Polymorphism and Divergence of Novel Gene Expression Patterns in *Drosophila melanogaster*

Julie M. Cridland,[1] Alex C. Majane, Hayley K. Sheehy, and David J. Begun
Department of Evolution and Ecology, University of California, Davis, California 95616

ABSTRACT Transcriptomes may evolve by multiple mechanisms, including the evolution of novel genes, the evolution of transcript abundance, and the evolution of cell, tissue, or organ expression patterns. Here, we focus on the last of these mechanisms in an investigation of tissue and organ shifts in gene expression in *Drosophila melanogaste*r. In contrast to most investigations of expression evolution, we seek to provide a framework for understanding the mechanisms of novel expression patterns on a short population genetic timescale. To do so, we generated population samples of *D. melanogaster* transcriptomes from five tissues: accessory gland, testis, larval salivary gland, female head, and first-instar larva. We combined these data with comparable data from two outgroups to characterize gains and losses of expression, both polymorphic and fixed, in *D. melanogaster*. We observed a large number of gain- or loss-of-expression phenotypes, most of which were polymorphic within *D. melanogaster*. Several polymorphic, novel expression phenotypes were strongly influenced by segregating *cis*-acting variants. In support of previous literature on the evolution of novelties functioning in male reproduction, we observed many more novel expression phenotypes in the testis and accessory gland than in other tissues. Additionally, genes showing novel expression phenotypes tend to exhibit greater tissue-specific expression. Finally, in addition to qualitatively novel expression phenotypes, we identified genes exhibiting major quantitative expression divergence in the *D. melanogaster* lineage.

KEYWORDS *Drosophila*; evolution; RNA-seq; polymorphism; testis; accessory gland

While there is broad agreement that qualitatively novel phenotypes contribute to adaptation and the evolutionary success of certain lineages, the diversity of mechanisms underlying the origin and evolution of novelties remain a subject of intense investigation. One potential source of novel traits is novel genes. Such genes may originate through a diversity of mechanisms including gene duplication (Ohno 1970; Zhang 2003), gene fusion (Wang *et al.* 2004), and *de novo* gene evolution (Begun *et al.* 2006). A second, and potentially more common mechanism underlying novel phenotypes is the evolution of novel gene regulatory patterns or novel domains of gene expression of ancestral genes (*e.g.*, True and Carroll 2002). Such novel expression domains may result from *cis*-acting variation, such as the origin or modification of enhancers (Wray 2007; Koshikawa *et al.* 2015), or from *trans*-acting

changes, such as cooption of ancestral pathways into new expression domains via novel transcription factor expression (Doniger and Fay 2007; Rebeiz *et al.* 2011; Glassford *et al.* 2015; Long *et al.* 2016).

Several case studies deriving from the genetic analysis of novel phenotypes have provided evidence that such phenotypes may often result from the deployment of ancient genes or pathways in new developmental contexts. Some examples include butterfly eye spots (Brakefield *et al.* 1996), *Drosophila* wing spots (Gompel *et al.* 2005), *Drosophila* sex combs (Kopp 2011), *Drosophila* optic lobe expression (Rebeiz *et al.* 2011), and *Drosophila* and mammalian pigmentation (Wittkopp *et al.* 2003; Mallarino *et al.* 2016; Roeske *et al.* 2018).

An alternative line of investigation into the evolution of novelty focuses more strongly on novel gene expression phenotypes *per se*, many of which could be favored by selection pressures unrelated to visibly obvious phenotypes. One such form of novel expression is gene recruitment or cooption, whereby an ancient gene becomes expressed in an organ or tissue where it was not expressed ancestrally (True and Carroll 2002). A classic example of recruitment is the deployment of soluble enzymes into expression in the lens, leading

to crystallins (Wistow and Piatigorsky 1988; Wistow 1993), which has occurred multiple times in diverse animal clades (Tomarev and Piatigorsky 1996). More recently, the use of large-scale transcriptomic/genomic approaches has led to additional examples of recruitment, including the cooption of genes into venoms (Wong and Belov 2012; Martinson et al. 2017) and the recruitment of digestive enzymes to the female reproductive tract of a butterfly (Meslin et al. 2015). Nevertheless, general comparative and population genetic investigations of single-copy gene recruitment across tissues and organs remain rare.

Within Drosophila there are a few studies that address tissue-based patterns of gene expression. Dickinson (1980) showed that in a group of 27 Hawaiian Drosophila species for which quantitative estimates of protein abundance for five genes and five tissues were obtained, frequent gains and losses of tissue expression had occurred. He later demonstrated that the rate of evolution of tissue gains and losses of expression was lower in the virilis group than in the Hawaiian Drosophila (Thorpe et al. 1993), suggesting that expression evolution rate may vary among clades. In the Drosophila melanogaster subgroup, gains and losses of tissue expression have been observed for Gld (Ross et al. 1994) and Est-6 (Oakeshott et al. 1990). Begun and Lindfors (2005) used a semiquantitative approach to document a dramatically reduced expression of Acp24A4 in the accessory gland of D. melanogaster relative to D. simulans. Rebeiz et al. (2011) used in situ RNA hybridization for 20 candidate genes in larval imaginal discs in the D. melanogaster subgroup to describe the evolution of novel tissue-specific expression domains.

While there is a substantial amount of comparative transcriptomic work in Drosophila, it has tended to use either a pairwise (i.e., nonphylogenetic) approach (Brown et al. 2014) or use data collected from whole males, whole females, testes, or ovaries (Ma et al. 2018). These limitations compromise opportunities to investigate the evolution of gene cooption and qualitative changes in tissue patterns of expression, given that gene expression tends to vary dramatically over tissues (Chintapalli et al. 2007). Our goal here is to characterize qualitatively novel patterns of tissue expression in D. melanogaster, including the analysis of segregating variation influencing these novel patterns.

We focus on six D. melanogaster genotypes that are part of the Drosophila Genetic Reference Panel (DGRP) lines from Raleigh, NC (Mackay et al. 2012). Each of these genotypes was sampled for five tissues: the testis, the accessory gland + anterior ejaculatory duct (hereafter AG), the female head, the third-instar salivary gland (hereafter SG), and first-instar larvae. Our choice of tissues provides several functional comparisons. For example, inclusion of testis and the AG enables a comparison of two tissues that function specifically in male reproduction, but one tissue is primarily germline while the other is somatic. While the rapid evolution of AG-expressed genes has often been attributed to directional selection relating to male reproduction (Begun et al. 2000; Mueller et al. 2005), it is worth noting that any unusual evolutionary properties of such genes could also be attributable to glandular function. Thus, the inclusion of SG enables a comparison of two glands, one of which functions in reproduction and one of which does not. The head and first-instar larvae are neither primarily reproductive nor glandular, and thus serve as potentially useful comparisons for the other sampled tissues, although we acknowledge that these comparisons are weakened by the fact that the head and first-instar larvae represent complex mixtures of several tissue types.

Using expression data from these D. melanogaster genotypes and their F1s, along with comparable data from outgroup species D. simulans and D. yakuba, we identify both polymorphic and fixed gains and losses of tissue expression that have evolved in D. melanogaster since the D. melanogaster/ D. simulans common ancestor. Our focal species represent very recent evolutionary history—D. melanogaster diverged from D. simulans and D. yakuba only ∼1.4–3.4 MYA (Obbard et al. 2012)—thus, we address the evolution of tissue expression on short timescales, which has received little attention. In addition to identifying recent, qualitative evolution of expression across diverse D. melanogaster tissues, we identify the evolution of major quantitative changes within tissues by identifying genes that show dramatic recent evolution of transcript abundance in D. melanogaster.

Materials and Methods

Flies, tissues, RNA isolation, library construction, and sequencing

We sequenced the transcriptomes of six D. melanogaster strains from the DGRP population resource (Mackay et al. 2012), hereafter "RAL" lines (Table 1). In addition, we sequenced the transcriptomes of each of three unique F1 crosses derived from the six RAL strains with each RAL strain contributing to one F1 cross (Supplemental Material, Table S1). Finally, we also sequenced between one and three strains each of D. simulans and D. yakuba per tissue, which were used as outgroups. For each D. melanogaster genotype we prepared libraries from five tissues, AG, female head, first-instar larvae, wandering third-instar larva SG, and testis. For each genotype and tissue we prepared and sequenced one library.

Tissues from 30 flies per strain were used for the head, first-instar larvae, and SG. Two-day-old virgin females were collected in 1.5-ml centrifuge tubes and flash-frozen in liquid nitrogen. Tubes were scraped across a tube rack multiple times and contents were collected on a clean petri dish placed on top of a piece of dry ice. Heads were collected using a clean paint brush and placed in Trizol. SGs from unsexed wandering third-instar larvae, identified as having branched anterior spiracles and dark orange rings at the tip of the posterior spiracles, were dissected in cold 1× PBS and then placed directly into Trizol. Unsexed first-instar larvae were collected by allowing females to oviposit on food for 1 hr and then collecting larvae 24 hr later using a dissecting needle. Testis

samples were dissected from 24-hr-old virgin males and immediately placed in Trizol on ice. Between 100 and 200 testes were collected per strain. AG samples were dissected from 48-hr-old virgin males (~50 per strain) and immediately placed in Trizol on ice. All tissue samples were stored at $-80°$ until isolation of RNA.

Tissues were homogenized in 200 μl Trizol using a sterile pestle and the volume then adjusted to 1 ml. Next, 200 μl chloroform was added and shaken for 20 sec, followed by room temperature incubation for 5 min. Samples were then centrifuged at $4°$ and 13,000 rpm for 15 min and the upper phase collected. After addition of 1 μl glycogen, 500 μl isopropanol was added and mixed by gentle inversion. Samples were left at $-20°$ for 1 hr, after which nucleic acids were pelleted and then washed with 70% ethanol, followed by drying and resuspension in nuclease-free water. All samples were subjected to DNase digestion using the TURBO DNA-free kit (Ambion) following the manufacturer's protocol and the qualities of the resulting RNAs were checked using RNA Nano chip on a Bioanalyzer (Agilent).

Libraries were prepared using the NEBNext Ultra RNA Library Prep Kit for Illumina (New England Biolabs, Beverly, MA) with 1 μg total RNA as input. The manufacturer's protocol was used with minor modifications. All AMPure bead elution steps were performed with PCRClean DX beads (Aline Biosciences). Qualities of libraries were estimated using the High Sensitivity DNA chip on a Bioanalyzer (Agilent). Libraries were sequenced on an Illumina HiSeq2000 to generate 100-bp paired-end reads.

The focal set of genes used in all expression analyses is composed of 7356 1:1:1 orthologs (Table S2) identified in *D. melanogaster*, *D. simulans*, and *D. yakuba* using MCL (van Dongen and Abreu-Goodger 2012) which uses a Markov cluster algorithm for assigning genes into families (M. Hahn and G. Thomas, personal communication). Di- and polycistronic genes were omitted from the analysis, as were genes with one or more exons that were shared perfectly with another gene.

