Gene protein sequence evolution can predict the rapid divergence of ovariole numbers in *Drosophila*

Research Article

Authors: Carrie A. Whittle[1,3], Cassandra G. Extavour[1,2,3]

Affiliations:

[1]Department of Organismic and Evolutionary Biology, Harvard University, 16 Divinity Avenue, Cambridge MA 02138, USA

[2]Department of Molecular and Cellular Biology, Harvard University, 16 Divinity Avenue, Cambridge MA 02138, USA

[3]Howard Hughes Medical Institute, Chevy Chase MD

Corresponding author: Cassandra G. Extavour

Email: extavour@oeb.harvard.edu

Phone: (617)-496-1935

Fax: (617)-496-9507

ORCIDs

C.A. Whittle 0000-0002-9331-0520

C.G. Extavour 0000-0003-2922-5855

Abstract

Ovaries play key roles in fitness and evolution: they are essential female reproductive structures that develop and house the eggs in sexually reproducing animals. In *Drosophila*, the mature ovary contains multiple tubular egg-producing structures known as ovarioles. Ovarioles arise from somatic cellular structures in the larval ovary called terminal filaments, formed by terminal filament cells and subsequently enclosed by sheath cells. As in many other insects, ovariole number per female varies extensively in *Drosophila*. At present however, there is a striking gap of information on genetic mechanisms and evolutionary forces that shape the well-documented rapid interspecies divergence of ovariole numbers. To address this gap, here we studied genes associated with *D. melanogaster* ovariole number or functions based on recent experimental and transcriptional datasets from larval ovaries, including terminal filaments and sheath cells, and assessed their rates and patterns of molecular evolution in five closely related species of the *melanogaster* subgroup that exhibit species-specific differences in ovariole numbers. From comprehensive analyses of protein sequence evolution (dN/dS), branch-site positive selection, expression specificity (*tau*) and phylogenetic regressions (PGLS), we report evidence of 42 genes that showed signs of playing roles in the genetic basis of interspecies evolutionary change of *Drosophila* ovariole number. These included the signalling genes *upd2* and *Ilp5* and extracellular matrix genes *vkg* and *Col4a1*, whose dN/dS predicted ovariole numbers among species. Together, we propose a model whereby a set of ovariole-involved gene proteins have an enhanced evolvability, including adaptive evolution, facilitating rapid shifts in ovariole number among *Drosophila* species.

Keywords: Ovariole number, *Drosophila*, genetic mechanism, phenotype, dN/dS, adaptive evolution, *tau*

Significance Statement: Ovaries in *Drosophila*, like in other insects, contain egg producing structures, known as ovarioles. The number of ovarioles per female varies among *Drosophila* species, but little is known about the genes and evolutionary dynamics that may shape interspecies changes in ovariole numbers. Here, used *a priori* experimental and transcriptome data from *D. melanogaster* to identify genes involved in ovariole formation and functions, and studied their molecular evolution among its closely related species within the *melanogaster* subgroup. Using a multi-layered analysis consisting of protein sequence divergence (dN/dS), adaptive evolution, expression breadth, and phylogenetic regressions, we identified 42 genes whose molecular evolution patterns were well linked to ovariole numbers divergence. Further, gene protein sequence divergence was often predictive of species ovariole numbers.

Introduction

Ovarian development is a process that is poised to play key roles in organismal evolutionary biology, as the female gonads form and house the oocytes and/or eggs that are central to fertility and reproductive success of a species, and thus affect their fitness (Miller, et al. 2014; Macagno, et al. 2015). In insects, the most well-studied model with respect to ovarian development and genetics is the fruit fly *Drosophila melanogaster* (Dansereau and Lasko 2008; Eliazer and Buszczak 2011; Li, et al. 2014; Slaidina, et al. 2020; Lebo and McCall 2021). The mature ovary in *D. melanogaster*, as in other species of insects, is comprised of tubular egg-producing structures known as ovarioles (King, et al. 1968; Dansereau and Lasko 2008; Lebo and McCall 2021), which are a central factor shaping organismal reproductive output (Montague , et al. 1981; Starmer, et al. 2003; Church, et al. 2021). The number of ovarioles contained in the ovaries is highly variable within the genus *Drosophila* (Kambysellis and Heed 1971; Hodin and Riddiford 2000; Starmer, et al. 2003; Markow, et al. 2009; Sarikaya, et al. 2019; Church, et al. 2021). As an example, within the *melanogaster* subgroup, *D. melanogaster* has typically about 19 ovarioles per ovary, while its closely related sister species *D. sechellia* has only about 8 to 9 ovarioles per ovary (Hodin and Riddiford 2000). A broad range of ovariole numbers has been observed across the family *Drosophilidae*, from one to more than 50 per ovary across the genus *Drosophila* (Sarikaya, et al. 2019; Church, et al. 2021). At present, however, we know little about the genetic basis of the evolution of ovariole number within insects (Hodin and Riddiford 2000; Markow, et al. 2009; Sarikaya, et al. 2019).

