 15.4% and 31.4% of

735 all genes. In contrast, in stage 5 crosses we found conserved transcript abundance in
736 between 4% and 8% of all genes that are either zygotic-only or are mostly zygotic (see
737 Methods for definitions). While there is a large difference in the percentage of
738 conserved genes between the two stages, our stage 5 analysis is limited to those genes
739 with much higher expression at the zygotic stage in comparison to the maternal stage of
740 development. There may be more genes that are mostly zygotic or zygotic-only that are
741 misregulated at this stage in development relative to all of the genes that are expressed
742 at stage 5.

743 In contrast to maternal expression patterns, we found more genes that have an
744 additive mode of inheritance or that are misexpressed in the hybrids for zygotic genes
745 (Figure 6, Table S6). Previous studies indicate that additive inheritance is associated
746 with *cis* regulatory divergence (Lemos *et al.* 2008; McManus *et al.* 2010). This is
747 consistent with our findings that a larger proportion of genes at the zygotic stage have
748 expression divergence due, in part or wholly, to *cis* regulatory changes and that more
749 zygotic genes show an additive pattern of inheritance. Higher levels of misexpression at
750 the zygotic stage, taken together with lower conservation of transcript levels at the
751 zygotic stage, suggests that zygotic genes may contribute more to genetic
752 incompatibilities than maternal genes.

753

754 **DISCUSSION**

755

756 In this study, we asked whether evolution of gene regulation differs at different
757 developmental stages. We found striking differences in the proportions of *cis* and *trans*

758 regulatory changes between the stage of embryogenesis where all transcripts are
759 maternally derived, and a stage just a few hours later after the zygotic genome has
760 been activated. Between the species examined, we uncovered an overwhelming
761 number of *trans* regulatory changes resulting in differential maternal transcript levels,
762 whereas a complex mix of *cis*, *trans*, and the combination of the two were responsible
763 for changes in zygotic transcription of mostly zygotic genes (see Methods). Here, we
764 propose that the differences in the patterns of gene regulatory evolution between the
765 stages we examined may be due to fundamental differences in the biological context
766 and regulatory architecture producing the transcriptomes present at these stages.
767 Maternal transcripts are produced by support cells called nurse cells during
768 oogenesis and are either transported by microtubule-dependent mechanisms or
769 dumped into the oocyte along with the cytoplasmic contents of the nurse cells upon
770 apoptosis (Kugler and Lasko 2009). Many aspects of maternal provisioning have been
771 well-studied in *D. melanogaster*, including transport of transcripts into the oocyte
772 (Mische *et al.* 2007), localization of transcripts within the oocyte (Theurkauf and
773 Hazelrigg 1998), post-transcriptional regulation of maternal gene products (Salles *et al.*
774 1994), and subsequent degradation of maternal transcripts (Tadros *et al.* 2007; Bushati
775 *et al.* 2008; Laver *et al.* 2015). Surprisingly, how transcription is regulated in the nurse
776 cells is not well understood. Nurse cells are polyploid, and are able to rapidly transcribe
777 a large quantity of RNA that represents a large proportion of the genome (Tadros *et al.*
778 2007; De Renzis *et al.* 2007; Thomsen *et al.* 2010; Lott *et al.* 2011; Vastenhouw *et al.*
779 2019) to provide the oocyte with the large stock of transcripts needed. The oocyte itself
780 is thought to be largely transcriptionally silent (Navarro-Costa *et al.* 2016). What we

781 found here, and what was also found in another study investigating binding motifs in
782 maternal factors across the *Drosophila* genus using computational methods (Omura
783 and Lott 2020), is that maternal transcription is associated with *trans* factors annotated
784 to be insulators and that interact with topologically associated domains (TADs). This
785 provides evidence that maternal transcription may be controlled broadly at the level of
786 chromatin state. In this context, we predict that changes in only a few *trans* factors can
787 be responsible for the bulk of the between-species changes in maternal transcription.
788 Thus, changes in the levels of *trans* regulators at this stage may easily be responsible
789 for changes in transcription level for a number of genes.

790 In contrast to the large proportion of regulatory changes in *trans* at the maternal
791 stage, differences in zygotic gene transcription for genes that are mostly zygotic (see
792 Methods) between these species is predominantly explained by a combination of
793 changes in *cis*, *trans*, *cis* + *trans*, and *cis* x *trans*. Zygotic gene transcription for genes
794 without a maternal complement is fundamentally different than maternal gene
795 transcription. Unlike the bulk transcription that takes place in the nurse cells, zygotic
796 gene transcription is precisely regulated at the spatial and temporal level across the
797 embryo with enhancer regions playing a large role in where, when and how much genes
798 are expressed (Haines and Eisen 2018). Due to these fundamental differences in gene
799 regulation, the embryo at the zygotic stage may be more sensitive to changes in gene
800 expression than at the maternal stage. Specifically, changes in *trans* regulation, which
801 can affect the expression of many genes, may be detrimental to the developing
802 organism at this stage. In contrast, changes in *cis* regulation are gene-specific and may
803 only affect gene expression in a subset of the embryo, which might be a more suitable

804 way of fine-tuning zygotic expression. We propose that fundamental differences in the
805 regulatory landscape and developmental purpose of the maternal versus the zygotic
806 stage likely explain why the evolution of gene expression occurs through different
807 mechanisms for transcripts that are maternally deposited and genes that are primarily
808 zygotic.

809 While this study was directed at understanding the regulatory basis of evolution
810 in gene expression at the maternal and zygotic stages of embryogenesis, it also
811 provides insight into the relative conservation of gene expression, both between species
812 and between parent and hybrid offspring. Here, in both the analysis of regulatory
813 changes and the analysis of modes of inheritance, we found more genes with
814 conserved transcript levels among those that are maternally deposited relative to those
815 that are zygotically transcribed. This is in agreement with previous studies that identified
816 high conservation of maternally deposited transcripts relative to those transcribed
817 zygotically between species (Heyn *et al.* 2014; Atallah and Lott 2018) and indicates that
818 the maternal stage is highly conserved. We observe lower conservation of transcript
819 levels at the zygotic stage. A caveat our gene expression analysis at the zygotic stage
820 is that we had to remove genes that still have a large maternal component at this stage
821 (roughly 50% of total transcript pool at late stage 5 is maternally derived; Lott *et al.*
822 2014). Thus, our finding is best viewed as genes whose transcripts are primarily zygotic
823 at stage 5, have a higher rate of evolutionary change. Additionally, the large proportion
824 of genes with conserved transcript levels at the maternal stage may be unexpected
825 considering that there is substantial post-transcriptional regulation of maternally
826 deposited factors (Tadros *et al.* 2007; Rouget *et al.* 2010; Barckmann and Simonelig

827 2013), so it is not clear that a high degree of conservation at the transcript level should
828 be necessary to maintaining conservation at the protein level. Alternatively, if the
829 maternal genome is primarily regulated at the level of chromatin state, this may be a
830 mechanistic constraint on evolution at the level of gene expression. It may be
831 functionally difficult for a gene located in a region of open chromatin to be repressed, or
832 for a gene in a region of heterochromatin to gain expression. Thus, it may be easier to
833 evolve differences in expression over evolutionary time via post-transcriptional
834 mechanisms for maternal genes. Further study is needed to disentangle conservation at
835 the transcript and protein levels of maternal factors across species.

836 In addition to the differences in conservation between stages, we also found
837 differences in the patterns of inheritance of gene expression between species at the
838 maternal and zygotic stages of embryogenesis. The zygotic stage has a larger
839 proportion of additive differences, which some previous theory (Gibson *et al.* 2004) and
840 empirical studies (Lemos *et al.* 2008; McManus *et al.* 2010) have suggested may be
841 more likely to be changes in *cis* regulation. This would be consistent both with the
842 increased relative role of *cis* changes at the zygotic stage compared to the maternal
843 stage found here, as well as what is known about zygotic gene regulation more
844 generally (Mannervik 2014). In addition, a larger proportion of changes at the zygotic
845 stage fall into the broad category of misregulation (underdominant, overdominant),
846 which have been proposed to increase with divergence time (Coolon *et al.* 2014) and
847 may be a potential source of hybrid incompatibility between species (Michalak 2003;
848 Ortíz-Barrientos *et al.* 2006; Moehring *et al.* 2007; Mack *et al.* 2016).

849 In this study, we found that differences between species in levels of maternally
850 deposited transcripts and zygotically transcribed genes evolve via different patterns of
851 regulatory change. We found that maternal transcript abundance is more conserved but
852 when changes do occur, they occur more frequently through *trans* regulation in
853 comparison to zygotic complements. Regulatory organization, constraints, and
854 developmental processes that are specific to each developmental stage likely play a
855 large role in determining how gene regulation can evolve at these two embryonic
856 timepoints. Further study is needed to characterize the molecular basis of evolved
857 changes in transcript level on a single gene level, and more generally to determine what
858 is controlling the regulatory landscape at each stage in development.
859
860
861 **ACKNOWLEDGMENTS**
862
863 We would like to thank members of the Lott lab for comments on the manuscript, as well
864 as members of the UC Davis fly community for their contributions. We would also like to
865 thank Amanda Crofton and Anthony Le for their assistance with genotyping and RNA
866 extraction and Graham Coop for helpful discussions. This work was supported by the
867 National Institute of General Medical Sciences of the National Institutes of Health grant
868 R01GM111362 and the Floyd and Mary Schwall Fellowship.
869
870

871 **REFERENCES**

872

873 Ali T., R. Renkawitz, and M. Bartkuhn, 2016 Insulators and domains of gene
874 expression. Current Opinion in Genetics & Development 37: 17–26.
875 https://doi.org/10.1016/j.gde.2015.11.009
876 Ali-Murthy Z., S. E. Lott, M. B. Eisen, and T. B. Kornberg, 2013 An essential role for
877 zygotic expression in the pre-cellular Drosophila embryo. PLoS genetics 9:
878 e1003428. https://doi.org/10.1371/journal.pgen.1003428
879 Ashburner M., 1989 *Drosophila: A Laboratory Handbook*(C. spring harbor laboratory
880 press, Ed.).
881 Atallah J., and S. E. Lott, 2018 Evolution of maternal and zygotic mRNA complements
882 in the early Drosophila embryo, (K. A. Dyer, Ed.). PLoS genetics 14: e1007838.
883 https://doi.org/10.1371/journal.pgen.1007838
884 Bailey T. L., and C. Elkan, 1994 Fitting a mixture model by expectation maximization to
885 discover motifs in bipolymers
886 Barckmann B., and M. Simonelig, 2013 Control of maternal mRNA stability in germ cells
887 and early embryos. Biochimica et Biophysica Acta (BBA) - Gene Regulatory
888 Mechanisms 1829: 714–724. https://doi.org/10.1016/j.bbagrm.2012.12.011
889 Baroux C., D. Autran, C. S. Gillmor, D. Grimanelli, and U. Grossniklaus, 2008 The
890 Maternal to Zygotic Transition in Animals and Plants. Cold Spring Harbor Symposia
891 on Quantitative Biology 73: 89–100. https://doi.org/10.1101/sqb.2008.73.053
892 Bownes M., 1975 A photographic study of development in the living embryo of
893 Drosophila melanogaster. Development 33: 789–801.