Expression measures

Reads were initially aligned to the appropriate reference genome; *D. melanogaster* version 6.19, *D. simulans* version 2.02, and *D. yakuba* version 1.05 using Hisat2 (Kim *et al.* 2015) with default parameters. The *D. melanogaster* F1 transcriptomes were additionally aligned to parent-specific reference genomes. We downloaded the RAL line-specific reference sequences available on the *Drosophila* NEXUS platform (Lack *et al.* 2015, https://www.johnpool.net/genomes.html) and aligned each F1 transcriptome simultaneously to both parents of the cross, as above. We then kept reads that reported alignment exclusively to one parental genome, identified by having a quality score of 60. These reads were then used to generate new fastq files, which were then realigned to the NEXUS parent-specific reference as before.

We used Stringtie (Pertea *et al.* 2016) to calculate transcripts per million (TPM) for each gene in each transcriptome, including in the parent-specific transcriptomes generated by

Table 1 Number of genotypes sequenced per tissue

Species	AG	Head	1st instar	SG	Testis
D. melanogaster	6	6	6	6	6
D. melanogaster F1s	3	3	3	3	3
D. simulans	1	2	2	2	3
D. yakuba	1	2	2	2	2

AG, accessory gland plus ejaculatory duct; SG, salivary gland.

separating identifiable reads in a cross. We used the -e option to restrict the analysis to previously identified and annotated genes identified in GTF files obtained from FlyBase (www.flybase.org) on December 1, 2018, and that were also in our list of 1:1:1 orthologs between the three species. Dicistronic and polycistronic genes were removed from the analysis, as were genes that shared one or more exons perfectly with another gene.

We used two approaches to investigate the possible effect of coverage variation across libraries on gene expression calls. First, we investigated whether the number of reads was correlated with the number of genes called as expressed at cutoffs TPM $\geq$ 1 and TPM $\geq$ 2. We found no relationship (linear regression, $P > 0.05$), either overall or in analyses of each tissue. Second, because the range of reads per library varied from 12 to 63 million, with most libraries having at least 20 million reads, we downsampled higher-coverage libraries to both 12 million and 20 million read pairs and reestimated TPMs. This downsampling had a negligible effect (~3 genes per library at 20 million reads and ~17 genes per library at 12 million reads) on the total number of genes identified as expressed at a TPM $\geq$ 2 per library. The downsampled data exhibited less than one fewer neomorph per library compared to the actual data. We conclude from these results that coverage variation has a negligible effect on our inferences.

Gain- and loss-of-expression phenotypes

Genes that had TPM $<$ 0.1 in all strains for both outgroup species for a given tissue were inferred to exhibit no expression in that tissue as the *D. melanogaster* ancestral state. We further used an expression threshold of TPM $\geq$ 2 to identify a gene as expressed. This expression cutoff is appropriate based on previous studies (Wagner *et al.* 2013; Thompson *et al.* 2019). We defined candidate gains of tissue expression in *D. melanogaster* as genes for which at least one RAL genotype expressed the gene above a threshold TPM of $\geq$2, and for which there was no evidence of expression in the outgroup species. We refer to these gain-of-expression phenotypes as neomorphs for convenience, in spite of the fact that we have no evidence that they behave as gain-of-function variants/phenotypes (Figure 1A). We categorized genes exhibiting fixed expression gains as those for which all RAL genotypes had TPM of $\geq$2; all other neomorphs were considered polymorphic. A substantial number of such genes exhibited at least one RAL line with TPM = 0. Using this logic, the genes for which some RAL strains have TPM = 0 and
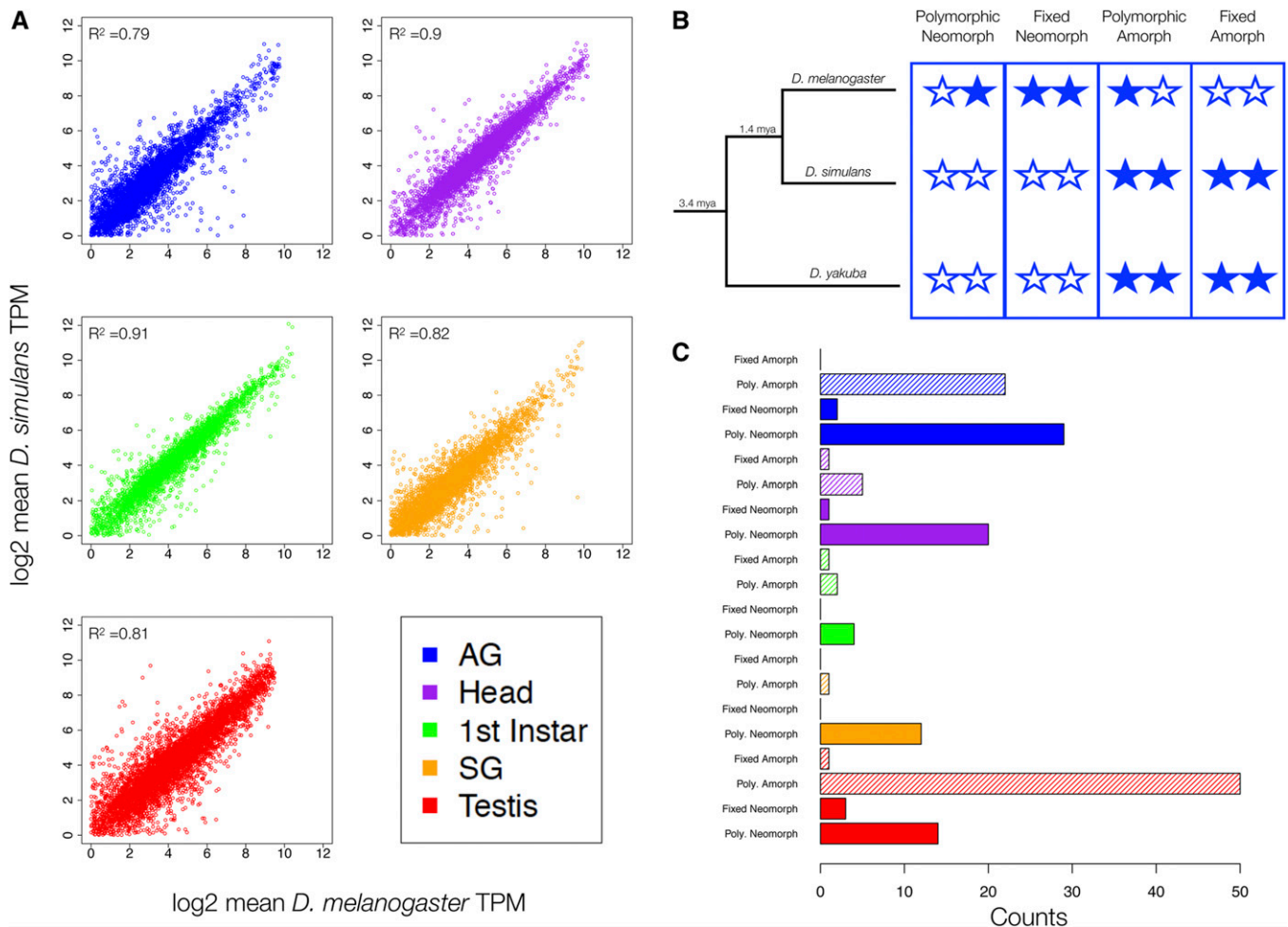
Figure 1 Gene expression polymorphism and divergence in *D. melanogaster*. (A) Correlations of log2 TPM (transcripts per million) estimates between *D. simulans* and *D. melanogaster* for several tissues for 1:1:1 orthologs in *D. melanogaster*, *D. simulans*, and *D. yakuba*. Only genes in the 99th percentile for *D. melanogaster* were used. AG, accessory gland plus ejaculatory duct (blue); head, female head (purple); first instar, first-instar larvae (green); SG, third-instar larval salivary gland (orange; and testis (blue). (B) Parsimony was used to define polymorphic (Polym.) and fixed *D. melanogaster* neomorphs and amorphs. Representation of gene expression patterns for a gene in two strains from each of three species. Filled star, expressed gene (TPM ≥ 2); empty star, nonexpressed gene (TPM ≤ 0.1). Ages of hypothetical common ancestors are from Obbard *et al.* (2012). (C) Numbers of observed neomorphs and amorphs per tissue. Tissues denoted by colors used in (A).

others have TPM ≥ 2 are the most conservative set of polymorphic neomorphs. Using the same basic approach, we identified losses of tissue expression in *D. melanogaster* (hereafter, amorphs) by identifying genes for which one or more RAL lines had TPM < 0.1 for a gene in a given tissue, and all outgroup strain TPM estimates exceeded a threshold value of TPM ≥ 2.

Tissue specificity of candidate genes

We used the FlyAtlas2 *D. melanogaster* gene expression data set (Leader *et al.* 2018) to estimate the tissue specificity of our candidate genes. We downloaded the data from the source in SQL database format [motif.gla.ac.uk/downloads/FlyAtlas2_18.05.25.sql (accessed July 2, 2018)]. We then used an R script to estimate the index of adult tissue specificity, τ (Yanai *et al.* 2005), separately for male and female tissues for every gene/transcript. Expression enrichment in each tissue for each gene/transcript was estimated by dividing the

fragments per kilobase of transcript per million mapped reads (FPKM) for that tissue by the FPKM for the whole-body samples of adult males or females. We used a threshold of FPKM ≥ 2 to call a gene expressed in a particular tissue; in the event of a whole-body FPKM < 2, we set it equal to 2 for enrichment estimates to avoid large values of τ in very lowly expressed genes (David Leader, personal communication).

To investigate whether ancestral tissue patterns of expression might influence the probability of gain-of-expression in a new tissue, we identified for testis and AG neomorphs the expression patterns in nonfocal tissues (*i.e.*, "not-testis" and "not-AG"), and then compared neomorph expression properties to those of other (*i.e.*, ancestrally) AG- or testis-expressed genes. To do so, we recorded for testis or AG neomorphs the nonfocal tissue that showed the highest level of expression. To determine the comparable pattern expected for ancestrally AG- or testis-expressed genes we used a resampling approach. We selected at random among the genes expressed

at FPKM $\geq$ 2 in the focal tissue in FlyAtlas2, the number of genes equal to the number identified as neomorphs. For each sampled gene we recorded the tissue with the greatest expression other than the focal tissue. We repeated this 1000 times to generate an empirical distribution of the non-focal tissues that tend to show the greatest expression.