A central factor that may underlie the rapid interspecies transitions in ovariole numbers in *Drosophila* is the evolvability of ovariole-related protein-coding genes, that is, the propensity of the proteins encoded by these genes to diverge and/or undergo adaptive sequence changes (Wagner and Zhang 2011; Cutter and Bundus 2020). Functional amino acid changes in protein-coding DNA and associated selection pressures (measured as nonsynonymous to synonymous changes, or dN/dS (Yang 1997;

Bielawski and Yang 2005; Cutter and Bundus 2020)) can play a significant role in shaping interspecies divergence of developmental processes and other key phenotypes (Hoekstra and Coyne 2007). For instance, dN/dS of specific genes or sets of genes has been correlated with the divergence of sperm length in *Drosophila* (Chebbo, et al. 2021), sperm head size (Luke, et al. 2014) and testis size (Ramm, et al. 2008) in rodents, plumage color in toucans (Corso, et al. 2016), and brain mass in primates (Montgomery, et al. 2011), as well as other species traits (Swanson and Vacquier 2002; Hoekstra and Coyne 2007; Clark, et al. 2009; Cutter and Bundus 2020). Several lines of evidence indicate that ovariole number may also be a phenotype whose interspecies evolution in *Drosophila* is shaped by gene protein sequence changes and associated selection pressures (dN/dS, (Yang and Nielsen 2002; Bielawski and Yang 2005; Yang 2007)). Specifically, ovariole number is highly heritable and polygenic (Coyne, et al. 1991; Wayne and McIntyre 2002; Bergland, et al. 2008; Green and Extavour 2012; Sarikaya and Extavour 2015; Lobell, et al. 2017; Kumar, et al. 2020), and thus genetic mechanisms exist wherein changes in ovariole-related gene protein products could lead to interspecies differences in ovariole numbers. Further, in *Drosophila*, sexual (positive) selection pressures have been commonly observed and mating behaviors are variable among taxa (Kaneshiro and Boake 1987; Singh, et al. 2002; Singh and Singh 2014; Lupold, et al. 2016; Wigby, et al. 2020). These factors have been linked to accelerated interspecies protein sequence evolution in reproduction-related gene proteins and reproductive characteristics (Markow 2002; Swanson, et al. 2004; Jagadeeshan and Singh 2005; Haerty, et al. 2007; Kang, et al. 2016), that may potentially include ovariole numbers. Natural adaptive selection may also influence ovariole number evolution in *Drosophila*. For example, ovariole numbers and/or functions among species have been correlated with local environmental conditions and with oviposition and larval substrates in the *melanogaster* subgroup, as well as in the Hawaiian *Drosophila* (Kambysellis and Heed 1971; Kambysellis, et al. 1995; Sarikaya, et al. 2019). Finally, species-specific ovariole number may also be partly influenced by neutral protein sequence

77 changes via random genetic drift (Kimura 1989; Kambysellis, et al. 1995). For these reasons, we sought
78 to investigate whether evolutionary pressures on changes in proteins (dN/dS) involved in ovariole
79 formation and function, especially in those genes that exhibit signs of evolvability and adaptive evolution,
80 could underlie or even predict interspecies divergence in ovariole number, as is the case for certain other
81 fitness-related phenotypes in animals (Montgomery, et al. 2011; Wagner and Zhang 2011; Luke, et al.
82 2014; Corso, et al. 2016; Chebbo, et al. 2021).