894 Britten R. J., and E. H. Davidson, 1969 Gene Regulation for Higher Cells - a Theory.
895 Science (New York, N.Y.) 165: 349-+. https://doi.org/10.1126/science.165.3891.349
896 Buchberger, Reis, Lu, and Posnien, 2019 Cloudy with a Chance of Insights: Context
897 Dependent Gene Regulation and Implications for Evolutionary Studies. Genes 10:
898 492–21. https://doi.org/10.3390/genes10070492
899 Bushati N., A. Stark, J. Brennecke, and S. M. Cohen, 2008 Temporal reciprocity of
900 miRNAs and their targets during the maternal-to-zygotic transition in Drosophila.
901 Current Biology 18: 501–506. https://doi.org/10.1016/j.cub.2008.02.081
902 Campos-Ortega J. A., and V. Hartenstein, 2013 *The Embryonic Development of*
903 *Drosophila melanogaster*. Springer Science & Business Media.
904 Carroll S. B., 1995 Homeotic Genes and the Evolution of Arthropods and Chordates.
905 Nature 376: 479–485. https://doi.org/10.1038/376479a0
906 Chan Y. F., M. E. Marks, F. C. Jones, G. Villarreal, M. D. Shapiro, *et al.*, 2010 Adaptive
907 Evolution of Pelvic Reduction in Sticklebacks by Recurrent Deletion of a Pitx1
908 Enhancer. Science (New York, N.Y.) 327: 302–305.
909 https://doi.org/10.1126/science.1182213
910 Chen K., J. Johnston, W. Shao, S. Meier, C. Staber, *et al.*, 2013 A global change in
911 RNA polymerase II pausing during the Drosophila midblastula transition. Elife 2:
912 e00861. https://doi.org/10.7554/elife.00861
913 Clark A. G., M. B. Eisen, D. R. Smith, C. M. Bergman, B. Oliver, *et al.*, 2007 Evolution of
914 genes and genomes on the Drosophila phylogeny. Nature 450: 203–218.
915 https://doi.org/10.1038/nature06341

916 Coolon J. D., C. J. McManus, K. R. Stevenson, B. R. Graveley, and P. J. Wittkopp,
917 2014 Tempo and mode of regulatory evolution in Drosophila. Genome Research 24:
918 797–808. https://doi.org/10.1101/gr.163014.113
919 Crofton A. E., E. L. Cartwright, A. A. Feitzinger, and S. E. Lott, 2018 Effect of Larval
920 Nutrition on Maternal mRNA Contribution to the Drosophila Egg. G3:
921 Genes|Genomes|Genetics 8: 1933–1941. https://doi.org/10.1534/g3.118.200283
922 Davidson E. H., and I. S. Peter, 2015 *Genomic control process*.
923 De Renzis S. D., O. Elemento, S. Tavazoie, and E. F. Wieschaus, 2007 Unmasking
924 Activation of the Zygotic Genome Using Chromosomal Deletions in the Drosophila
925 Embryo. PLOS Biology 5: e117. https://doi.org/10.1371/journal.pbio.0050117
926 Eichhorn S. W., A. O. Subtelny, I. Kronja, and J. K. Elife, 2016 mRNA poly (A)-tail
927 changes specified by deadenylation broadly reshape translation in Drosophila
928 oocytes and early embryos. cdn.elifesciences.org.
929 https://doi.org/10.7554/elife.16955.001
930 Farrell J. A., and P. H. O'Farrell, 2014 From Egg to Gastrula: How the Cell Cycle Is
931 Remodeled During the Drosophila Mid-Blastula Transition. Annual Review of
932 Genetics, Vol 48 48: 269–294. https://doi.org/10.1146/annurev-genet-111212-
933 133531
934 Fischer B. E., E. Wasbrough, L. A. Meadows, O. Randlet, S. Dorus, *et al.*, 2012
935 Conserved properties of Drosophila and human spermatozoal mRNA repertoires.
936 Proc Royal Soc B Biological Sci 279: 2636–2644.
937 https://doi.org/10.1098/rspb.2012.0153

938 Gibson G., R. Riley-Berger, L. Harshman, A. Kopp, S. Vacha, *et al.*, 2004 Extensive
939 Sex-Specific Nonadditivity of Gene Expression in Drosophila melanogaster. Genetics
940 167: 1791–1799. https://doi.org/10.1534/genetics.104.026583
941 Glaser-Schmitt A., A. Zečić, and J. Parsch, 2018 Gene Regulatory Variation in
942 Drosophila melanogaster Renal Tissue. Genetics genetics.301073.2018-46.
943 https://doi.org/10.1534/genetics.118.301073
944 Graze R. M., L. M. McIntyre, B. J. Main, M. L. Wayne, and S. V. Nuzhdin, 2009
945 Regulatory Divergence in Drosophila melanogaster and D. simulans, a Genomewide
946 Analysis of Allele-Specific Expression. Genetics 183: 547–561.
947 https://doi.org/10.1534/genetics.109.105957
948 Grossniklaus U., R. K. Pearson, and W. J. Gehring, 1992 The Drosophila sloppy paired
949 locus encodes two proteins involved in segmentation that show homology to
950 mammalian transcription factors. Genes & Development 6: 1030–1051.
951 https://doi.org/10.1101/gad.6.6.1030
952 Haines J. E., and M. B. Eisen, 2018 Patterns of chromatin accessibility along the
953 anterior-posterior axis in the early Drosophila embryo, (J. M. Vaquerizas, Ed.). PLoS
954 genetics 14: e1007367-25. https://doi.org/10.1371/journal.pgen.1007367
955 Hart C. M., O. Cuvier, and U. K. Laemmli, 1999 Evidence for an antagonistic
956 relationship between the boundary element-associated factor BEAF and the
957 transcription factor DREF. Chromosoma 108: 375–383.
958 https://doi.org/10.1007/s004120050389
959 Heinz S., C. Benner, N. Spann, E. Bertolino, Y. C. Lin, *et al.*, 2010 Simple Combinations
960 of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements

961 Required for Macrophage and B Cell Identities. Molecular cell 38: 576–589.
962 https://doi.org/10.1016/j.molcel.2010.05.004
963 Heyn P., M. Kircher, A. Dahl, J. Kelso, P. Tomancak, *et al.*, 2014 The Earliest
964 Transcribed Zygotic Genes Are Short, Newly Evolved, and Different across Species.
965 CellReports 6: 285–292. https://doi.org/10.1016/j.celrep.2013.12.030
966 Jäckle H., D. Tautz, R. Schuh, E. Seifert, and R. Lehmann, 1986 Cross-regulatory
967 interactions among the gap genes of Drosophila. Nature 324: 668–670.
968 https://doi.org/10.1038/324668a0
969 Johnston D. S., and C. Nüsslein-Volhard, 1992 The Origin of Pattern and Polarity in the
970 Drosophila Embryo. Cell 68: 201–219. https://doi.org/10.1016/0092-8674(92)90466-p
971 King M. C., and A. C. Wilson, 1975 Evolution at two levels in humans and chimpanzees.
972 Science (New York, N.Y.) 188: 107–116. https://doi.org/10.1126/science.1090005
973 Kugler J.-M., and P. Lasko, 2009 Localization, anchoring and translational control of
974 oskar, gurken, bicoid and nanos mRNA during Drosophila oogenesis. Fly 3: 15–28.
975 https://doi.org/10.4161/fly.3.1.7751
976 Kvon E. Z., O. K. Kamneva, U. S. Melo, I. Barozzi, M. Osterwalder, *et al.*, 2016
977 Progressive Loss of Function in a Limb Enhancer during Snake Evolution. Cell 167:
978 633-642.e11. https://doi.org/10.1016/j.cell.2016.09.028
979 Lachaise D., J. R. David, F. Lemeunier, L. Tsacas, and M. Ashburner, 1986 The
980 Reproductive Relationships of Drosophila sechellia with D. mauritiana, D. simulans,
981 and D. melanogaster from the Afrotropical Region. Evolution 40: 262–271.
982 https://doi.org/10.1111/j.1558-5646.1986.tb00468.x