F1 validation of parental expression in D. melanogaster

The identification of polymorphic neomorphs in *D. melanogaster* depends fundamentally on the premise that the observations of extremely low expression values (0 < TPM < 0.1) in RAL genotypes are real rather than technical artifacts. To address this issue, we used gene expression estimates from RAL F1 genotypes. For each F1 we compared the TPM estimates for the parental strains to the TPM estimates of the F1 and to the number of parent-separated reads that mapped to each parental chromosome (as described below). We examined crosses that exhibited zero reads in one of the two RAL parents used to create an F1, and which in the F1 met our minimum number of 10 mappable parent-separated reads. Under the assumption that a low-expression phenotype is strictly *cis*-mediated and real rather than artifactual, RAL chromosomes that exhibited zero reads from a RAL parent should also exhibit zero reads from the same chromosome carried in an F1 (Zhao *et al.* 2015). We identified the number of such cases across all tissues and then compared this number to the expected number of cases under the null hypothesis that TPM = 0 observations from RAL parents are technical artifacts. We generated this expectation by randomly assigning TPM estimates across RAL genotypes for each tissue, and then determining how many RAL TPM = 0 cases were recapitulated in the F1 as zero reads deriving from the corresponding parental chromosome. This was done 1000 times to generate the expected value under the null hypothesis that TPM = 0 estimates in RAL parents are false negatives.

Comparisons with external data

Because our transcriptome data included only one replicate of each sample, we sought support for our conclusions from other *D. melanogaster* data sets. First, we compared the results of our testis neomorph and amorph analysis with an analysis of a replicated testis transcriptome data set (four replicates per genotype from each of two *D. melanogaster* genotypes) published by Yang *et al.* (2018). For each of these replicated testis transcriptomes, reads were aligned to the *D. melanogaster* version 6 reference genome using HISAT2 (Kim *et al.* 2015) as above, and TPM was estimated using Stringtie (Pertea *et al.* 2016) as above. TPM was estimated for each gene in each of the four replicates for each genotype. We considered a gene expressed in a genotype if the gene had a mean TPM $\geq$ 2 (Table S3). Second, we compared our neomorph and amorph expression inferences to expression estimates from multiple tissues found in FlyAtlas2 (Leader *et al.* 2018, http://flyatlas.gla.ac.uk/FlyAtlas2/index.html, last accessed March 2020) (Table S4).

General expression changes in the D. melanogaster lineage

While our main goal was to identify derived expression changes that one can reasonably classify as qualitatively discrete, we also sought to identify the genes that showed the greatest quantitative expression change in the *D. melanogaster* lineage without conditioning on extremely low (TPM < 0.1) expression estimates from *D. simulans* and *D. yakuba*. We identified for each tissue the genes that exhibited a $\geq$1.25-fold difference between *D. melanogaster* and *D. simulans* and a fold difference of <1.25 between *D. simulans* and *D. yakuba* (in all cases using the mean TPM estimate from each species), which should enrich for genes showing substantial changes of expression in *D. melanogaster*.

Genetics of gene expression variation

We used parental RAL TPM estimates and corresponding estimates from their F1s to partition AG and testis expression variation into *cis* and *trans* effects, generally following McManus *et al.* (2010). We used a fold-change cutoff of 1.25 to call differences in (1) expression between RAL parents, (2) between parent-specific estimates in hybrids, and (3) between the observed overall F1 expression and the expected F1 expression assuming additivity. We also required that at least one parent or F1 exhibit TPM $\geq$ 2 to be included in the analysis. To identify differences between read counts derived from each parent within an F1 we performed a binomial test followed by Bonferroni correction. Reads from an F1 were identified as deriving from one parent by aligning the entire transcriptome to a combined parent-specific reference genome composed of the two NEXUS (Lack *et al.* 2015) genotype-specific references, with additional masking so that any region that was masked in one parent was also masked in the other parent. We compared the ratios of reads aligned uniquely to each parent to identify any bias before using the parent-separated reads in downstream analyses (Table S5). The RAL307 $\times$ RAL304 AG library did not have roughly equal reads derived from each parent and was dropped from the analysis. Once parent-specific reads were identified, new fastq files were generated and the parent-specific reads were realigned, as described previously, to the *D. melanogaster* reference. We categorized genes into groups using the following criteria.

Cis: fold change difference between parental TPMs, fold change difference in parent-specific reads from the F1 in the same direction as the parents, and no fold change difference between F1 TPM and expected F1 TPM.

Trans: fold change difference between parental TPMs, no difference in F1, and observed F1 TPM is different from expected F1 TPM (either lower or higher).

Cis + trans: fold change difference between parental TPMs, fold change in the F1 in the same direction as the parents, and observed F1 TPM is higher than expected F1 TPM.

Cis − Trans: fold change difference between parental TPMs, fold change in the F1 in the same direction as the parents, and observed F1 TPM is lower than expected F1 TPM.

Cis * Trans: fold change difference between parental TPMs, fold change in the F1 in the opposite direction as the parents, and observed F1 TPM is different from expected F1 TPM.

Compensatory: no fold change difference between parental TPMs, fold change in the F1, and observed F1 TPM does not differ from expected F1 TPM.

Conserved: no fold change difference between parental TPMs, no fold change in the F1, and observed F1 TPM does not differ from expected F1 TPM.

Ambiguous: all other patterns of expression between parental TPMs and F1s.

Data availability

The supplemental material are as follows: Table S1 (sequenced tissues) contains the number of reads sequenced for each RNA-sequencing (RNA-seq) experiment; Table S2, FBgn list of all orthologs; Table S3, TPMs of neomorphs and amorphs in Yang *et al.* (2018) testis data; Table S4, FPKM in neomorphs measured in FlyAtalas2; Table S5, parent-specific read-pair counts for allele specific expression (ASE) analysis; Table S6, TPMs in all 1:1:1 orthologs (this table contains the expression means and medians for all 7356 genes; Table S7, number of genes expressed in each tissue (this table includes the number of genes expressed in each tissue or group of tissues; Table S8, R^2 comparing mean TPMs in *D . simulans vs. D. melanogaster*; Table S9, regulatory mechanisms for neomorphs and amorphs (this table contains expression levels for parental lines and crosses as well as the predicted regulatory mechanisms); Table S10, neomorphs (this table contains expression measurements for all candidate neomorphs); Table S11, amorphs (this table contains expression measurements for all candidate amorphs); Table S12, AG/testis top-expressing gene list [gene ontology (GO) analysis for genes expressed most highly in the AG and testis based on FlyAtlas2 data; and Table S13, genes showing quantitative expression differences in *D. melanogaster* (this table contains the set of genes with increases or decreases in *D. melanogaster* expression relative to *D. simulans* and *D. yakuba*). Raw sequence data for all experiments is available from the Sequence Read Archive, PRJNA575046 and PRJNA210329. Supplemental material available at figshare: https://figshare.com/s/1bf8cf2433db8dacfe0c.

Results

General patterns of gene expression

Our analysis included 7356 single-copy ortholog sets in the three species. We estimated expression of these genes in *D. melanogaster, D. simulans*, and *D. yakuba* (Table S6) for five tissues: AG, female head, first-instar larvae, SG, and testis. Expression patterns within tissues were generally similar between species (Table 2 and Table S7). For example, when comparing orthologs with a mean TPM $\geq$ 2 in a given tissue across genotypes for a species, a large proportion of 1:1:1

orthologs were expressed in all five tissues studied: *D. melanogaster* 48%, *D. simulans* 49%, and *D. yakuba* 39%. While fewer genes were detected in all five tissues in *D. yakuba* at a TPM $\geq$ 2, we observe greater similarity in the proportions across species using either genes expressed in four or five tissues, or if we use a cutoff of TPM $\geq$ 1. Additionally, most genes (59% at TPM $\geq$ 2 and 64% at TPM $\geq$ 1) were expressed in the same set of tissues across the three species. A minority of genes (20% in *D. melanogaster*, 18% in *D. simulans*, and 18% in *D. yakuba*) were expressed in only one tissue. This pattern is consistent with those from FlyAtlas2 data (Leader *et al.* 2018), in which 10.5% of *D. melanogaster* genes exhibited FPKM $\geq$ 2 in only one tissue and 41% of genes were expressed in all 30 tissues measured, including adult and larval tissues. As observed in previous studies (Graveley *et al.* 2011), the testis is highly unusual in that the vast majority of genes expressed in only one tissue are expressed in the testis for all three species ($\sim$13%). Both glandular tissues exhibit very few tissue-specific genes for all three species.

To characterize general patterns of expression divergence across tissues we estimated the *D. melanogaster vs. D. simulans* log2 mean TPM correlation. We initially examined all genes in the bottom 99th percentile of *D. melanogaster* expression, with expression $\geq$1 (Figure 1A). The top 1% of *D. melanogaster* genes were excluded to avoid inflating R^2 values due to extreme expression values. The variation in TPMs at the tail of the distribution prompted us to then further separate our genes into the bottom 95% and top 95–99% percentiles of expressed genes in *D. melanogaster* (Table S8). We found that similar to the <99% set of genes, for the bottom 95% of genes, R^2 values were lower for the testis, AG, and SG (0.76, 0.7, and 0.74, respectively) than for the other tissues (0.86 for head and 0.88 for first-instar larvae). Thus, while previous studies have provided evidence of unusually rapid expression divergence for male-biased genes, which tend to be testis-biased (Meiklejohn *et al.* 2003; Parisi *et al.* 2004; Graveley *et al.* 2011), the fact that we observe similar interspecific expression correlations for male reproductive tissue and the SG suggest the possibility that several tissues may evolve quickly in *Drosophila*. The 95–99% percentile *D. melanogaster* genes generally exhibited weaker correlations of between-species TPM relative to the bottom 95% of genes, with the testis exhibiting the biggest difference ($R^2 = 0.76$ *vs.* $R^2 = 0.26$).

F1 validation of absence of D. melanogaster expression

Inferences of polymorphic gains of tissue expression may be subject to error because this categorization requires reliable observations of very low expression, TPM < 0.1, in some *D. melanogaster* strains and higher levels, TPM $\geq$ 2, in other strains. To investigate whether observed TPMs $\leq$ 0.1 in RAL strains may be artifactual, we compared the expression measures for RAL strains to the expression data from the F1 animals. We find that expression in the F1s corresponds well to the expression in the parental strains, even under the simplifying assumption of additivity, deviations from which will weaken the correlations between parental and F1 expression phenotypes.

In the set of 10,730 observations (pooled across genes, F1s, and tissues) for which both parents in a cross exhibit TPM = 0, the vast majority (10,130 or 94.4%) of F1 observations also exhibit TPM < 0.1, supporting the conclusion that lack of expression in a tissue reflects real variation. The small proportion of cases that deviate from this pattern may be false negatives in the RAL parents or may be a consequence of *cis* × *trans* interaction effects in the F1, suggesting that our estimate provides an upper bound on the false-negative rate in the RAL parents. Moreover, focusing specifically on the subset of polymorphic neomorphs in which some RAL strains exhibit TPM = 0, our permutation test (*Materials and Methods*) indicates that these cases of zero expression in one parent are not artifacts but instead, as a group, reflect real extreme expression difference between parental chromosomes (binomial test, $P = 4.31\mathrm{e}^{-8}$). Because RAL chromosomes may exhibit true TPM = 0 in the parental strains and true nonzero TPM in the F1 due to *trans* (or *cis* × *trans* interaction) effects in the F1, our binomial tests are conservative with respect to the biological reality of TPM < 0.1 observations in the RAL parents. Additional results, described below, support the idea that many polymorphic neomorphs are explained by major *cis* effects, providing further support for the reality of the parental RAL lack-of-expression calls. Table S9 shows 39/132 instances of F1 crosses that support *cis* effects that contribute to a novel expression phenotype. This accounts for 20 neomorphs and 9 amorphs, with one or more crosses having evidence of *cis* effects, including 10 of 31 AG neomorphs and 10 of 17 testis neomorphs. Pure *cis* effects are found in nine neomorphs, including three where one parent has a TPM = 0 (Table S9).