83 The most crucial developmental period that determines ovariole number in *D. melanogaster* is the
84 larval stage (fig. 1) (King, et al. 1968; Godt and Laski 1995; Hodin and Riddiford 2000; Sarikaya, et al.
85 2012; Sarikaya and Extavour 2015; Slaidina, et al. 2020). Somatic gonad precursors specified during
86 embryogenesis give rise to many different somatic ovarian cell types in the larval stage, and the numbers
87 and behaviours of these somatic cells largely determine final ovariole number (Extavour and Akam 2003;
88 Clark, et al. 2007; Dansereau and Lasko 2008). Specifically, the number of terminal filaments (TFs; fig.
89 1A) , which are stacks of flattened intercalated terminal filament cells in the anterior ovary at the late third
90 larval instar stage (LL3), determines adult ovariole number (King, et al. 1968; Godt and Laski 1995;
91 Dansereau and Lasko 2008; Sarikaya, et al. 2012; Sarikaya and Extavour 2015). Each TF is the starting
92 point for formation of a single ovariole (Sahut-Barnola, et al. 1996; Sarikaya, et al. 2012), which contains
93 an anterior germarium housing germ line stem cells, and egg chambers that form the oocytes in an anterior
94 to posterior pattern of oocyte maturation (Sahut-Barnola, et al. 1996; Eliazer and Buszczak 2011;
95 Sarikaya, et al. 2012; Lebo and McCall 2021; Slaidina, et al. 2021). Single-celled RNA sequencing (sc-
96 RNA seq) data (Slaidina, Banisch et al. 2020) suggest that LL3 TFs have anterior (TFa) and posterior
97 (TFp) subgroups with distinct transcriptional profiles (fig. 1A). Another key somatic cell type are the
98 sheath cells, also located at the anterior of the LL3 ovary (fig. 1A), and sub-categorized based on sc-RNA
99 seq into anterior sheath cells (SHa) and migrating sheath cells (SHm). The latter cells migrate in an anterior

to posterior direction between the TFs, depositing basement membrane that partitions the remaining cells of the ovary (germ cells and posterior somatic cells) into the developing ovarioles (King, et al. 1968; King 1970; Slaidina, et al. 2020). Additional somatic cells in the LL3 ovary include intermingled cells, which are interspersed between the germ cells and are involved in their proliferation (Gilboa and Lehmann 2006), cap cells, which form the adult germ line stem cell niche (Song, et al. 2002), follicle stem cell precursors, which give rise to adult follicle stem cells (Slaidina, et al. 2020; Slaidina, et al. 2021), and swarm cells, whose precise functions largely remain to be ascertained (Slaidina, et al. 2020) (fig. 1A). In this regard, understanding the interspecies evolution of ovariole number in *Drosophila* requires consideration of the genes and proteins regulating cell behaviour in the larval ovary, and particularly the behaviours of the TF and SH cells, which are instrumental to determining ovariole numbers in *D. melanogaster*.

Until recently, research on the relationships between divergence in gene sequences and ovariole numbers in *Drosophila* was challenged by the lack of data on the identity of protein-coding genes expressed in somatic cells of the larval ovary that regulate ovariole number (Sarikaya, et al. 2012; Sarikaya and Extavour 2015). Recently available large-scale functional genetic and cell type-specific expression data from *D. melanogaster*, however, now provide a means to systematically identify genes linked to ovariole numbers, and in turn, assess their molecular evolution across species. A large-scale RNAi screen of 463 signalling genes from 14 conserved animal signalling pathways revealed that TF-mediated ovariole number determination is regulated by all conserved animal signalling pathways (Kumar, et al. 2020). Another study using bulk-RNA seq expression data from FACS-separated germ cells and somatic cells revealed additional genes differentially expressed throughout TF formation, suggesting their potential involvement in ovariole number regulation (Tarikere, et al. 2022). In addition to those studies, a recent sc-RNA seq study yielded unique transcriptional profiles for all of the known cell types in the *D. melanogaster* LL3 ovaries (fig. 1), providing a novel resource to identify and study the evolution of genes

transcribed in TF and SH cells, the two crucial cell types in determining ovariole number (Slaidina, et al. 2020).