983 Landry C. R., P. J. Wittkopp, C. H. Taubes, J. M. Ranz, A. G. Clark, *et al.*, 2005
984 Compensatory cis-trans evolution and the dysregulation of gene expression in
985 interspecific hybrids of Drosophila. Genetics 171: 1813–1822.
986 https://doi.org/10.1534/genetics.105.047449
987 Lasko P., 2012 mRNA Localization and Translational Control in Drosophila Oogenesis.
988 Cold Spring Harbor Perspectives in Biology 4: a012294–a012294.
989 https://doi.org/10.1101/cshperspect.a012294
990 Laver J. D., X. Li, D. Ray, K. B. Cook, N. A. Hahn, *et al.*, 2015 Brain tumor is a
991 sequence-specific RNA-binding protein that directs maternal mRNA clearance during
992 the Drosophila maternal-to-zygotic transition. Genome biology 16: 94.
993 https://doi.org/10.1186/s13059-015-0659-4
994 Lemos B., L. O. Araripe, P. Fontanillas, and D. L. Hartl, 2008 Dominance and the
995 evolutionary accumulation of cis- and trans-effects on gene expression. Proceedings
996 of the National Academy of Sciences 105: 14471–14476.
997 https://doi.org/10.1073/pnas.0805160105
998 León-Novelo L. G., L. M. McIntyre, J. M. Fear, and R. M. Graze, 2014 A flexible
999 Bayesian method for detecting allelic imbalance in RNA-seq data. BMC Genomics
1000 15: 920. https://doi.org/10.1186/1471-2164-15-920
1001 Li H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, *et al.*, 2009 The Sequence
1002 Alignment/Map format and SAMtools. Bioinformatics 25: 2078–2079.
1003 https://doi.org/10.1093/bioinformatics/btp352

1004 Li J., and D. S. Gilmour, 2013 Distinct mechanisms of transcriptional pausing
1005 orchestrated by GAGA factor and M1BP, a novel transcription factor. The EMBO
1006 Journal 32: 1829–1841. https://doi.org/10.1038/emboj.2013.111
1007 Liang J., L. Lacroix, A. Gamot, S. Cuddapah, S. Queille, *et al.*, 2014 Chromatin
1008 Immunoprecipitation Indirect Peaks Highlight Long-Range Interactions of Insulator
1009 Proteins and Pol II Pausing. Molecular cell 53: 672–681.
1010 https://doi.org/10.1016/j.molcel.2013.12.029
1011 Lim J., M. Lee, A. Son, H. Chang, and V. N. Kim, 2016 mTAIL-seq reveals dynamic
1012 poly(A) tail regulation in oocyte-to-embryo development. Genes & Development 30:
1013 1671–1682. https://doi.org/10.1101/gad.284802.116
1014 Liu M. M., J. W. Davey, D. J. Jackson, M. L. Blaxter, and A. Davison, 2014 A conserved
1015 set of maternal genes? Insights from a molluscan transcriptome. Int J Dev Biology
1016 58: 501–511. https://doi.org/10.1387/ijdb.140121ad
1017 Lott S. E., J. E. Villalta, G. P. Schroth, S. Luo, L. A. Tonkin, *et al.*, 2011 Noncanonical
1018 Compensation of Zygotic X Transcription in Early Drosophila melanogaster
1019 Development Revealed through Single-Embryo RNA-Seq. PLOS Biology 9:
1020 e1000590. https://doi.org/10.1371/journal.pbio.1000590
1021 Lott S. E., J. E. Villalta, Q. Zhou, D. Bachtrog, and M. B. Eisen, 2014 Sex-specific
1022 embryonic gene expression in species with newly evolved sex chromosomes., (H. S.
1023 Malik, Ed.). PLoS genetics 10: e1004159.
1024 https://doi.org/10.1371/journal.pgen.1004159

1025 Mack K. L., P. Campbell, and M. W. Nachman, 2016 Gene regulation and speciation in
1026 house mice. Genome Research gr.195743.115.
1027 https://doi.org/10.1101/gr.195743.115
1028 Macmanes M. D., 2014 On the optimal trimming of high-throughput mRNA sequence
1029 data. Frontiers in genetics 5: 13. https://doi.org/10.3389/fgene.2014.00013
1030 Mannervik M., 2014 Control of Drosophila embryo patterning by transcriptional co-
1031 regulators. Experimental Cell Research 321: 47–57.
1032 https://doi.org/10.1016/j.yexcr.2013.10.010
1033 Martin M., 2011 Cutadapt removes adapter sequences from high-throughput
1034 sequencing reads. EMBnet.journal 17: 10–12. https://doi.org/10.14806/ej.17.1.200
1035 Matzat L. H., and E. P. Lei, 2014 Surviving an identity crisis: A revised view of
1036 chromatin insulators in the genomics era. Biochimica et Biophysica Acta (BBA) -
1037 Gene Regulatory Mechanisms 1839: 203–214.
1038 https://doi.org/10.1016/j.bbagrm.2013.10.007
1039 McDermott S. R., and R. M. Kliman, 2008 Estimation of Isolation Times of the Island
1040 Species in the Drosophila simulans Complex from Multilocus DNA Sequence Data.
1041 PLoS ONE 3: e2442. https://doi.org/10.1371/journal.pone.0002442
1042 McManus C. J., J. D. Coolon, M. O. Duff, J. Eipper-Mains, B. R. Graveley, *et al.*, 2010
1043 Regulatory divergence in Drosophila revealed by mRNA-seq. Genome Research 20:
1044 816–825. https://doi.org/10.1101/gr.102491.109
1045 Mi H., A. Muruganujan, X. Huang, D. Ebert, C. Mills, *et al.*, 2019 Protocol Update for
1046 large-scale genome and gene function analysis with the PANTHER classification
1047 system (v.14.0). Nature Protocols 1–21. https://doi.org/10.1038/s41596-019-0128-8

1048 Michalak P., 2003 Genome-Wide Patterns of Expression in Drosophila Pure Species
1049 and Hybrid Males. Molecular Biology and Evolution 20: 1070–1076.
1050 https://doi.org/10.1093/molbev/msg119
1051 Mische S., M. Li, M. Serr, and T. S. Hays, 2007 Direct Observation of Regulated
1052 Ribonucleoprotein Transport Across the Nurse Cell/Oocyte Boundary, (S. Wente,
1053 Ed.). Molecular Biology of the Cell 18: 2254–2263. https://doi.org/10.1091/mbc.e06-
1054 10-0959
1055 Moehring A. J., K. C. Teeter, and M. A. F. Noor, 2007 Genome-Wide Patterns of
1056 Expression in Drosophila Pure Species and Hybrid Males. II. Examination of Multiple-
1057 Species Hybridizations, Platforms, and Life Cycle Stages. Molecular Biology and
1058 Evolution 24: 137–145. https://doi.org/10.1093/molbev/msl142
1059 Navarro-Costa P., A. McCarthy, P. Prudêncio, C. Greer, L. G. Guilgur, *et al.*, 2016 Early
1060 programming of the oocyte epigenome temporally controls late prophase I
1061 transcription and chromatin remodelling. Nat Commun 7: 12331.
1062 https://doi.org/10.1038/ncomms12331
1063 Nolte V., R. V. Pandey, R. Kofler, and C. Schlötterer, 2013 Genome-wide patterns of
1064 natural variation reveal strong selective sweeps and ongoing genomic conflict in
1065 Drosophila mauritiana. Genome Research 23: 99–110.
1066 https://doi.org/10.1101/gr.139873.112
1067 Nüsslein-Volhard C., and E. Wieschaus, 1980 Mutations affecting segment number and
1068 polarity in Drosophila. Nature 287: 795–801. https://doi.org/10.1038/287795a0
1069 Omura C. S., and S. E. Lott, 2020 The conserved regulatory basis of mRNA
1070 contributions to the early Drosophila embryo differs between the maternal and

1071 zygotic genomes. Plos Genet 16: e1008645.
1072 https://doi.org/10.1371/journal.pgen.1008645
1073 Ortíz-Barrientos D., B. A. Counterman, and M. A. F. Noor, 2006 Gene expression
1074 divergence and the origin of hybrid dysfunctions. Genetica 129: 71–81.
1075 https://doi.org/10.1007/s10709-006-0034-1
1076 Paris M., J. E. Villalta, M. B. Eisen, and S. E. Lott, 2015 Sex Bias and Maternal
1077 Contribution to Gene Expression Divergence in Drosophila Blastoderm Embryos., (A.
1078 G. Clark, Ed.). PLoS genetics 11: e1005592.
1079 https://doi.org/10.1371/journal.pgen.1005592
1080 Perry M. W., A. N. Boettiger, and M. Levine, 2011 Multiple enhancers ensure precision
1081 of gap gene-expression patterns in the Drosophila embryo. Proceedings of the
1082 National Academy of Sciences 108: 13570–13575.
1083 https://doi.org/10.1073/pnas.1109873108
1084 Prud'homme B., N. Gompel, A. Rokas, V. A. Kassner, T. M. Williams, *et al.*, 2006
1085 Repeated morphological evolution through cis -regulatory changes in a pleiotropic
1086 gene. Nature 440: 1050–1053. https://doi.org/10.1038/nature04597
1087 Quinlan A. R., and I. M. Hall, 2010 BEDTools: a flexible suite of utilities for comparing
1088 genomic features. Bioinformatics 26: 841–842.
1089 https://doi.org/10.1093/bioinformatics/btq033
1090 R Core Team, 2017 *R: A language and environment for statistical computing*.
1091 Ramírez F., V. Bhardwaj, L. Arrigoni, K. C. Lam, B. A. Grüning, *et al.*, 2018 High-
1092 resolution TADs reveal DNA sequences underlying genome organization in flies.
1093 Nature Communications 9: 1–15. https://doi.org/10.1038/s41467-017-02525-w