Identification and expression of neomorphs and amorphs

We identified 85 neomorphs in *D. melanogaster*, six (7%) of which were classified as fixed (expressed at TPM ≥ 2 in all RAL lines) yet showed no expression in outgroup strains (Figure 1B and Table S10). The majority of these fixed neomorphs were expressed either in the AG (two genes) or testis (three genes), and had expression means well above our cutoffs (Table 3). The remaining neomorphs (*n* = 79) were classified as polymorphic (Figure 1C). Fifty of these 79 genes (63%) exhibited a novel tissue expression domain in a single RAL genotype at TPM ≥ 2. We compared expression of neomorphs in the RAL lines to expression in the F1 crosses. For 118 of 136 observations for which either one or both RAL parents expressed at TPM ≥ 2, we observed neomorph expression at TPM ≥ 2 in the F1.

Overall, of the five tissues examined, the AG exhibited the greatest number of neomorphs (31), while the first-instar larva exhibited the fewest (4) (Figure 1C). Testis neomorphs were expressed at TPM ≥ 2 in an average of 3.3 RAL genotypes, while other tissue neomorphs were expressed in an average of 1.2–1.9 RAL genotypes. In general then, male reproductive tissues exhibited more novel expression patterns (AG) or expressed such novel phenotypes more consistently (testis) across *D. melanogaster* genotypes. For all tissues,

Table 2 Mean number of genes expressed by tissue

	D. melanogaster	*D. simulans*	*D. yakuba*
Genes expressed only in tissue			
AG	31	16	11
Head	133	104	117
1st Instar	172	167	174
SG	24	25	24
Testis	941	857	848
All Tissues	3277	3274	2610
Total	6806	6638	6673
All Genes Expressed in Tissue			
AG	4040	4121	3775
Head	5338	5268	5318
1st Instar	5380	5360	5383
SG	3699	3657	2984
Testis	5713	5620	5496

AG, accessory gland plus ejaculatory duct; SG, salivary gland.

neomorphs were expressed at lower median TPMs than other genes expressed ancestrally in that tissue (Student's *t*-test, $P < 0.01$ for all comparisons) (Table 4).

We observed 83 amorphs exhibiting lack of expression in a single tissue, most of which (80) were polymorphic (Table S11). An additional six amorphs exhibited lack of expression in two or more tissues. Fifty-seven amorphs (69%) were identified in only one of the six RAL genotypes. While we chose a cutoff of TPM < 0.1 to identify a candidate gene as not expressed in a given strain, we found that for 66% of candidates, one or more strains had TPM = 0. Indeed, 26 amorphs (31%) were expressed at TPM < 2 in all six RAL lines, and expression of amorphs in the RAL lines that retained expression was lower than that of ancestrally expressed genes in the testis and AG (Student's *t*-test, $P < 1.7\ \mathrm{e}^{-5}$ for both tests). Similar to neomorphs, amorphs identified in the RAL strains generally showed reduced expression in the F1s. For 32 of 63 instances where one RAL parent exhibited TPM = 0, expression in the F1 was reduced relative to the expressing parent. Of the 17 cases where both RAL parents had TPM = 0, 13 F1s exhibited TPM = 0.

As we observed for the polymorphic neomorphs, the testis and the AG differ from the other three tissues in the number of polymorphic amorphs. Indeed, the testis exhibits 67% of all polymorphic loss-of-expression phenotypes, with another 25% in the AG. Moreover, a large proportion of the polymorphic testis expression losses derived from a single line, RAL360. The testis RNA-seq data from this genotype do not differ from those of other genotypes with respect to number of reads or overall alignment statistics. The RAL360 chromosomes express typically in the testis when heterozygous against RAL517, which suggests that the genetic basis of the loss-of-expression phenotypes for RAL360 is not generally additive (Glaser-Schmitt *et al.* 2018), consistent with previous observations of frequent nonadditivity of expression in *Drosophila* (Gibson *et al.* 2004; Wayne *et al.* 2004; Massouras *et al.* 2012). The first-instar larvae exhibited only three amorphs, one of which was fixed.

Table 3 Mean TPM of top candidate fixed neomorphs

Gene	Ch	Tissue	D. melanogaster	F1s	D. simulans	D. yakuba
Marf1	2L	AG	14.28	8.69	0.02	0.01
CG32816	X	AG	9.80	5.41	0.05	0.00
CG15824	2L	Head	8.07	5.30	0.01	0.00
CG44227	3R	Testis	83.76	76.22	0.00	0.00
CG14662	3R	Testis	10.20	10.97	0.04	0.05
lmd	3R	Testis	5.67	7.76	0.07	0.03

AG, accessory gland plus ejaculatory duct; Ch, chromosome.

Table 4 Median expression of neomorphs and amorphs

	All orthologs		Neomorphs		Amorphs	
Tissue	n	Median	n	Median	n	Median
AG	4495	7.71	31	4.07	12	4.09
Head	5525	23.43	21	4.92	2	4.62
1st instar	5561	19.27	4	3.32	2	44.87
SG	4224	6.86	12	4.71	1	4.1
Testis	6074	16.78	17	5.49	40	10.62

For each gene expressed at TPM ≥ 2 in one or more RAL strain, the median value of the expressing strains was calculated. The median of all expressed genes per tissue was then calculated. AG, accessory gland plus ejaculatory duct; SG, salivary gland; TPM, transcripts per million.

Comparisons with other data sets

Many polymorphic neomorphs are expressed at relatively low levels in the tissues where they are detected and most (63%) were expressed in only one line. Similarly, most amorphs are polymorphic and are also expressed at low levels in the genotypes where they are expressed. To seek additional support for the novel expression patterns of these genes in the appropriate tissues, we investigated the expression of our candidate neomorphs and amorphs in both FlyAtlas2 (Leader *et al.* 2018) and in the Yang *et al.* (2018) testis data set.

For 11 of 17 testis neomorphs we observed mean expression at TPM ≥ 2 in at least one of the *D. melanogaster* strains used by Yang *et al.* (2018) (Table S3). For seven testis neomorphs we observed mean TPM ≥ 2 in both genotypes. These neomorphs were generally expressed at moderate levels in the Yang *et al.* (2018) data set with mean TPM = 14.2. For testis amorphs we found that 30 of 51 had a mean TPM of < 0.1 in both strains from Yang *et al.* (2018), and an additional 6 had a mean TPM < 0.1 in one strain.

Similarly, we found that 10 of 17 testis neomorphs exhibited expression at FPKM ≥ 1 in the FlyAtlas2 testis data (Table S4). For 18 out of 31 AG neomorphs the FlyAtlas2 FPKM estimates were ≥1. In the same database we found evidence of expression for only 2 out of 21 head neomorphs and 1 out of 12 SG neomorphs, though this is not surprising given that many were expressed in only one RAL line. For amorphs, we found that 29 of 83 have an FPKM < 1 in FlyAtlas2, including 16 out of 22 AG amorphs. Overall, the results from external data sets strongly support the conclusions from analysis of our own data.

To address the possibility that we have erroneously inferred instances of parallel loss-of-tissue expression in *D. yakuba* and *D. simulans* as testis or AG gains, we used existing testis RNA-seq data from *D. ananassae* (Yang *et al.* 2018; GSE99574) to determine whether our parsimony-based criterion for inferring *D. melanogaster* gains in our three-species analysis was consistent with data from an additional outgroup. The existing *D. ananassae* testis data included replicated FPKM estimates for 15 of the 18 testis neomorphs. Of the six fixed neomorphs for which *D. ananassae* data were available, all had mean TPM < 0.1, strongly supporting our inference that these genes were not expressed in the testis of the *D. melanogaster*–*D. simulans* ancestor.

Biological attributes of neomorphs

GO analysis of neomorphs and amorphs revealed no enrichments, which is not unexpected for such short gene lists. However, we do find a number of interesting gene features.

Multiple AG neomorphs have potential transcription factor activity. Two polymorphic AG neomorphs, *CG34026* and *CG30413*, are predicted to have multifactor bridgin protein 2 transcription activator domains, while a third, *kumgang*, is also a transcription factor. *CG34244*, which is polymorphically expressed in the AG, contains a Kazal domain, which is found in several genes ancestrally expressed in the AG that are also thought to be accessory gland proteins (Begun *et al.* 2006; Dottorini *et al.* 2007; Sirot *et al.* 2008). The predicted protein of *CG34244* has a strongly predicted signal sequence (Nielsen 2017) and thus is likely secreted, raising the possibility that it represents a transition from a secreted function in other tissues [*e.g.*, it is expressed in the prepupal SG (FlyBase)] to a secreted function in the AG. Consistent with FlyBase, we observed low levels of SG expression (0.1 < TPM < 2) in four of six RAL lines. The gene *Ejaculatory bulb protein II* (*EpbII*) is, despite its name, expressed at a very high level in the accessory gland of RAL399 (TPM = 1395), as well as at low-to-intermediate levels in the AGs of two other RAL lines and all three F1 crosses. Thus, the range of TPM estimates for this gene across RAL strains is 0–1395. The substantial AG expression of this gene in our data, in at least some strains, is consistent with expression estimates from the modENCODE (model organism ENCyclopedia Of DNA Elements) tissue expression data (The modENCODE Consortium *et al.* 2010; Graveley *et al.* 2011; Brown *et al.* 2014).

Our interest in the evolution of the AG transcriptome, including the component that codes for secreted proteins (von Heijne 1990), motivated us to investigate the possible general role of secretion in neomorph evolution. We used SignalP version 4.1 (Nielsen 2017) to identify predicted signal sequences for all neomorphs for each tissue and compared the relative abundances of predicted signal sequences to those of all expressed orthologs. Given their glandular function, we had hypothesized that we would find enrichment of predicted signal sequences in the AG and SG, but were surprised to find enrichment among neomorphs in each of the five tissues (Table 5). However, the two fixed AG neomorphs have no

predicted signal sequences, suggesting they are unlikely to be transferred to females in the seminal fluid. This raises interesting questions about what their functional effects might be and how they might influence fitness.