Collectively these datasets provide valuable empirical data from which to *a priori* identify sets of genes that regulate ovariole numbers or functions in *Drosophila*, and in turn, to evaluate which of these genes exhibit elevated or otherwise unusual rates of interspecies protein sequence evolution, including adaptive evolution, suggesting them as candidates for driving interspecies divergence of ovariole numbers in *Drosophila*. For example, by assessing dN/dS, we may ask whether ovariole-related gene protein sequences typically have been under strict purifying selection, which could mean that phenotypes regulated by these genes are likely to show high pleiotropy and low evolvability, and to have minimal potential to diverge neutrally or adaptively (Fisher 1930; Otto 2004; Wagner and Zhang 2011; Cutter and Bundus 2020; Munds, et al. 2021). If, in contrast, some ovariole-related genes have been subjected to relaxed selection and/or have commonly experienced adaptive changes, we might expect high phenotypic evolvability and adaptability (Otto 2004; Larracuente, et al. 2008; Clark, et al. 2009; Mank and Ellegren 2009; Montgomery, et al. 2011; Luke, et al. 2014; Corso, et al. 2016; Chebbo, et al. 2021). In this regard, the study of the evolution of protein-coding genes (from dN/dS) that are pre-screened for likely roles in ovariole numbers and/or functions by studies like the ones described above (Kumar, et al. 2020; Slaidina, et al. 2020; Tarikere, et al. 2022) provides a novel pathway to advance our understanding of the genetic factors and evolutionary forces that shape rapid interspecies divergence in ovariole numbers.

In the present study, we rigorously assess the molecular evolutionary patterns of genes that regulate ovariole numbers and/or functions, that were identified *a priori* based on one or both of functional genetic evidence (Kumar, et al. 2020) or transcriptional activity (Slaidina, et al. 2020; Tarikere, et al. 2022). We focus on the molecular evolution of ovariole-related genes within five species of the *melanogaster* subgroup of *Drosophila*, that is a closely related species clade that includes *D. melanogaster*, diverged

from a common ancestor about 13 Mya (Tamura, et al. 2004), and exhibits substantial interspecies variation in ovariole numbers (Hodin and Riddiford 2000; Starmer, et al. 2003; Markow, et al. 2009). From our assessments, we identify 42 genes that are high confidence candidates for contributing to the genetic basis of interspecies divergence in ovariole numbers. We hypothesize that evolved changes in these genes are apt to underlie ovariole number divergence among taxa given that they exhibit an ovariole-related function (Kumar, et al. 2020; Slaidina, et al. 2020; Tarikere, et al. 2022), have a propensity to diverge in protein sequence, or high evolvability, show a high frequency of adaptive sequence evolution events in branches of the phylogeny, and are associated with low pleiotropy (Yanai, et al. 2005). Further, phylogenetic regressions show gene dN/dS has predictive associations to ovariole numbers. Collectively, our findings provide a genetic framework to explain the rapid interspecies divergence of ovariole numbers in *Drosophila*, which we propose is largely mediated by selection pressures shaping the evolution of functional protein sequences, and thus ovariole numbers.

Results

The Clade Under Study, the *melanogaster* subgroup

For our study, we focused on a multi-layered analysis of the molecular evolution of ovariole-related genes across five species from the *melanogaste*r subgroup of *Drosophila*: *D. simulans* (Dsim), *D. sechellia* (Dsec), *D. melanogaster* (Dmel), *D. yakuba* (Dyak), and *D. erecta* (Dere) (fig. 2; *D. ananassae* of the *melanogaster* group was used as an outgroup for phylogeny construction, see "*Drosophila* Phylogeny" section; the abbreviated names were used in tables and figures). Using this closely-related species clade, we hypothesize that if genes with demonstrated roles in regulating ovariole numbers or formation are involved in the interspecies divergence of ovariole numbers, then they will exhibit relatively rapid evolution (dN/dS) as compared to the genome, as well as interspecies variation in dN/dS, signs of

positive selection, and low pleiotropy (as inferred by high *tau* across tissues, table S1). We further hypothesize that, if evolutionary variation in these genes contributes to the genetic basis of evolved shifts in ovariole number, that dN/dS values for these genes may predict species ovariole numbers.