1094 Rouget C., C. Papin, A. Boureux, A.-C. Meunier, B. Franco, *et al.*, 2010 Maternal mRNA
1095 deadenylation and decay by the piRNA pathway in the early Drosophila embryo.
1096 Nature 467: 1128-U144. https://doi.org/10.1038/nature09465
1097 Salles F. J., M. E. Lieberfarb, C. Wreden, J. P. Gergen, and S. Strickland, 1994
1098 Coordinate initiation of Drosophila development by regulated polyadenylation of
1099 maternal messenger RNAs. Science (New York, N.Y.) 266: 1996–1999.
1100 https://doi.org/10.1126/science.7801127
1101 Schulz K. N., and M. M. Harrison, 2019 Mechanisms regulating zygotic genome
1102 activation. Nature Reviews Genetics 1–14. https://doi.org/10.1038/s41576-018-0087-
1103 x
1104 Small S., A. Blair, and M. Levine, 1992 Regulation of Even-Skipped Stripe-2 in the
1105 Drosophila Embryo. Embo Journal 11: 4047–4057. https://doi.org/10.1002/j.1460-
1106 2075.1992.tb05498.x
1107 Supek F., M. Bošnjak, N. Škunca, and T. Šmuc, 2011 REVIGO Summarizes and
1108 Visualizes Long Lists of Gene Ontology Terms, (C. Gibas, Ed.). PLoS ONE 6:
1109 e21800-9. https://doi.org/10.1371/journal.pone.0021800
1110 Tadros W., A. L. Goldman, T. Babak, F. Menzies, L. Vardy, *et al.*, 2007 SMAUG Is a
1111 Major Regulator of Maternal mRNA Destabilization in Drosophila and Its Translation
1112 Is Activated by the PAN GU Kinase. Developmental Cell 12: 143–155.
1113 https://doi.org/10.1016/j.devcel.2006.10.005
1114 Tadros W., and H. D. Lipshitz, 2009 The maternal-to-zygotic transition: a play in two
1115 acts. Development 136: 3033–3042. https://doi.org/10.1242/dev.033183

1116 Theurkauf W. E., and T. I. Hazelrigg, 1998 In vivo analyses of cytoplasmic transport and
1117 cytoskeletal organization during Drosophila oogenesis: characterization of a multi-
1118 step anterior localization pathway. Dev Camb Engl 125: 3655–66.
1119 Thomsen S., S. Anders, S. C. Janga, W. Huber, and C. R. Alonso, 2010 Genome-wide
1120 analysis of mRNA decay patterns during early Drosophila development. Genome
1121 biology 11: R93. https://doi.org/10.1186/gb-2010-11-9-r93
1122 Torres-Oliva M., I. Almudi, A. P. McGregor, and N. Posnien, 2016 A robust (re-
1123)annotation approach to generate unbiased mapping references for RNA-seq-based
1124 analyses of differential expression across closely related species. BMC Genomics
1125 17: 392. https://doi.org/10.1186/s12864-016-2646-x
1126 Trapnell C., A. Roberts, L. Goff, G. Pertea, D. Kim, *et al.*, 2012 Differential gene and
1127 transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks.
1128 Nature Protocols 7: 562. https://doi.org/10.1038/nprot.2012.016
1129 Van der Auwera G. A. V. der, M. O. Carneiro, C. Hartl, R. Poplin, G. del Angel, *et al.*,
1130 2013 Current Protocols in Bioinformatics. Curr Protoc Bioinform Ed Board Andreas D
1131 Baxevanis Et Al 43: 11.10.1-11.10.33.
1132 https://doi.org/10.1002/0471250953.bi1110s43
1133 Vastenhouw N. L., W. X. Cao, and H. D. Lipshitz, 2019 The maternal-to-zygotic
1134 transition revisited. Development 146: dev161471-20.
1135 https://doi.org/10.1242/dev.161471
1136 Wittkopp P. J., B. K. Haerum, and A. G. Clark, 2004 Evolutionary changes in cis and
1137 trans gene regulation. Nature 430: 85–88. https://doi.org/10.1038/nature02698

1138 Xu G., J. Lyu, Q. Li, H. Liu, D. Wang, *et al.*, 2020 Adaptive evolution of DNA methylation
1139 reshaped gene regulation in maize. Biorxiv 2020.03.13.991117.
1140 https://doi.org/10.1101/2020.03.13.991117
1141 Zabidi M. A., C. D. Arnold, K. Schernhuber, M. Pagani, M. Rath, *et al.*, 2015 Enhancer–
1142 core-promoter specificity separates developmental and housekeeping gene
1143 regulation. Nature 518: 556–559. https://doi.org/10.1038/nature13994
1144

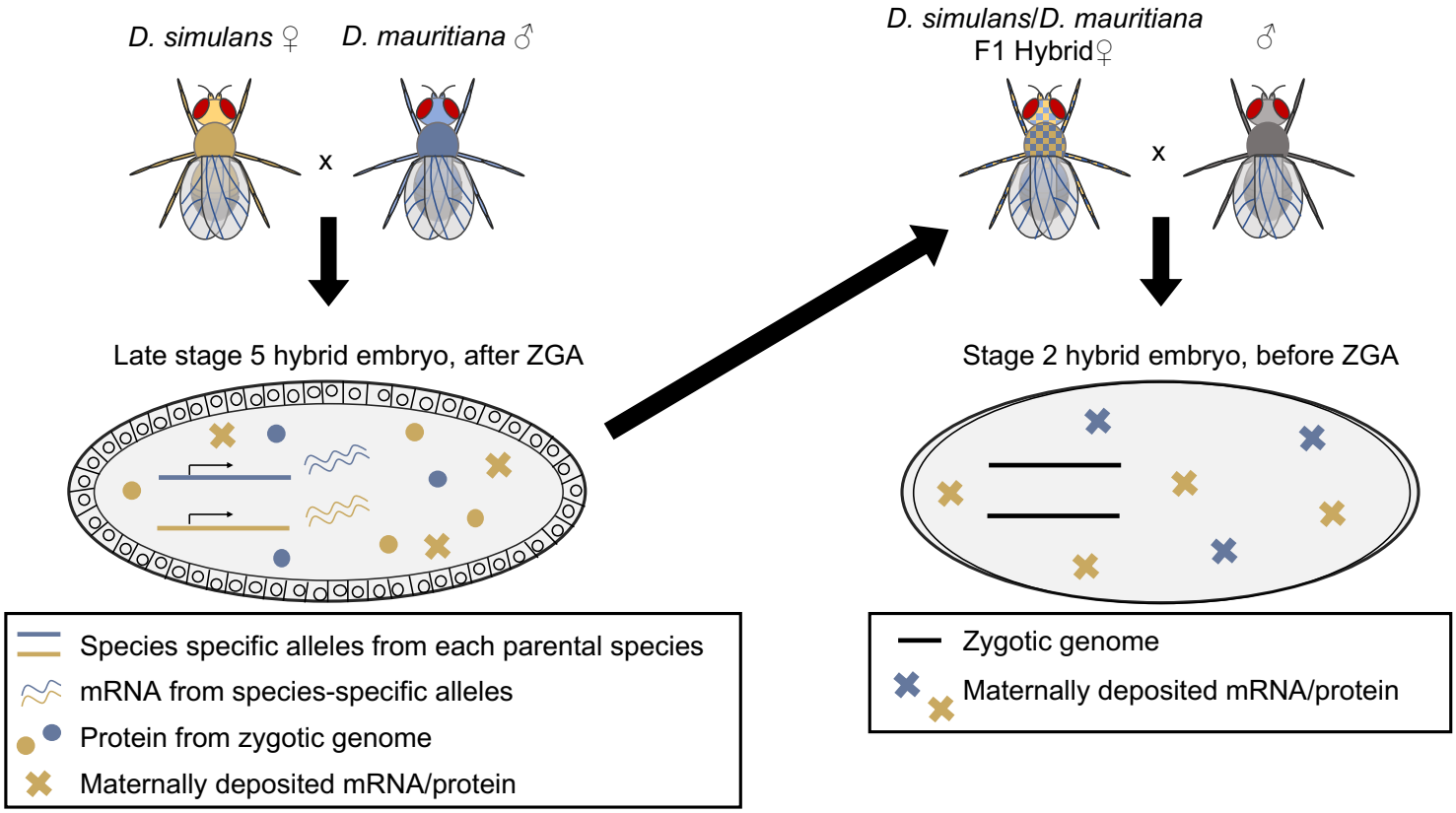


Figure 1: Crosses to produce hybrid embryos for the zygotic and maternal stages. To look at changes in regulation for zygotic genes, hybrid stage 5 embryos (left) were produced by crossing two parental species and collecting their eggs at the appropriate stage (late stage 5). To look at regulatory changes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