The fixed testis neomorphs are the transcription factor *lame duck* and two genes of unknown function, *CG44227* and *CG14662*; all three genes are germline expressed (Witt *et al.* 2019). These genes all express at a consistently high level across all six RAL strains (with mean RAL TPMs for the three genes of 5.7, 83.8, and 10.2, respectively), and yet show no evidence of outgroup testis expression. Among the strongest polymorphic testis neomorphs, defined as exhibiting at least one RAL strain with TPM = 0 and expressing at TPM > 2 in multiple RAL strains, are *CG43315*, *phyllopod*, *CG44037*, *Osi14*, *Osi23*, *CG7031*, *ppk25*, and *CG8960*. Most of these genes have no known function, with the exception of *phyllopod* and *ppk25*. *Phyllopod* plays a role in ubiquitination, regulates *Notch*, *Wnt*, and *sevenless* signaling pathways (Dickson 1995; Nagaraj and Banerjee 2009), and is germline expressed (Witt *et al.* 2019). Interestingly, although the function of *CG8960* is unknown, it physically interacts with *CG5289* (Guruharsha *et al.* 2011), which also plays a role in ubiquitin-related proteolysis (Wójcik and DeMartino 2002). The gene *ppk25* is a sodium channel that is important in male courtship behavior (Lin *et al.* 2005).

The appearance of six polymorphic chorion protein genes (whose known functions are to contribute to the structure of the egg) in the head is unexpected and deserving of further attention. Four of these genes are expressed in RAL304, with three of the four genes exhibiting TPM > 10 and low-to-modest expression (TPM < 2) in some other RAL lines. Two of the four genes, *Cp16* and *Cp19*, are located in tandem on *3L* while two others, *Cp36* and *Cp38*, are arranged tandemly on the *X* chromosome. The other two head-expressed chorion protein genes are *Cp7Fc* and *Cp7Fb*, both expressed in RAL360 and again, located tandemly on the *X* chromosome. To investigate whether there is independent evidence of polymorphic expression of chorion protein genes in female heads we used the data from Osada *et al.* 2017. Their results, obtained by crossing 18 RAL strains to a standard reference strain, followed by RNA-seq analysis of female heads, also provided repeatable support for expression in chorion protein genes in the heads of significant numbers of genotypes, in some cases at levels similar to those observed here. Finally, we point out that one gene, *Reepl1*, which plays a role in male-germline cyst formation (Yang *et al.* 2017), exhibited TPM = 30.8 in the head of one line, but TPM near 0.1 in the other five lines (and TPM = 0 in outgroup strains); the biological significance, if any, of such rare, high-expression phenotypes remains to be determined.

Tissue specificity in neomorphs and amorphs

Both neomorphs and amorphs exhibit significantly higher tissue specificity bias, τ (Yanai *et al.* 2005) (Table 6), than one-to-one orthologs expressed in the same tissue. This difference persists when we examine the differences in τ in each

Table 5 Predicted signal sequence in neomorphs

Tissue	Neomorphs with predicted signal sequence/total neomorphs	Orthologs with predicted signal sequence/total orthologs	P-value
AG	17/31	435/4040	1.05E−09
Head	13/21	734/5338	4.03E−08
1st Instar	4/4	796/5380	0.00E+00
SG	10/12	385/3699	1.67E−10
Testis	8/17	636/5713	2.69E−05

P-values represent the probability that the proportion of neomorphs with predicted signal sequences is the same as the proportion of all neomorphs with predicted signal sequences. AG, accessory gland plus ejaculatory duct; SG, salivary gland.

sex separately. For example, the male τ estimate for both testis-expressed neomorphs and testis amorphs were significantly higher than for all testis-expressed orthologs (Student's *t*-tests, $P = 1.56e^{-7}$ for neomorphs and $P = 2.84e^{-9}$ for amorphs). A similar pattern of higher τ for neomorphs and amorphs was observed for the AG (Student's *t*-test, neomorphs $P = 8.38e^{-18}$ and amorphs $P = 1.33e^{-9}$). Female heads showed a significantly elevated τ for neomorphs *vs.* orthologs (Student's *t*-test $P = 2.87e^{-11}$), but no difference for amorphs (Student's *t*-test, $P = 0.165$). We did not observe a difference in τ for SG neomorphs *vs.* orthologs expressed in the SG (Student's *t*-test, $P = 0.064$). We did not perform comparisons for the SG amorphs or the first-instar larvae due to small sample sizes. Overall, we conclude that genes with narrower ancestral tissue expression patterns appear to be more likely to evolve neomorphic and amorphic expression.

To investigate whether expression in ancestral tissues is predictive of expression gains in a novel tissue, we identified for each testis and AG neomorph (both fixed and polymorphic) a ranked descending list of expression in each tissue as estimated from FlyAtlas2 (Leader *et al.* 2018), other than the tissue in which the neomorph was expressed in our data. We compared these observed ranked tissue FPKMs to an empirical distribution of ranked tissue FPKMs generated by resampling genes from the set of orthologs expressed in that same tissue in FlyAtlas2 (Leader *et al.* 2018). We found that only 3 of the 26 AG neomorphs with FlyAtlas2 data showed the testis as the tissue (AG excluded) with the highest level of expression, fewer than expected when compared to the resampled genes expressed in the AG (binomial test, $P = 0.035$). None of the testis neomorphs showed the AG as the tissue with the highest expression. Indeed, four of eight testis neomorphs with FlyAtlas2 expression showed expression only in the testis in the FlyAtlas2 data set. These genes also either showed no or very low expression in *D. simulans* and *D. yakuba* for all tissues; for these genes, we have no insight into their ancestral expression pattern. Overall, there appears to be no strong relationship between AG and testis regarding neomorph expression. We also find that for four AG neomorphs and one testis neomorph, the tissue with the second highest level of expression is the brain. The midgut is the tissue with the second highest expression for four AG neomorphs.

Table 6 Tissue specificity (τ) in candidate gene sets

Sex	Orthologs	Neomorph		Amorph	
		Mean τ	*P*-value	Mean τ	*P*-value
Female	0.67	0.90	3.11E−29	0.89	5.99E−17
Male	0.72	0.93	1.43E−37	0.90	2.82E−13

To determine whether these features of AG and testis neomorphs differ from those of all orthologs, we used expression data from 12 male tissues from FlyAtlas2 to identify the genes showing the highest expression level in the AG, and the second highest in the testis and vice versa in all male-expressed genes. For genes expressed at the highest level in the testis, 22.4% have their second highest expression in the AG. Similarly, for genes expressed at the highest level in AG, 26% have their second highest expression in the testis. Thus, for all orthologs the correlation between these two tissues is quite strong and highly significant against the null hypothesis that gene expression in the two tissues is uncorrelated ($P = 2.29e^{-56}$ for testis and $P = 5.76e^{-51}$ for AG). We performed GO analysis (Huang *et al.* 2009a,b) and found that this combined set of genes is enriched for several terms including multicellular organism reproduction, extracellular space, and proteasome complex (Table S12). The fact that the correlation between testis and AG is stronger for all orthologs than for neomorphs suggests that neomorphic male reproductive expression is likely not strongly related to ancestral male reproductive function. Other patterns of association between tissues were also identified. For the genes expressing at the highest level in testis, 16% exhibited their second highest expression in male brain, while 15% had their second highest expression in the malpighian tubules. For the 30% of genes most highly expressed in the AG, their second highest expression was in the adult male SG, suggesting correlated functions of these two glandular tissues.

Genetics underlying expression differences

To characterize the regulatory genetics of neomorphs and amorphs, we used data from six RAL parents and their three distinct F1s to partition expression variation into *cis*-acting components *vs.* *trans*-acting components in testis and AG tissues following McManus *et al.* (2010), with some differences due to experimental design (see *Materials and Methods*). We had sufficient information (see *Materials and Methods*) to make calls about the regulatory genetics for one or more crosses in 31 of 48 (65%) of neomorphs and 41 of 73 (56%) of amorphs (Table S9). We found that observations indicating both *cis* and *trans* effects as contributing to novel expression phenotypes were common: 43% of AG neomorph and 34% of testis neomorph observations exhibited both *cis* and *trans* effects. Purely *cis* effects accounted for a smaller proportion of neomorph observations, representing 14% of AG and 24% of testis observations. One candidate neomorph exhibiting strong cis-effects is *phyl* (Figure 2), the data from which suggest that it is associated with novel, polymorphic

testis enhancer. In total, we find that 57% of AG neomorphs exhibit evidence of *cis* effects in one or more crosses, as do 58% of testis neomorphs. Thus, neomorphic expression appears to be influenced by both substantial *cis* and substantial *trans* effects. Amorphs in both the AG and testis showed much higher proportions of pure *trans* effects; 54% of AG observations and 58% of testis observations. The low population frequencies of amorphic expression patterns along with the larger *trans* component of regulatory variation relative to neomorphs could be consistent with the idea that *trans*-mediated drivers of novel expression may be more deleterious, on average, than novel *cis* effects.

Major interspecific quantitative expression divergence within tissues

Expanding our analysis to reveal genes exhibiting major *D. melanogaster*-specific quantitative within-tissue expression variation, but that did not necessarily represent qualitatively novel phenotypes, we identified all genes exhibiting a *D. simulans vs. D. melanogaster* fold difference ≥1.25 and a *D. simulans vs. D. yakuba* fold difference < 1.25. We identified 64 such genes (Table S13), only nine of which showed evolution of lower expression in *D. melanogaster*, many fewer than expected under the null hypothesis that major up- and downregulation was equally likely ($P = 1.77e^{-9}$). Of these 64 genes, three were candidate neomorphs. For 55 genes, expression was at least 3.3 times higher in *D. melanogaster* than *D. simulans*, with a maximum fold difference of 378. Using a more conservative criterion of TPM > 2 in all RAL lines preserved 39 out of 55 major expression increase candidates in *D. melanogaster*. Twenty-one of the 55 candidate genes (38%) exhibited the major *D. melanogaster*-specific expression increase in the testis (binomial test, $P = 0.006$), consistent with previous results suggesting that testis-related phenotypes have unusual evolutionary properties (Parisi *et al.* 2004; Graveley *et al.* 2011).

Of the 18 genes showing at least 10-fold greater mean TPM in *D. melanogaster* than in *D. simulans*, and that were not among our neomorphs, seven showed expression divergence in the testis, seven in the AG, 2 in the head, and 3 in the first-instar larva. Thus, the broad patterns of major quantitative tissue expression evolution are comparable to patterns of candidate neomorphs. While some of the 55 candidate genes exhibit very low expression in *D. simulans/D. yakuba* and modest expression in *D. melanogaster*, 28 exhibited mean expression in *D. melanogaster* of TPM ≥ 10, around the median expression of all genes (Table S6). A good example of such a gene is *CG6106*, which is annotated as metal-dependent hydrolase (www.flybase.org) and shows very low *D. melanogaster* expression across multiple tissues in the modENCODE and FlyAtlas2 data. However, in our data this gene was expressed in the AG at a mean TPM of 24.2 across the six *D. melanogaster* genotypes, a 158-fold increase relative to *D. simulans*. The lack of significant *D. melanogaster* AG expression for this gene in modENCODE and FlyAtlas2 likely reflects the high within-species expression variance for this gene: TPM ranged

from 0.2 to 41 across the six RAL genotypes. Thus, we speculate that the strains used to generate the data used for the public databases happened to be low-expressing genotypes. Other genes with highly unusual AG expression increases in *D. melanogaster* were *CG43210* (92-fold greater than *D. simulans*), *CG43678* (65-fold greater than *D. simulans*), and *CG6793* (43-fold greater than *D. simulans*).