The five-species clade of *melanogaster* subgroup had the following advantages for our study: (1) all species within the clade are very closely related to *D. melanogaster* (Tamura, et al. 2004; Obbard, et al. 2012), the species for which experimental and transcriptome data on genes associated with ovariole numbers or functions are available (Kumar, et al. 2020; Slaidina, et al. 2020; Tarikere, et al. 2022), and thus we hypothesize are likely to share similarities in the genetic pathways affecting ovariole numbers, more so than we would be expect for distantly related species; (2) the clade exhibits substantial variation in ovariole numbers among species, typically about 39.2 (per female) for *D. melanogaster* and 17.0 for *D. sechellia* and intermediate values for *D. simulans* (33.9), *D. yakuba* (25.8) and *D. erecta* (27.0) (fig. 2; see values and variability (Hodin and Riddiford 2000; Starmer, et al. 2003; Markow, et al. 2009), and includes both some species with similar ovariole numbers and others that markedly differ; (3) the phylogeny is highly resolved (fig. 2 (Cutter 2008; Obbard, et al. 2012), unlike some other *Drosophila* clades and branches (Finet, et al. 2021)), and the five species are very closely related to each other (Tamura, et al. 2004; Cutter 2008). We made this choice to minimize biological differences other than ovariole numbers among taxa, and to facilitate the detection of putative cause-effect relationships (here, dN/dS and ovariole number (Felsenstein 1985; Bromham, et al. 1996; Whittle and Johnston 2003; Thomas, et al. 2010; Symonds and Blomberg 2014). The close relatedness of species is more conducive to accurate alignments, and retains a larger set of orthologous genes, including rapidly evolving genes, for study, than when studying more divergent species, which often skews toward the identification of fewer and more slowly evolving orthologous gene sets (*cf.* (Stanley and Kulathinal 2016; Bubnell, et al. 2022)), and may exclude some rapidly evolving genes of interest.; (4) each species has a whole genome sequence available

(Gramates, et al. 2022) and; (5) the dN and dS values among the species in this subgroup have substantially diverged, yet are also unsaturated in the frequency of substitutions, and thus are within the ideal range for dN/dS analysis (Castillo-Davis, et al. 2004; Larracuente, et al. 2008; Treangen and Rocha 2011) (for example, from M0 dN/dS values (that is, the single clade-wide measure of dN/dS, (Stanley and Kulathinal 2016)), we found that the 95th percentile for M0 dN=0.235 and M0 dS=0.791 for the 9,232 genes that had orthologs in all five species and M0 values). In sum, this closely related taxonomic group has multiple benefits for the study of the evolution of ovariole-related genes.

Identification of Rapidly-Evolving Ovariole-Related Genes for Follow-up Study

To identify genes associated with ovariole numbers or functions for study, we focused on three recently available datasets from *D. melanogaster*. The first gene set we designate as the SIGNALC dataset, defined here as the signalling and connector genes (connectors identified by protein interaction networks) that were identified as affecting ovariole or egg numbers in a *hpo[RNAi]* and/or a *hpo[+]* background (Kumar, et al. 2020). Among 463 signalling genes and additional connector genes studied, the authors reported 67 genes that affected ovariole number in a *hpo[RNAi]* background (named therein *hpo[RNAi]* Ovariole Number), 59 and 49 genes that affected egg laying in a *hpo[RNAi]* background (*hpo[RNAi]* Egg Laying) and a wild type (*wt*) background (Egg Laying *[wt]*) respectively, and 17 connector genes that altered ovariole or egg laying phenotypes (and passed screening of z>1; note that genes may belong to more than one category) (Kumar, et al. 2020). The second is the BULKSG dataset, based on bulk-RNA seq data obtained from pooled larval ovarian somatic cells or germ cells from the early (72 hours after egg laying = 72h AEL), mid (96h AEL) and late (120h AEL) TF developmental stages (Tarikere, et al. 2022) and identified differentially expressed genes (P-values were from DeSeq2 (Love, et al. 2014)). The third is the SINGLEC dataset (Slaidina, et al. 2020), a sc-RNA seq dataset that provided expression data for

each of the cell types of the *D. melanogaster* LL3 larval ovary (fig. 1) (Slaidina, et al. 2020). The SINGLEC study assessed average standardized expression to identify differentially expressed genes among cell types (P-values from Seurat v.2; some genes were upregulated in more than one cell type using the criteria therein (Slaidina, et al. 2020)).