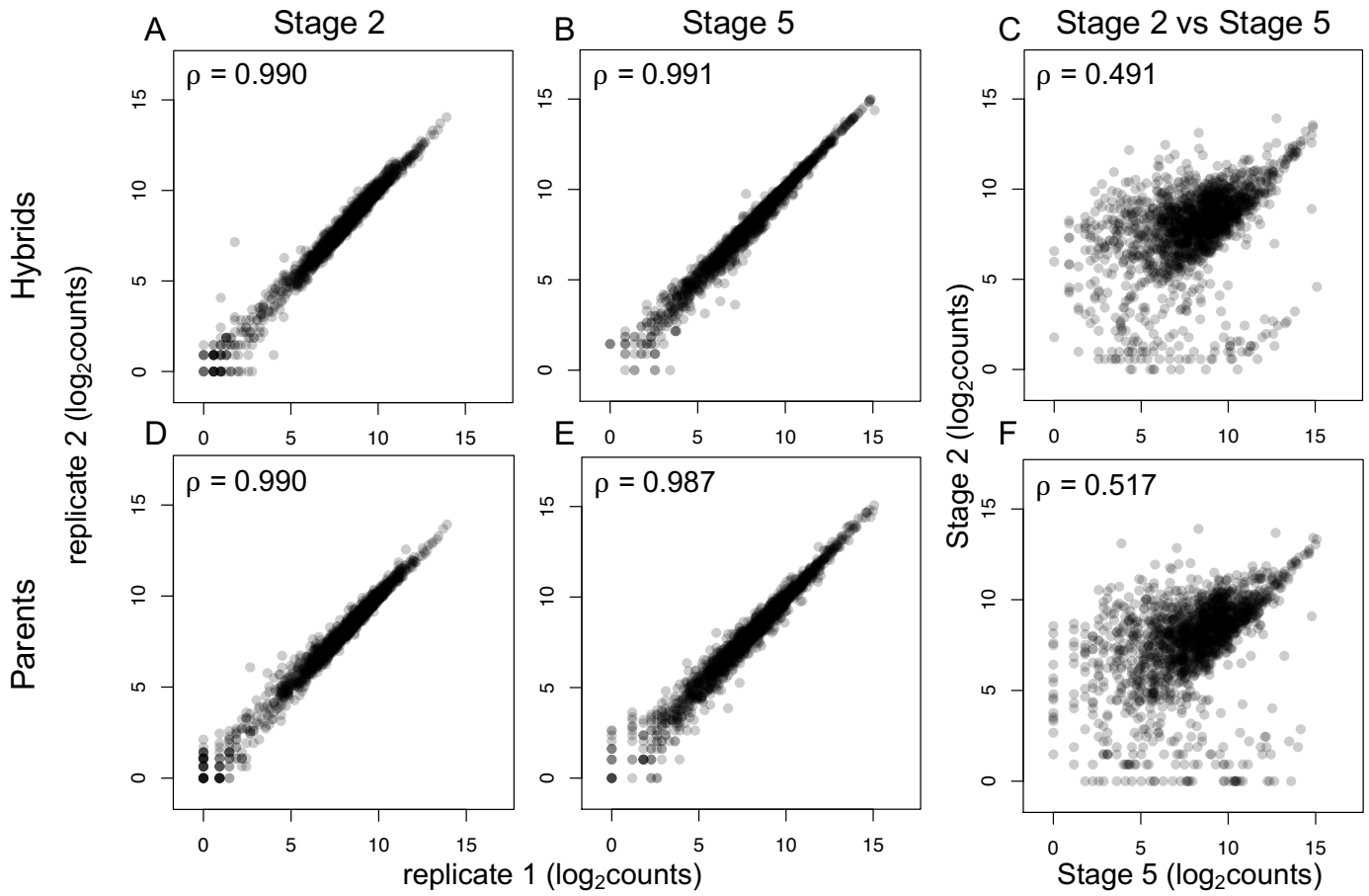


Figure 2: Hybrid and parental species single embryo transcript levels are highly reproducible. (A,B,D,E) Spearman's rank correlation coefficients are high when counts from replicate transcriptomes of the same stage and genotype are compared. Correlation coefficients are similarly high in parental species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

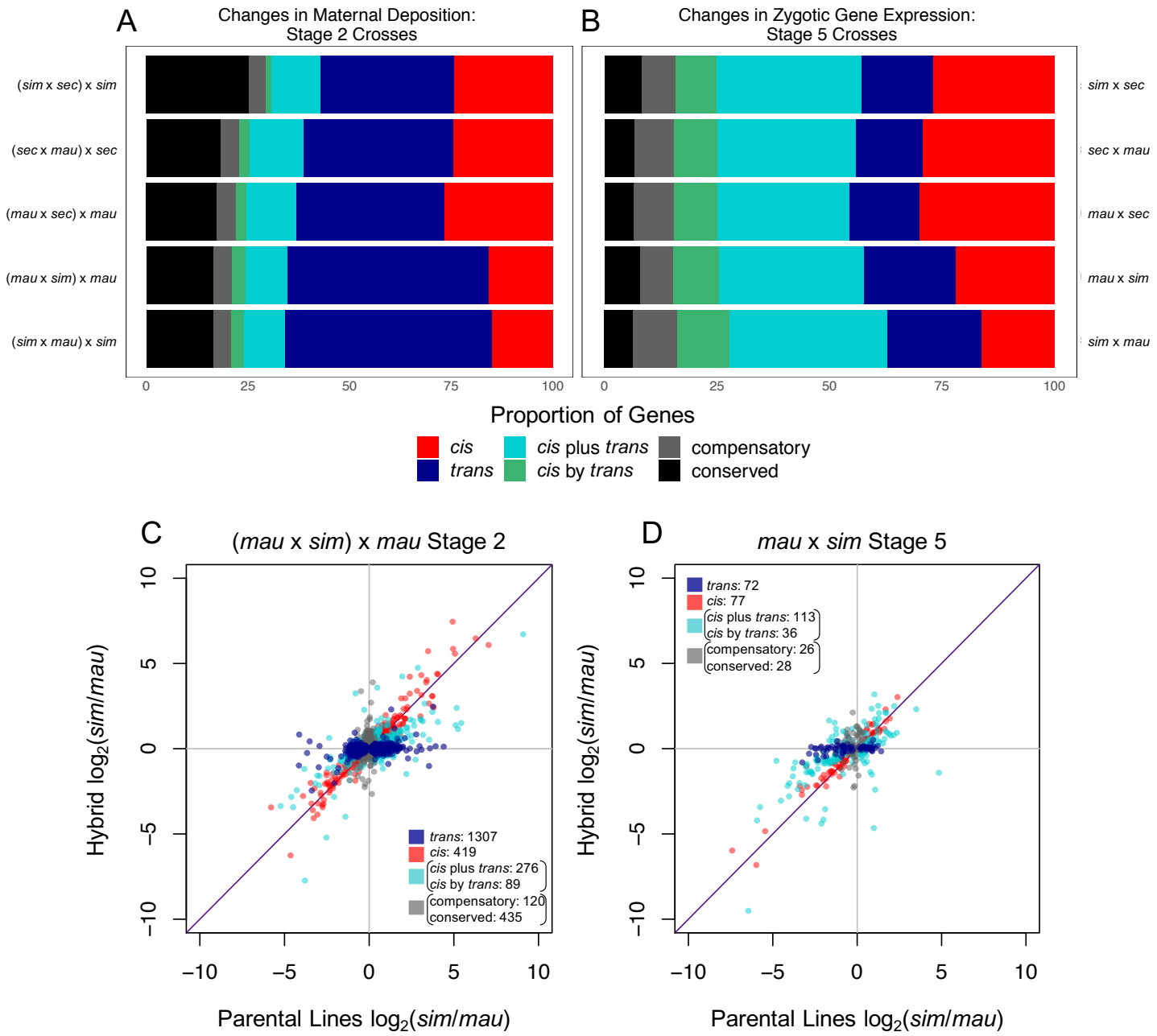


Figure 3: Different types of evolved regulatory changes dominate in maternal transcript deposition vs. zygotic transcription. Proportion of genes that fall into categories of regulatory change for each cross are shown for both the maternal transcript deposition (A) and zygotic gene transcription (B),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