The genes showing the greatest *D. melanogaster* testis fold changes were *Hr51* (21-fold), *CG43470* (20-fold), *fs(1)N* (16-fold), and *CG43402* (15-fold), which happens to be located 15 kb downstream of one of the testis candidate neomorphs (*CG43401*). While *fs(1)N* is named based on its maternal effect sterile phenotype, our results are consistent with FlyAtlas2, which also shows moderate enrichment of the gene in the testis. Also notable is *CG43402*, due to its high mean TPM estimate of 89.4 compared to roughly TPM = 6 in *D. simulans* and *D. yakuba*. Among the other genes showing substantially increased expression in *D. melanogaster* are *Vps28*, which codes for a ubiquitin-dependent protein that plays a role in sperm individualization; *maelstrom*, which plays a role in piwi-interacting RNA transposon silencing and stem cell regulation; the nuclear pore protein, *Nup160*; and *ppk6*, which codes for a cation channel.

The fact that testis neomorph *phyl* and testis upregulation candidate *Vps28* both function in ubiquitination (Figure 2), and the fact that *CG8960*, also a testis neomorph, physically interacts with a partner implicated in ubiquitin-related catabolism, suggests the possibility that ubiquitin-mediated processes have experienced recent selection favoring novelty in the *D. melanogaster* testis.

Discussion

While understanding the evolutionary mechanisms underlying organ expression evolution requires building connections between variation observed within species and differences that accumulate between species, population level investigation of organ expression is rare for any organism, including *Drosophila*, with the notable exception of Dickinson (1975), who observed changes in developmental expression patterns in aldehyde oxidase via evolution of a polymorphic *cis*-regulatory element in *D. melanogaster*. Here, we have sought to contribute to this literature using population transcriptome data from five *D. melanogaster* tissues along with outgroup data from *D. simulans* and *D. yakuba* to investigate how tissue expression in a collection of 1:1:1 orthologs has evolved specifically in the *D. melanogaster* lineage, or roughly in the order of 1.4–3.4 million years (or less in the case of polymorphism) (Obbard *et al.* 2012).

We observed a large number of novel expression phenotypes with several notable properties. First, most are polymorphic. While many polymorphic neomorphs are expressed in only a single RAL genotype and thus may be rare in natural populations, 33% are expressed in at least two lines at a TPM ≥ 2 and thus may be common phenotypes in *D. melanogaster* populations. While replicated RNA-seq libraries would increase our

confidence in identifying polymorphic novel expression phenotypes, the use of conservative cutoffs and validation in orthogonal RNA-seq data from F1 animals strongly supports the biological reality of our observations. Also strongly supporting our general findings, for 32 out of 85 neomorphs we identified expression in the appropriate tissue in FlyAtlas2 (Leader *et al.* 2018) and observed expression of 11 out of 17 testis neomorphs in independent, previously published *D. melanogaster* testis transcriptome data (Yang *et al.* 2018). Second, a majority of novel expression phenotypes are influenced by *cis*-acting variation, though *trans* effects are also common. Third, neomorphic expression tends to arise from genes that are ancestrally more tissue-biased in their expression. Fourth, and not surprisingly based on previous literature (Parisi *et al.* 2004; Graveley *et al.* 2011), male reproductive tract tissues exhibited the greatest number of novel expression phenotypes.

Phenotypic and fitness consequences of expression novelties

Among the outstanding questions that remain to be addressed are the possible role of selection in generating the novel expression phenotypes, and the related question of how or whether most of the expression variation observed here relates to intermediate and organismal phenotypes. The question of selection requires additional work. Our sample size of fixed neomorphs was small and several of these fixed neomorphs were located in regions of low recombination, where signatures of selection on these genes would be hard to disentangle from other properties of those genomic regions.

The issue of the selective spread of segregating neomorphs cannot be addressed incisively given the existing data or analyses for two main reasons. First, the complexity of the genetics of regulation of the observed neomorphs suggests that looking for evidence of recent selection near such genes resulting from fitness differences of regulatory elements will be challenging, as the targets of any selection may be widely distributed in the genome, and accordingly, several distinct genotypes may generate the same (or similar) polymorphic neomorphic phenotype. Second, because we observed relatively few intermediate frequency polymorphic neomorphs in our sample of six RAL strains, and because the expression phenotypes of additional chromosomes not studied here cannot be predicted from the sequences in or near candidate genes (given our limited understanding of the expression genetics), we cannot carry out proper population genetic tests for ongoing sweeps favoring expressing chromosomes in larger samples of RAL chromosomes. Moreover, there are no published whole-genome summaries of such sweep statistics available for *D. melanogaster*.

Are there other properties of the variation observed here that may shed additional light on evolutionary mechanisms? Because the average fitness effects of rare, common, or fixed variation may be heterogeneous, even our sample size of only six *D. melanogaster* genotypes may be informative. For example, many polymorphic neomorphs were observed only once in the six RAL genotypes. While the frequency of the observed
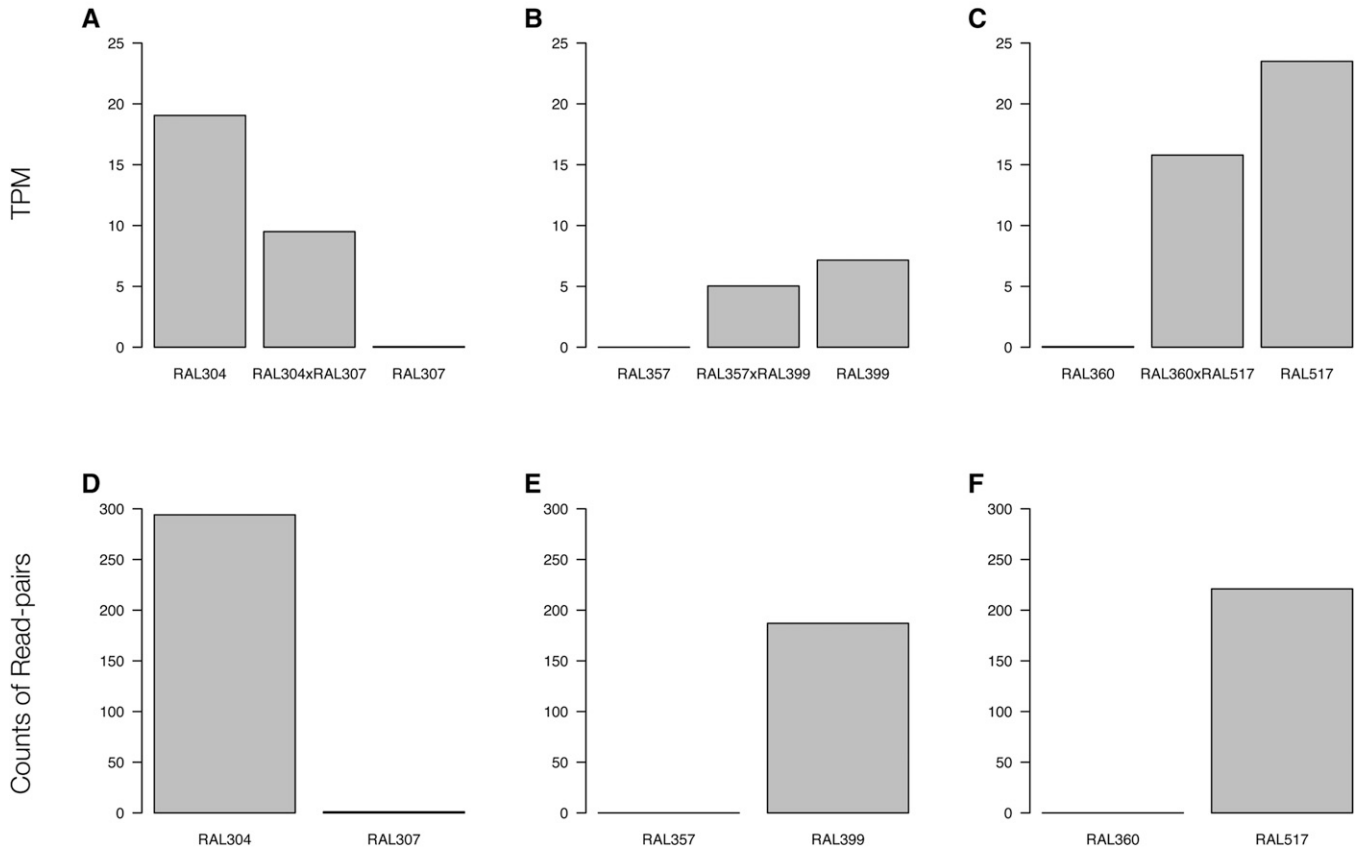
Figure 2 Expression genetics of phyllopod, a testis-expressed polymorphic neomorph. (A–C) Transcripts per million (TPM) estimates for *phyl* in parental genotypes and associated F1 genotypes. (D–F) Counts of parent-specific read-pairs identified for each of three F1 genotypes.

phenotypes cannot be used to estimate the frequencies of genetic variants underlying those phenotypes, one interpretation of the large number of singleton neomorphs is that they tend to reduce fitness. If this is the case, why would natural selection tolerate such abundant incidental gene expression? One possibility, for which there is some support in our data, is that neomorphs may in some cases result from epistatic (*e.g., cis × trans*) interactions. If incidental, deleterious expression phenotypes often result from epistatic interactions among a substantial collection of low-frequency variants that have small main effects, such variants would not effectively be removed from populations as they would rarely be exposed to selection.

It is also worth noting that for the singleton chromosomes whose novel phenotypes are not recapitulated in the F1 as strong allelic imbalance, we should entertain the possibility that each inbred line likely harbors deleterious mutations, at least a fraction of which are likely strongly recessive. This too, may explain some of the rarely observed novel phenotypes. For example, RAL360 is the line in which testis amorphs are identified for 27 genes, which constitutes 63% of all the polymorphic loss-of-testis-expression phenotypes among the RAL strains. However, the RAL360 genome generally

expresses normally in the testis of the F1. A plausible explanation for this observation is the presence of one or more recessive deleterious mutations that influence expression levels of several testis expressed genes in this strain. In contrast to rare, novel expression phenotypes, common novel phenotypes are less likely to be deleterious, though whether or how they might influence fitness is an open question.