The SIGNALC, BULKSG and SINGLEC gene sets were screened for further study using their clade-wide M0 dN/dS values (Yang 2007), that reflects the rate of protein divergence and the potential types of selective pressures that may have affected a gene (Yang 1997, 2007). Values of dN/dS <1 suggest a history of purifying selection on protein sequences, =1 infer neutral evolution, and >1 suggest a history of positive selection (Yang 1997, 2007); however, even when dN/dS <1 across an entire gene (Yang 2007), elevated dN/dS values in one gene relative to another suggest an enhanced degree of positive selection and/or neutral evolution (Yang 1998, 2007; Buschiazzo, et al. 2012; Ho and Smith 2016; Mitterboeck, et al. 2017; Whittle, et al. 2021). We identified those ovariole-related genes with an M0 dN/dS value at least 1.5-fold (SIGNALC; lower cut-off due to conserved nature of signalling genes, see Materials and Methods) or 2-fold (BULKSG and SINGLEC) higher than the genome-wide medians, and we then conducted a thorough follow-up analysis that included the M1 free-ratio species branch dN/dS (e.g., (Dorus, et al. 2004; Nadeau, et al. 2007; Clark, et al. 2009; Wlasiuk and Nachman 2010; Mensch, et al. 2013; Borges, et al. 2019; Kong, et al. 2019; LaBella, et al. 2021)), branch-site tests of positive selection (Zhang, et al. 2005; Yang 2007), *tau* (Yanai, et al. 2005) and phylogenetic regressions (R-Core-Team 2022) (see Materials and Methods).

Some Signalling Pathway Genes that Regulate Ovariole Number Have Evolved Rapidly

We report that for the ovariole-related SIGNALC gene set (Kumar, et al. 2020), that included signalling genes that affected ovariole number and/or egg laying, many genes exhibited very low M0

238 dN/dS (MWU-tests had P<0.05 versus the genome-wide values; fig. 3A). This suggests a history of strong
239 purifying selection on these highly conserved signalling genes, which may be partly due to their high
240 pleiotropy, given that all of these signalling pathways play multiple roles in development and homeostasis
241 (Kumar, et al. 2020). Consistent with this hypothesis, the *tau* values for these genes were statistically
242 significantly lower than the genome-wide values (fig. 3B; MWU-tests P<0.05), suggesting that broad
243 expression breadth may have acted to slow molecular evolution (Otto 2004; Kim, et al. 2007; Cui, et al.
244 2009; Mank and Ellegren 2009; Meisel 2011; Assis, et al. 2012; Masalia, et al. 2017; Whittle, et al. 2021).
245 Importantly however, our main goal herein was to identify whether any ovariole-related SIGNALC
246 genes evolved unusually rapidly, and showed signs of evolvability that could underlie interspecies
247 ovariole number divergence. As shown in table S2, we indeed identified 27 SIGNALC genes that had
248 elevated M0 dN/dS in at least one of the studied *Drosophila* taxon groups (≥1.5 fold higher than the
249 genome-wide median; table 1, table S2, see also Supplementary Text File S1 Results, and table 1 Notes
250 for *Paris*). The signalling pathways and example functions of each of these genes are provided in table
251 S3: we found they were preferentially involved in developmental and cytoskeletal roles. Thus, it is
252 apparent that while most of the ovariole number-related signaling genes evolved under strong purifying
253 selection (fig.3A), a subset of them exhibited a high rate of amino acid sequence changes, well above the
254 genome-wide median, in the *melanogaster* subgroup of *Drosophila*. This pattern shares similarities to the
255 previous finding that while most *D. melanogaster* developmental genes expressed at the phylotypic stage
256 of embryogenesis evolved under strong purifying selection (low dN/dS), a subset of genes expressed at
257 this stage exhibited a history of positive selection (Mensch, et al. 2013).
258
259 **Rapid and adaptive evolution of specific signalling genes coincides with ovariole number evolution**