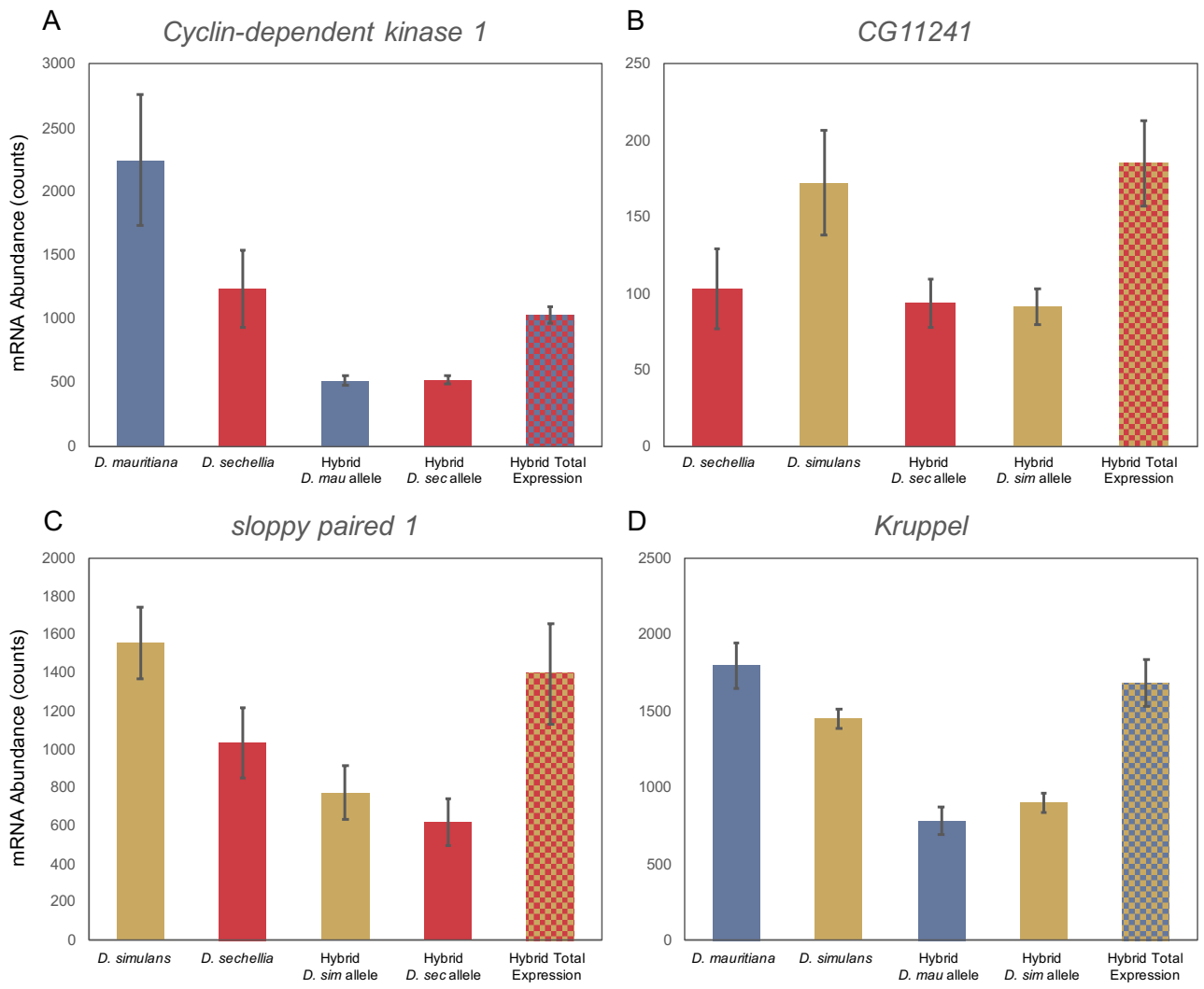


Figure 4: Examples of the type of regulatory changes observed, for individual genes. Transcript abundance, shown in counts, for each gene is plotted for the total mRNA abundance in both parental lines and for each parental allele within the hybrid; error bars shown represent the standard deviation. Total transcript abundance in the hybrid (the summation of levels from parental alleles in the hybrid) is shown as the last bar on the right in each graph. (A) Maternal transcript deposition of *Cyclin-dependent kinase 1 (Cdk1),* a critical cell cycle regulator in early development, changes in *trans* regulation between *D. mauritiana* and *D. sechellia*. Hybrid mRNA abundance is from the (*mau* x *sec*) x *mau* cross. (*Cdk1* also changes in *trans* regulation in the reciprocal cross comparison, (*sec* x *mau*) x *sec*.) (B) Maternal transcript deposition of *CG11241*, a gene of currently unknown function, changes in *trans* regulation between *D. sechellia* and *D. simulans*. Hybrid mRNA abundance is from the (*sim* x *sec*) x *sim* cross. (C) At stage 5 in development, *sloppy paired 1,* a critical pair-rule segmentation gene, changes in regulation through a combination of *cis* and *trans* regulatory changes (*cis* + trans) between *D. simulans* and *D. sechellia*. Hybrid expression is shown for the *sim* x *sec* cross. *Sloppy paired 2* also changes in *cis* regulation between these two species. (D) At stage 5, *Kruppel*, a gap gene crucial to segmentation changes in regulation through a combination of *cis* and *trans* regulatory elements (*cis* x *trans*) between *D. mauritiana* and *D. simulans*. Here, expression in the hybrid is from the *sim* x *mau* cross but in the reciprocal cross (*mau* x *sim*), *Kruppel* also changes in *cis* x *trans* regulation.

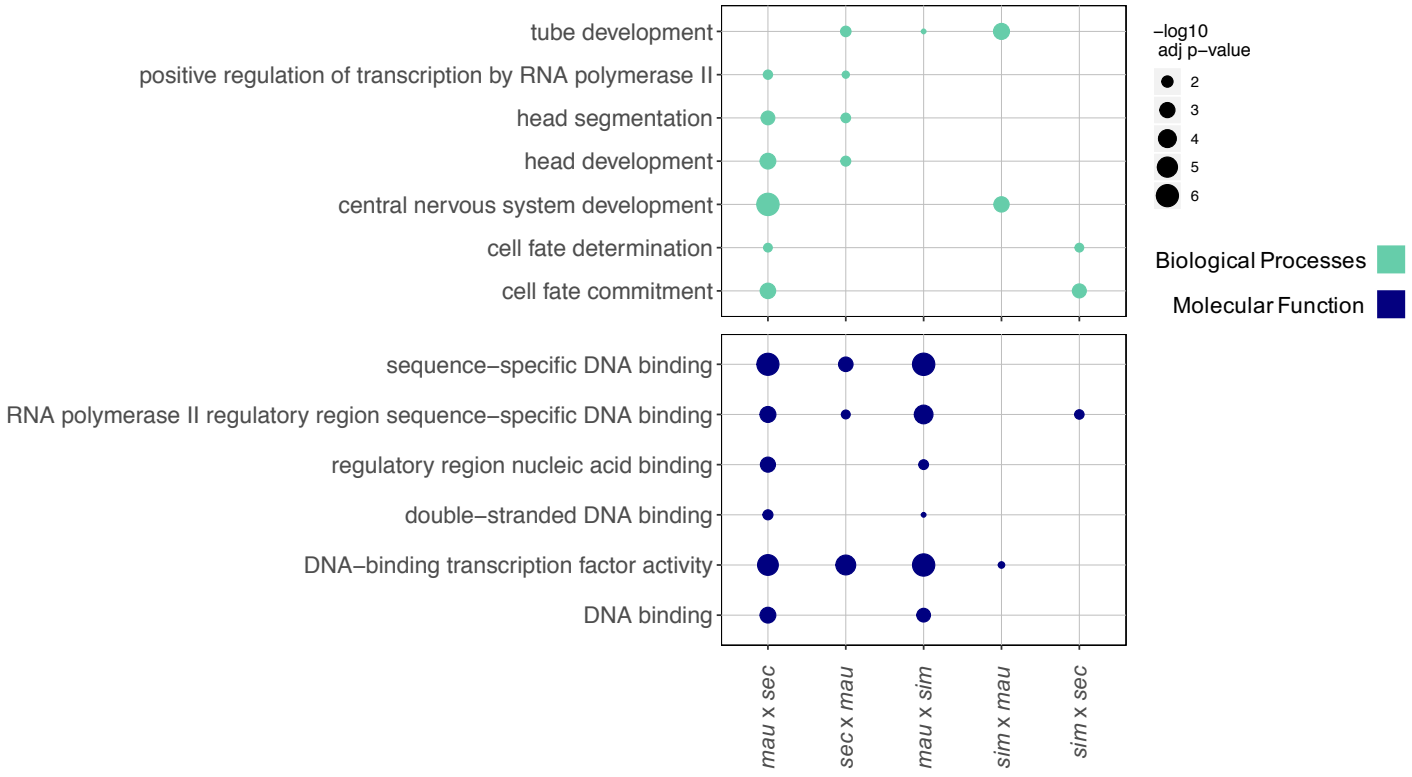


Figure 5: Gene ontology (GO) analysis identifies transcription factors that act in developmental processes as types of genes that change zygotically. Significantly enriched GO terms are listed for zygotically transcribed genes that change in *trans* regulation between each pair of species compared. Genes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

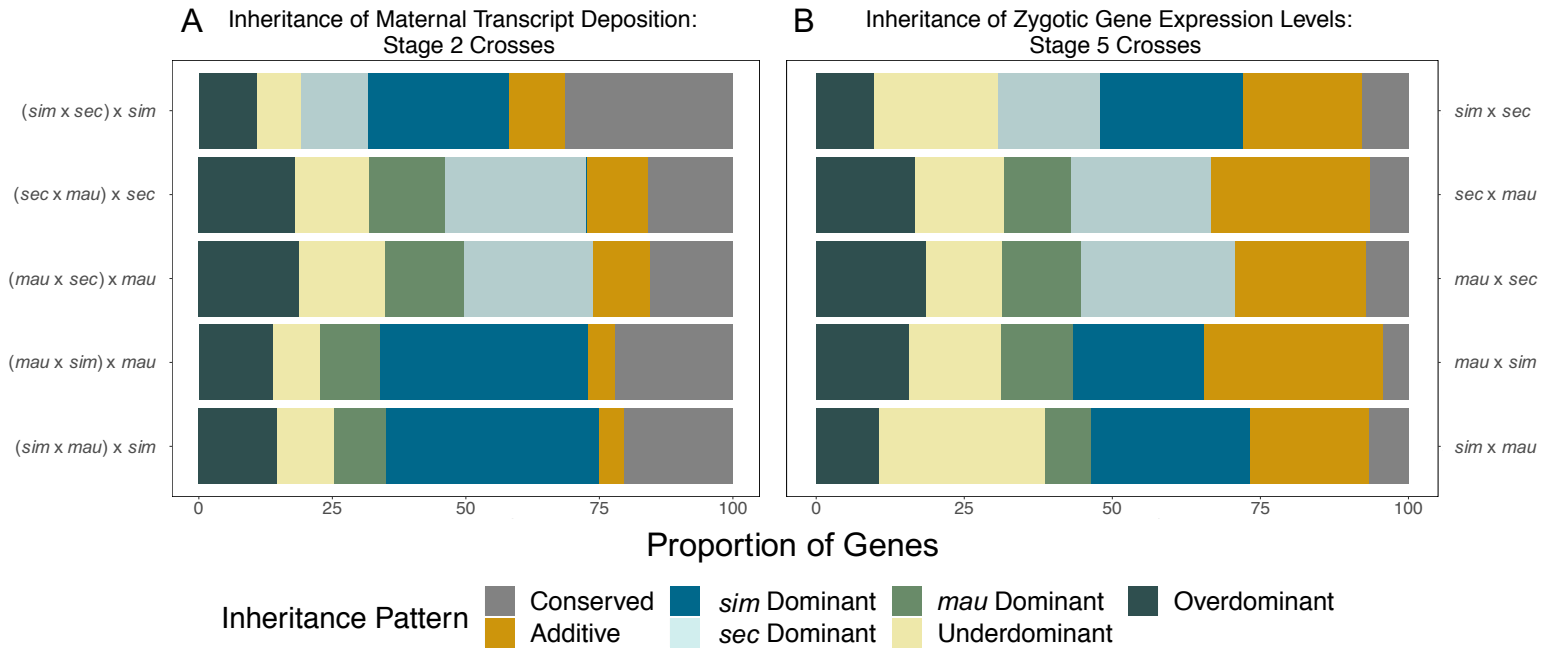


Figure 6: Patterns of inheritance show dominance of particular parental genomes at both stages. A) Shows patterns of inheritance for stage 2, over all genes and all crosses. B) Shows patterns of inheritance for stage 5, for mostly zygotic genes (see Methods) and all crosses. The maternal stage (A) shows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