Conclusions

The data and analytical approaches we have used here raise interesting questions regarding the difficulties of trying to understand the connection between tissue-based gene expression and population genetic and evolutionary mechanisms. For example, we used arbitrary criteria applied uniformly across genes, tissues, and species to identify candidate genes that may have evolved novel expression phenotypes. Our approach has a number of drawbacks.

First, it is unlikely that the arbitrary cutoffs used here are optimal for all genes and tissues, and the biological implications of a given expression level for a given gene are both unknown, and likely vary substantially between both genes and the cellular environment. However, the expression criteria we used seem conservative, and an exploration of other

expression cutoffs for identifying qualitatively novel phenotypes did not meaningfully change our conclusions.

Second, and related, while there is no doubt that data from dissected animals are better than whole animal data, the nature of some of our samples complicates some interpretations. For example, we noted differences in the properties of expression variation in female heads and first-instar larvae *vs.* testis, AG, and SG. While our data are a worthwhile contribution to the literature (Huylmans and Parsch 2014; Sanfilippo *et al.* 2017; Glaser-Schmitt *et al.* 2018), our comparisons are potentially weakened because the head and first-instar larvae are comprised of many different tissue types compared to the testis, AG, and SG. Indeed, heterogeneity in the biological complexity of our defined tissues may be driving a significant amount of the variation across "tissues" observed in our samples.

This raises the more general question of the level or levels of biological organization at which expression interacts with downstream phenotypes and natural selection, and thus the level(s) of organization at which such phenomena should be studied. For example, there are two possible explanations for why a gene may be expressed at the whole-organ level in one genotype or species and not another. A gene may be expressed in a similar manner across tissue or cell types within an organ across species, but at different expression levels. Alternatively, a gene may be expressed at a similar level for a given tissue type or cell type across genotypes, but genotypes may differ in the proportion of cells or tissue types present in an organ. These different phenotypes are biologically and developmentally distinct, and may interact in different ways with downstream biology and natural selection; thus our lack of knowledge on such issues is a major obstacle. More generally, the fascinating question as to whether most of the details of where in the organism a gene is expressed are the direct result of natural selection remains open.

The recent introduction of single-cell transcriptome analysis holds much promise as a tool to begin attacking these difficult problems. However, even single-cell transcriptomics will not be a panacea, as questions about protein abundance and activity, as well as cell- or tissue-specific phenotypes, will remain. The analysis of tissue- or cell type-specific knockdowns or knockouts, which is accessible in model systems such as *D. melanogaster* (Wheeler *et al.* 2004; Poe *et al.* 2019), will be vital for future advances.

Finally, our use of parsimony, while likely effective on the short timescales investigated here, is suboptimal and likely inappropriate for longer timescales or very rapid evolutionary change associated with the important investigation of repeated evolution of novel expression phenotypes. More sophisticated phylogenetic and statistical approaches will be preferable in such situations.

Acknowledgments

Literature Cited

Begun, D. J., and H. A. Lindfors, 2005   Rapid evolution of genomic amp complement in the *melanogaster* subgroup of *Drosophila*. Mol. Biol. Evol. 22: 2010–2021. https://doi.org/10.1093/molbev/msi201

Begun, D. J., P. Whitley, B. L. Todd, H. M. Waldrip-Dali, and A. G. Clark, 2000   Molecular population genetics of male accessory gland proteins in Drosophila. Genetics 156: 1879–1888.

Begun, D. J., H. A. Lindfors, M. E. Thompson, and A. K. Holloway, 2006   Recently evolved genes identified from *Drosophila yakuba* and *D. erecta* accessory gland expressed sequence tags. Genetics 172: 1675–1681. https://doi.org/10.1534/genetics.105.050336

Brakefield, P. M., J. Gates, D. Keys, F. Kesbeke, P. J. Wijngaarden *et al.*, 1996   Development, plasticity and evolution of butterfly eyespot patterns. Nature 384: 236–242. https://doi.org/10.1038/384236a0

Brown, J. B., N. Boley, R. Eisman, G. E. May, M. H. Stoiber *et al.*, 2014   Diversity and dynamics of the *Drosophila* transcriptome. Nature 512: 393–399. https://doi.org/10.1038/nature12962

Chintapalli, V. R., J. Wang, and J. A. T. Dow, 2007   Using FlyAtlas to identify better *Drosophila melanogaster* models of human disease. Nat. Genet. 39: 715–720. https://doi.org/10.1038/ng2049

Dickinson, W. J., 1975   A genetic locus affecting the developmental expression of an enzyme in *Drosophila melanogaster*. Dev. Biol. 42: 131–140. https://doi.org/10.1016/0012-1606(75)90319-X

Dickinson, W. J., 1980   Tissue specificity of enzyme expression regulated by diffusible factors: evidence in *Drosophila* hybrids. Science 207: 995–997. https://doi.org/10.1126/science.7352303

Dickson, B., 1995   Nuclear factors in *sevenless* signaling. Trends Genet. 11: 106–111. https://doi.org/10.1016/S0168-9525(00)89011-3

Doniger, S. W., and J. C. Fay, 2007   Frequent gain and loss of functional transcription factor binding sites. PLoS Comput. Biol. 3: e99 [corrigenda: PLoS Comput. Biol. 5 (2009)]. https://doi.org/10.1371/journal.pcbi.0030099

Dottorini, T. L., H. Nicolaides, D. W. Ranson, A. Rogers, A. Crisanti *et al.*, 2007   A genome-wide analysis in *Anopheles gambiae* mosquitoes reveals 46 male accessory gland genes, possible modulators of female behavior. Proc. Natl. Acad. Sci. USA 104: 16215–16220. https://doi.org/10.1073/pnas.0703904104

Gibson, G., R. Riley-Berger, L. Harshman, A. Kopp, S. Vacha *et al.*, 2004   Extensive sex-specific nonadditivity of gene expression in *Drosophila melanogaster*. Genetics 167: 1791–1799. https://doi.org/10.1534/genetics.104.026583

Glaser-Schmitt, A., A. Zečić, and J. Parsch, 2018   Gene regulatory variation in Drosophila melanogaster renal tissue. *Genetics* 210: 287–301. https://doi.org/10.1534/genetics.118.301073

Glassford, W. J., W. C. Johnson, N. R. Dall, S. J. Smith, Y. Liu *et al.*, 2015   Co-option of an ancestral Hox-regulated network underlies a recently evolved morphological novelty. Dev. Cell 34: 520–531. https://doi.org/10.1016/j.devcel.2015.08.005

Gompel, N., B. Prud'homme, P. J. Wittkopp, V. A. Kassner, and S. B. Carroll, 2005   Chance caught on the wing: *cis*-regulatory evolution and the origin of pigment patterns in *Drosophila*. Nature 433: 481–487. https://doi.org/10.1038/nature03235

Graveley, B. R., A. N. Brooks, K. W. Carlson, M. O. Duff, J. M. Landolin *et al.*, 2011   The developmental transcriptome of *Drosophila melanogaster*. Nature 471: 473–479. https://doi.org/10.1038/nature09715

Guruharsha, K. G., J. F. Rual, B. Zhai, J. Mintseris, P. Vaidya *et al.*, 2011   A protein complex network of *Drosophila melanogaster*. Cell 147: 690–703. https://doi.org/10.1016/j.cell.2011.08.047

Huang, D. W., B. T. Sherman, and R. A. Lempicki, 2009a   Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. 4: 44–57. https://doi.org/10.1038/nprot.2008.211

Huang, D. W., B. T. Sherman, and R. A. Lempicki, 2009b Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res. 37: 1–13. https://doi.org/10.1093/nar/gkn923

Huylmans, A. K., and J. Parsch, 2014 Population- and sex-biased gene expression in the excretion organs of Drosophila melanogaster. *G3 (Bethesda)* 4: 2307–2315. https://doi.org/10.1534/g3.114.013417

Kim, D., B. Langmead, and S. L. Salzberg, 2015 HISAT: a fast spliced aligner with low memory requirements. Nat. Methods 12: 357–360. https://doi.org/10.1038/nmeth.3317

Kopp, A., 2011 Drosophila sex combs as a model of evolutionary innovations. *Evol. Dev.* 13: 504–522. https://doi.org/10.1111/j.1525-142X.2011.00507.x

Koshikawa, S., M. W. Giorgianni, K. Vaccaro, V. A. Kassner, J. H. Yoder *et al.*, 2015 Gain of *cis*-regulatory activities underlies novel domains of wingless gene expression in Drosophila. Proc. Natl. Acad. Sci. USA 112: 7524–7529. https://doi.org/10.1073/pnas.1509022112

Lack, J. B., C. M. Cardeno, M. W. Crepeau, W. Taylor, R. B. Corbett-Detig *et al.*, 2015 The *Drosophila* genome nexus: a population genomic resource of 623 *Drosophila melanogaster* genomes, including 197 from a single ancestral range population. Genetics 199: 1229–1241. https://doi.org/10.1534/genetics.115.174664

Leader, D. P., S. A. Krause, A. Pandit, S. A. Davies, and J. A. T. Dow, 2018 FlyAtlas 2: a new version of the *Drosophila melanogaster* expression atlas with RNA-Seq, miRNA-Seq and sex-specific data. Nucleic Acids Res. 46: D809–D815. https://doi.org/10.1093/nar/gkx976

Lin, H., K. J. Mann, E. Starostina, R. D. Kinser, and C. W. Pikielny, 2005 A *Drosophila DEG/ENaC* channel subunit is required for male response to female pheromones. Proc. Natl. Acad. Sci. USA 102: 12831–12836. https://doi.org/10.1073/pnas.0506420102

Long, H. K., S. L. Prescott, and J. Wysocka, 2016 Ever-changing landscapes: transcriptional enhancers in development and evolution. Cell 167: 1170–1187. https://doi.org/10.1016/j.cell.2016.09.018

Ma, S., A. S. Avanesov, E. Porter, B. C. Lee, M. Mariotti *et al.*, 2018 Comparative transcriptomics across 14 *Drosophila* species reveals signatures of longevity. Aging Cell 17: e12740. https://doi.org/10.1111/acel.12740

Mackay, T. F. C., S. Richards, E. A. Stone, A. Barbadilla, J. F. Ayroles *et al.*, 2012 The *Drosophila melanogaster* genetic reference panel. Nature 482: 173–178. https://doi.org/10.1038/nature10811

Mallarino, R., C. Henegar, M. Mirasierra, M. Manceau, C. Schradin *et al.*, 2016 Developmental mechanisms of stripe patterns in rodents. Nature 539: 518–523. https://doi.org/10.1038/nature20109

Martinson, E. O., Mrinalini, Y. D. Kelkar, C.-H. Chang, and J. H. Werren, 2017 The evolution of venom by co-option of single-copy genes. Curr. Biol. 27: 2007–2013.e8. https://doi.org/10.1016/j.cub.2017.05.032