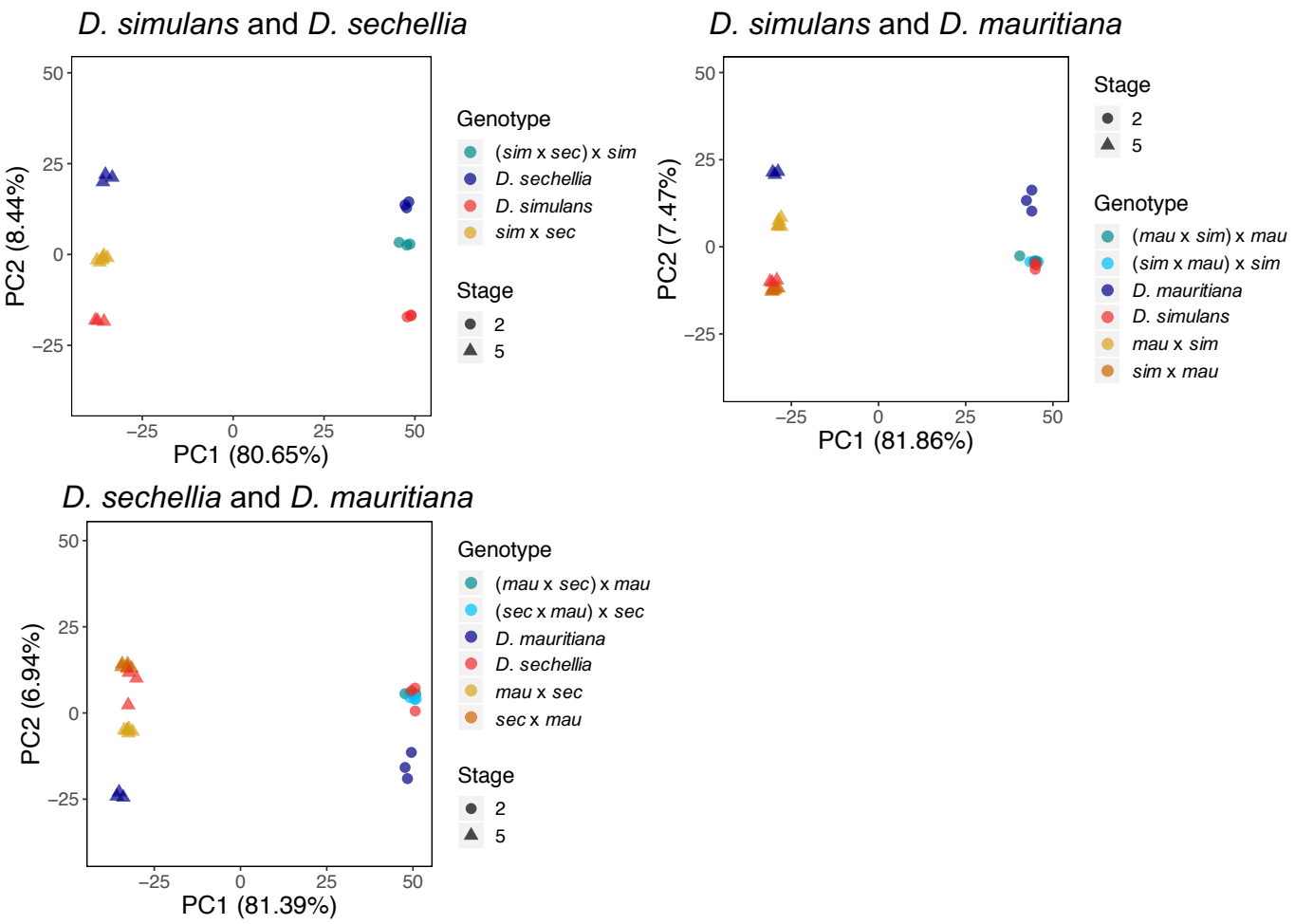


Figure S1: PCA plots for transcript abundance in all crosses. Samples of each stage, 2 or 5, cluster together. Samples of each genotype also cluster together, parental samples and hybrids. Proportion of variance explained by each principal component is listed on each axis.

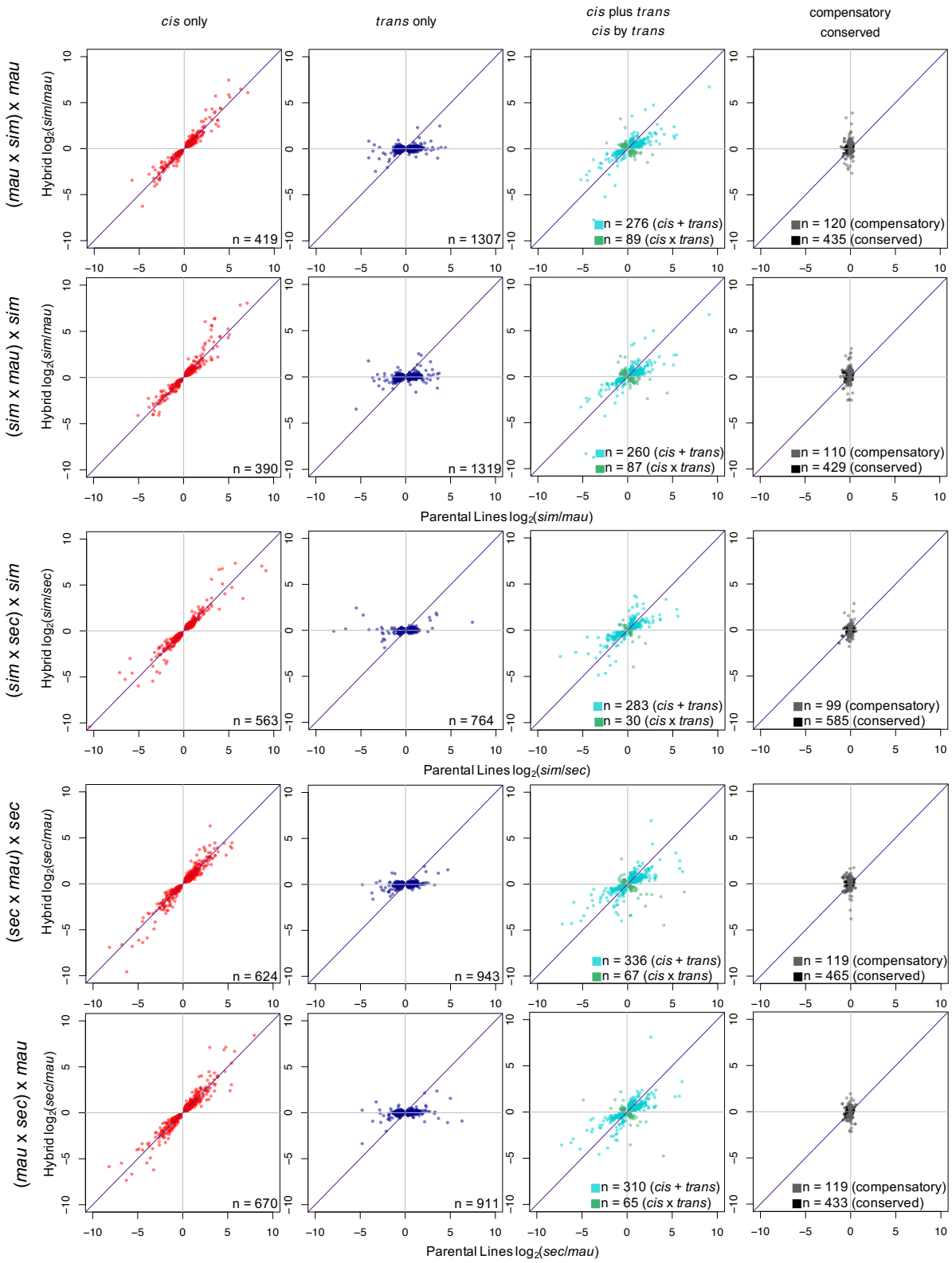


Figure S2: Regulatory changes in all pairwise comparisons for maternally deposited transcripts. Transcript level ratios between parental lines and within hybrids at stage 2 describe regulatory changes between species in each set of crosses. The number of genes in each category of regulatory change (n=)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

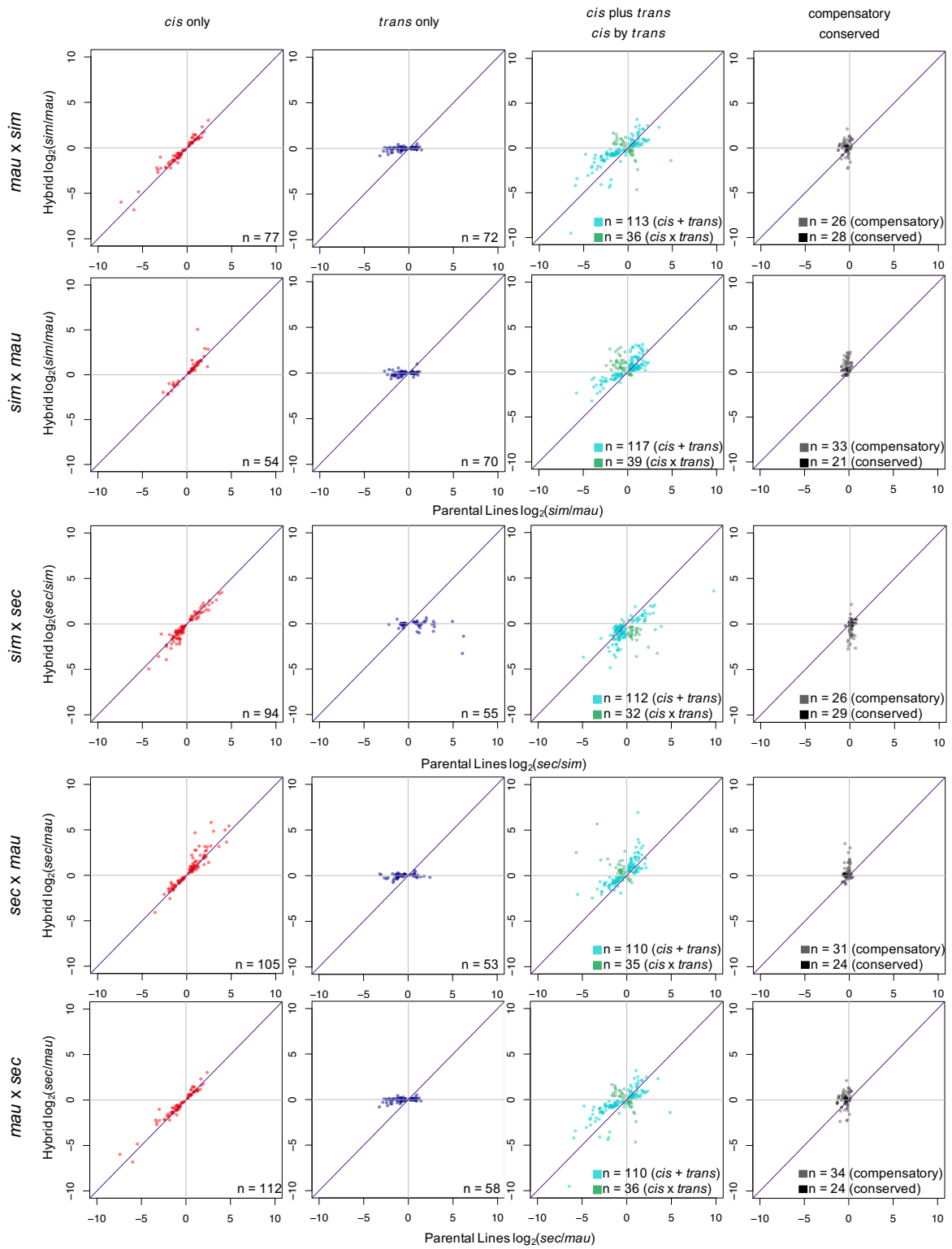


Figure S3: Regulatory changes in all pairwise comparisons for mostly zygotic genes. Transcript level ratios between parental lines and within hybrids at for mostly zygotic genes (see Methods) at stage 5 describe regulatory changes between species in each set of crosses. The number of genes in each category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

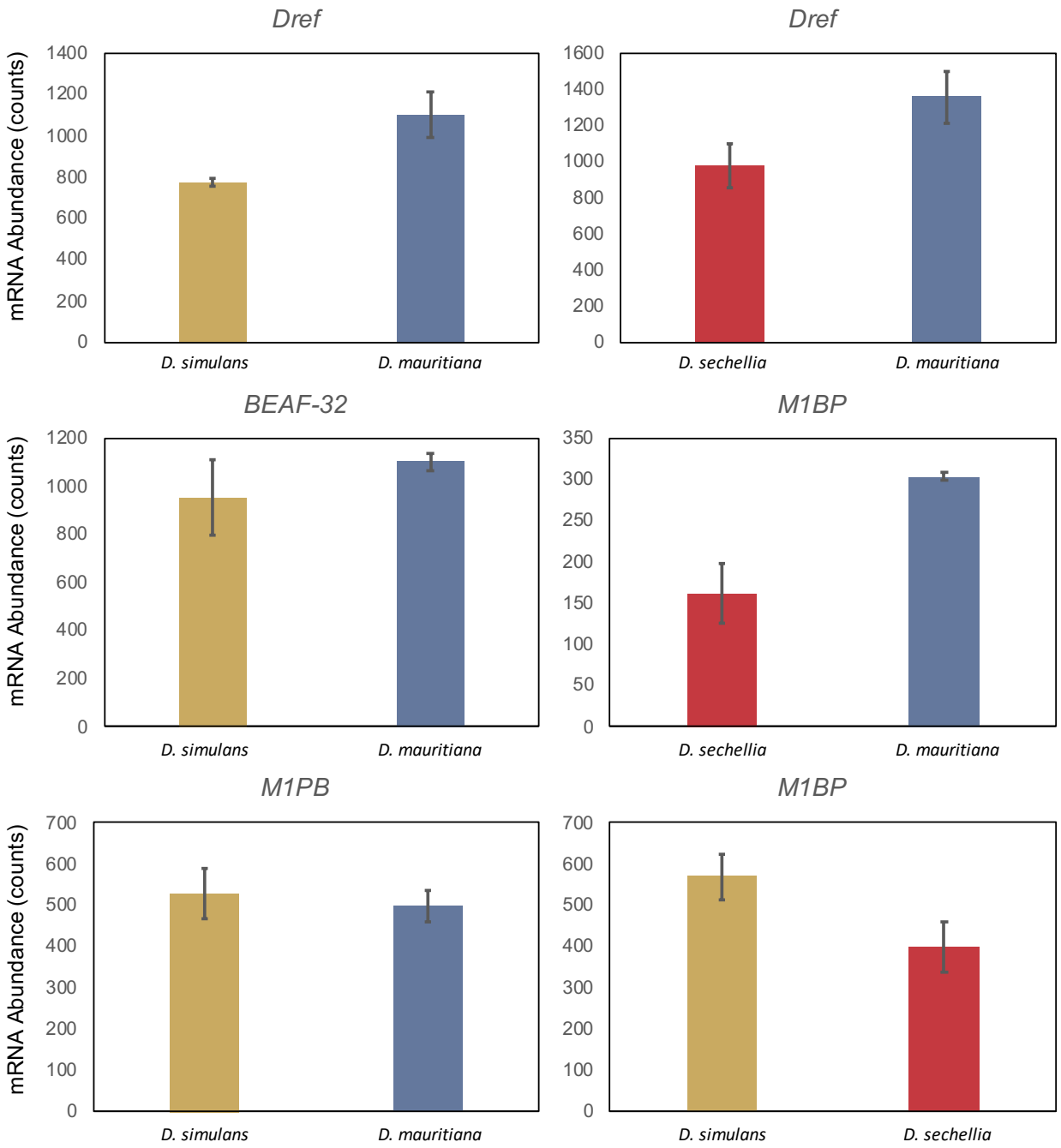


Figure S4: Transcript abundance from parental lines at stage 2 demonstrates differential maternal deposition of *M1BP*, *Dref* and *BEAF-32*. Counts for *D. simulans*/*D. mauritiana* and *D. simulans*/*D. sechellia* comparisons are averages across replicates from alignment to the *D. simulans genome*.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

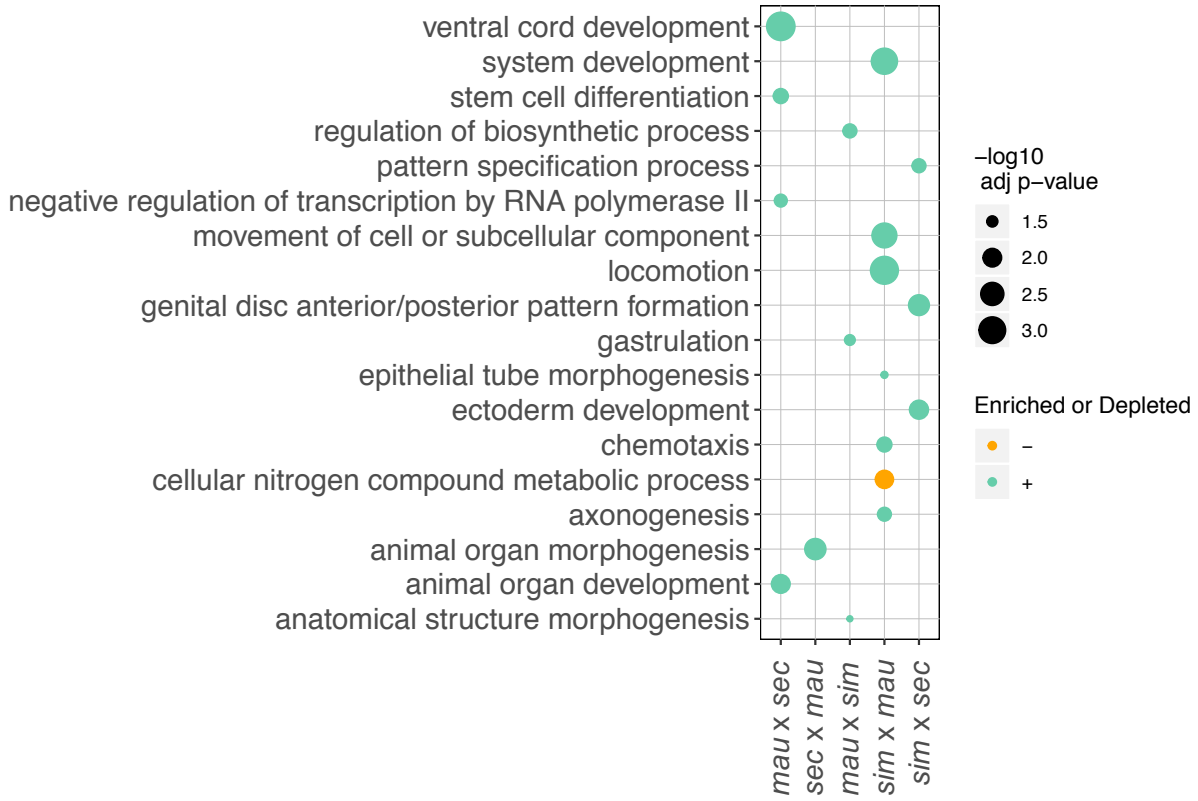


Figure S5: Gene ontology (GO) analysis for categories unique to a specific cross show enrichment for specific developmental processes. Significantly enriched GO terms are listed for zygotically transcribed genes that change in *trans* regulation between each pair of species compared. Again, zygotically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

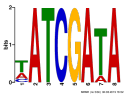
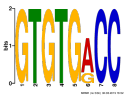
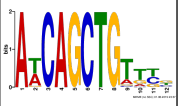
Motif	Predicted Binding Proteins	Enriched in upstream regions of *D. sechellia* genes that are conserved in expression (E-value)	Enriched in upstream regions of *D. sechellia* genes that are changing in expression (E-value)	Enriched in upstream regions of *D. simulans* genes that are conserved in expression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

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

Table S6: Number of genes within each category of inheritance pattern in the hybrids. Stage 2 hybrids are shown in the first five rows and stage 5 hybrids are shown in the last five rows of the table.