Massouras, A. S. M., M. Waszak, K. Albarda-Aguilera, W. Hens, W. Holcombe *et al.*, 2012 Genomic variation and its impact on gene expression in *Drosophila melanogaster*. PLoS Genet. 8: e1003055. https://doi.org/10.1371/journal.pgen.1003055

McManus, J. C., J. D. Coolon, M. O. Duff, J. Eipper-Mains, B. R. Graveley *et al.*, 2010 Regulatory divergence in *Drosophila* revealed by mRNA-seq. Genome Res. 20: 816–825 [corrigenda: Genome Res. 24: 1051 (2014)]. https://doi.org/10.1101/gr.102491.109

Meiklejohn, C. D., J. Parsch, J. M. Ranz, and D. L. Hartl, 2003 Rapid evolution of male-biased gene expression in Drosophila. Proc. Natl. Acad. Sci. USA 100: 9894–9899. https://doi.org/10.1073/pnas.1630690100

Meslin, C., M. S. Plakke, A. B. Deutsch, B. S. Small, N. I. Morehouse *et al.*, 2015 Digestive organ in the female reproductive tract borrows genes from multiple organ systems to adopt critical functions. Mol. Biol. Evol. 32: 1567–1580. https://doi.org/10.1093/molbev/msv048

The modENCODE Consortium, Roy, S., J. Ernst, P. V. Kharchenko, P. Kheradpour *et al.*, 2010 Identification of functional elements and regulatory circuits by *Drosophila* modENCODE. Science 330: 1787–1797. https://doi.org/10.1126/science.1198374

Mueller, J. L., K. Ravi Ram, L. A. McGraw, M. C. Bloch Qazi, E. D. Siggia *et al.*, 2005 Cross-species comparison of *Drosophila* male accessory gland protein genes. Genetics 171: 131–143. https://doi.org/10.1534/genetics.105.043844

Nagaraj, R., and U. Banerjee, 2009 Regulation of notch and wingless signaling by *phyllopod*, a transcriptional target of the EGFR pathway. EMBO J. 28: 337–346. https://doi.org/10.1038/emboj.2008.286

Nielsen, H., 2017 Predicting secretory proteins with SignalP, pp. 59–73 in *Protein Function Prediction*, edited by D. Kihara. Humana Press, Totowa, NY. 10.1007/978-1-4939-7015-5_6 https://doi.org/10.1007/978-1-4939-7015-5_6

Oakeshott, J. G., M. J. Healy, and A. Y. Game, 1990 Regulatory evolution of β-carboxyl esterases in *Drosophila*, pp. 359–387 in *Ecological and Evolutionary Genetics of Drosophila*, edited by J. S. F. Barker, W. T. Starmer, and R. J. MacIntyre. Springer-Verlag, Boston, MA.

Obbard, D. J., J. MacIennan, K.-W. Kim, A. Rabat, P. M. O'Grady *et al.*, 2012 Estimating divergence dates and substitution rates in the *Drosophila* phylogeny. Mol. Biol. Evol. 29: 3459–3473. https://doi.org/10.1093/molbev/mss150

Ohno, S., 1970 *Evolution by Gene Duplication*. Springer-Verlag, New York.

Osada, N., R. Miyagi, and A. Takahashi, 2017 Cis- and trans-regulatory effects on gene expression in a natural population of Drosophila melanogaster. *Genetics* 206: 2139–2148. https://doi.org/10.1534/genetics.117.201459

Parisi, M., R. Nuttall, P. Edwards, J. Minor, D. Naiman *et al.*, 2004 A survey of ovary-, testis-, and soma-biased gene expression in Drosophila melanogaster adults. *Genome Biol.* 5: R40. https://doi.org/10.1186/gb-2004-5-6-r40

Pertea, M., D. Kim, G. M. Pertea, J. T. Leek, and S. L. Salzberg, 2016 Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. Nat. Protoc. 11: 1650–1667. https://doi.org/10.1038/nprot.2016.095

Poe, A. R., B. Wang, M. L. Sapar, H. Ji, K. Li *et al.*, 2019 Robust CRISPR/Cas9-mediated tissue-specific mutagenesis reveals gene redundancy and perdurance in *Drosophila*. Genetics 211: 459–472. https://doi.org/10.1534/genetics.118.301736

Rebeiz, M., N. Jikomes, V. A. Kassner, and S. B. Carroll, 2011 Evolutionary origin of a novel gene expression pattern through co-option of the latent activities of existing regulatory sequences. Proc. Natl. Acad. Sci. USA 108: 10036–10043. https://doi.org/10.1073/pnas.1105937108

Roeske, M. K., E. M. Camino, S. Grover, M. Rebeiz, and M. W. Thomas, 2018 *Cis*-regulatory evolution integrated the *Bric-a-brac* transcription factors into a novel fruit fly gene regulatory network. Elife 7: e32273. https://doi.org/10.7554/eLife.32273

Ross, J. L., P. P. Fong, and D. R. Cavener, 1994 Correlated evolution of the *cis*-acting regulatory elements and developmental expression of the *Drosophila Gld* gene in seven species from the subgroup *melanogaster*. Dev. Genet. 15: 38–50. https://doi.org/10.1002/dvg.1020150106

Sanfilippo, P., J. Wen, and E. C. Lai, 2017 Landscape and evolution of tissue-specific alternative polyadenylation across *Drosophila* species. Genome Biol. 18: 229. https://doi.org/10.1186/s13059-017-1358-0

Sirot, L. K., R. L. Poulson, M. C. McKenna, H. Girnary, M. F. Wolfner *et al.*, 2008 Identity and transfer of male reproductive gland proteins of the dengue vector mosquito, *Ades aegypti*: potential tools for control of female feeding and reproduction. Insect Biochem. Mol. Biol. 38: 176–189. https://doi.org/10.1016/j.ibmb.2007.10.007

Tomarev, S. I., and J. Piatigorsky, 1996 Lens crystallins of invertebrates: diversity and recruitment from detoxification enzymes and

novel proteins. Eur. J. Biochem. 235: 449–465. https://doi.org/10.1111/j.1432-1033.1996.00449.x

Thompson, A., M. R. May, B. R. Moore, and A. Kopp, 2020. A hierarchical Bayesian mixture model for inferring the expression state of genes in transcriptomes. PNAS. 117: 19339–19346. https://doi.org/10.1073/pnas.1919748117

Thorpe, P. A., J. Loye, C. A. Rote, and W. J. Dickinson, 1993 Evolution of regulatory genes and patterns: relationships to evolutionary rates and to metabolic functions. J. Mol. Evol. 37: 590–599. https://doi.org/10.1007/BF00182745

True, J. R., and S. B. Carroll, 2002 Gene co-option in physiological and morphological evolution. Annu. Rev. Cell Dev. Biol. 18: 53–80. https://doi.org/10.1146/annurev.cellbio.18.020402.140619

van Dongen, S., and C. Abreu-Goodger, 2012 Using MCL to extract clusters from networks. Methods Mol. Biol. 804: 281–295.

von Heijne, G., 1990 The signal peptide. J. Membr. Biol. 115: 195–201. https://doi.org/10.1007/BF01868635

Wagner, G. P., K. Kin, and V. J. Lynch, 2013 A model based criterion for gene expression calls using RNA-seq data. Theory Biosci. 132: 159–164. https://doi.org/10.1007/s12064-013-0178-3

Wang, W., H. Yu, and M. Long, 2004 Duplication-degeneration as a mechanism of gene fission and the origin of new genes in *Drosophila* species. Nat. Genet. 36: 523–527. https://doi.org/10.1038/ng1338

Wayne, M. L., Y.-J. Pan, S. V. Nuzhdin, and L. M. McIntyre, 2004 Additivity and *trans*-acting effects on gene expression in male *Drosophila simulans*. Genetics. 168: 1413–1420. https://doi.org/10.1534/genetics.104.030973

Wheeler, D. B., S. N. Bailey, D. A. Guertin, A. E. Carpenter, C. O. Higgins *et al.*, 2004 RNAi living-cell microarrays for loss-of-function screens in Drosophila melanogaster cells. Nat. Methods 1: 127–132. https://doi.org/10.1038/nmeth711

Witt, E., S. Benjamin, N. Svetec, and L. Zhao, 2019 Testis single-cell RNA-seq reveals the dynamics of de novo gene transcription and germline mutational bias in *Drosophila*. Elife 8: e47138. https://doi.org/10.7554/eLife.47138

Wittkopp, P. J., S. B. Carroll, and A. Kopp, 2003 Evolution in black and white: genetic control of pigment patterns in *Drosophila*.

Trends Genet. 19: 495–504. https://doi.org/10.1016/S0168-9525(03)00194-X

Wistow, G., 1993 Lens crystallins: gene recruitment and evolutionary dynamism. Trends Biochem. Sci. 18: 301–306. https://doi.org/10.1016/0968-0004(93)90041-K

Wistow, G. J., and J. Piatigorsky, 1988 Lens crystallins: the evolution and expression of proteins for a highly specialized tissue. Annu. Rev. Biochem. 57: 479–504. https://doi.org/10.1146/annurev.bi.57.070188.002403

Wójcik, C., and G. N. DeMartino, 2002 Analysis of *Drosophila* 26 S proteasome using RNA interference. J. Biol. Chem. 277: 6188–6197. https://doi.org/10.1074/jbc.M109996200

Wong, E. S. W., and K. Belov, 2012 Venom evolution through gene duplications. Gene 496: 1–7. https://doi.org/10.1016/j.gene.2012.01.009

Wray, G., 2007 The evolutionary significance of *cis*-regulatory mutations. Nat. Rev. Genet. 8: 206–216. https://doi.org/10.1038/nrg2063

Yanai, I., H. Benjamin, M. Shmoish, V. Chalifa-Caspi, M. Shklar *et al.*, 2005 Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. Bioinform. 21: 650–659. https://doi.org/10.1093/bioinformatics/bti042

Yang, H., M. Jamie, M. Polihronakis, K. Kanegawa, T. Markow *et al.*, 2018 Re-annotation of eight *Drosophila* genomes. Life Sci. Alliance 1: e201800156. https://doi.org/10.26508/lsa.201800156

Yang, S. Y., Y.-C. Chang, Y. H. Wan, C. Whitworth, E. M. Baxter *et al.*, 2017 Control of a novel spermatocyte-promoting factor by the male germline sex determination factor PHF7 of *Drosophila melanogaster*. Genetics 206: 1939–1949. https://doi.org/10.1534/genetics.117.199935

Zhao, L., J. Wit, N. Svetec, and D. J. Begun, 2015 Parallel gene expression differences between low and high latitude populations of Drosophila melanogaster and D. simulans. PLoS Genet. 11: e1005184. https://doi.org/10.1371/journal.pgen.1005184

Zhang, J., 2003 Evolution by gene duplication: an update. Trends Ecol. Evol. 18: 292–298. https://doi.org/10.1016/S0169-5347(03)00033-8

Communicating editor: M. Nachman