To examine potential lineage-specific patterns of molecular evolution and pleiotropy of the 27 rapidly evolving ovariole-related genes, we assessed dN/dS per species branch (table 1), branch-site positive selection (table 1), and *tau* (table S3). We found that these 27 genes showed marked differences in dN/dS values per gene among the five-species terminal branches in the *melanogaster* subgroup (the distribution of dN/dS for all genome-wide genes per species branch are shown in box plots in fig. S1). In addition, we observed branch-site positive selection in at least one species branch for 19 of the 27 genes (table 1), which is consistent with potential high adaptability of these genes. Of particular note is the *D. sechellia* branch, as this species evolved a very low ovariole number (17 ovarioles per female, fig. 2), only half that of its most closely related sister species *D. simulans* (33.9 ovarioles per female, fig. 2), since diverging from their recent common ancestor. Among the five species terminal branches, the *D. sechellia* terminal branch had the highest dN/dS values for nine genes (table 1), namely *Zyx*, *elB*, *CG5504*, *CG3630*, *upd2*, *RpS6*, *Pdk1*,*Pyr* and *tefu*, with values ranging from 0.191 to >1. Further, five of these genes exhibited branch-site positive selection on amino acids in the *D. sechellia* branch (*elB*, *CG5504*, *unp2*, *RpS6*, *Pdk1*, branch-site P <0.05 for all genes (Zhang, et al. 2005; Yang 2007)), explicitly showing a propensity for adaptive evolution in this species branch. In total, six of the 27 genes 22.2%) exhibited branch-site positive selection in the *D. sechellia* terminal branch. This was nearly double the genome-wide frequency for this species, which was 12.0% of 9,232 genes (one tailed Chi-square P=0.05). Thus the *D. sechellia* lineage, with the lowest ovariole numbers (fig. 2), has a dynamic molecular evolutionary history of ovariole number-regulating genes, consisting of rapid gene-wide evolution (dN/dS), combined with a pervasiveness of positive selection events on such genes in that species branch.

In *D. sechellia*'s sister species *D. simulans* (fig. 2), eight genes had the highest dN/dS values in the *D. simulans* terminal branch (table 1), five of which also exhibited statistically significant branch-site positive selection (*Su(var)2-10*, *CkIIbeta*, *Gug*, *aPKC*, *CtBP*, P<0.05, table 1). In total, six of the studied

27 SIGNALC genes (22.2%) presented branch-site positive selection in the *D. simulans* branch, which was more than four-fold higher than the genome-wide frequency for the species (5.4%, Chi-square P<0.05). In turn, four of 27 genes had the highest dN/dS in the *D. melanogaster* branch, and four genes had branch-site positive selection in *D. melanogaster* (14.8%), which was more than triple its genome-wide frequency (4.1%; Chi-square P<0.05). *D. yakuba* and *D. erecta* had the highest dN/dS for three and two genes respectively, and had branch-site positive selection in three and four genes (table 1). In sum, for the *melanogaster* subgroup, all five species terminal branches showed signs of having the highest dN/dS values for at least two (*D. erecta*) and up to nine (*D. sechellia*) genes, exhibited signals of branch-site positive selection, and had particularly high rates of protein sequence divergence.

The patterns in table 1 support the hypothesis that protein sequence changes, including adaptive changes, in these ovariole-related genes may underlie the genetic basis for the marked divergence in interspecies ovariole numbers (fig. 2). For many of these genes, their known molecular and genetic mechanisms of action in tissue morphogenesis make them prime candidates for future analyses of how their diverged functions between species may have contributed to species-specific ovariole number evolution. For example, *Zyx* (*Zyxin*) is an actin cytoskeleton regulator and a signal transducer in the Hippo pathway, and mis-regulation of either actin cytoskeleton function (Li, et al. 2003) or Hippo signaling function (Sarikaya and Extavour 2015; Kumar, et al. 2020) during ovariole morphogenesis can alter ovariole number. We provide further discussion of some of these ovariole-related signalling genes in table 1 within the Supplementary Text File S1.

Multiple Genes Highly Upregulated in Larval Ovary Somatic Cells Have Evolved Rapidly

We identified genes whose high differential expression in the *D. melanogaster* larval ovary suggested a role in ovariole number regulation using the BULKSG RNA-seq datasets using pooled larval

ovarian somatic versus pooled germ cells from different stages of TF formation (Tarikere, et al. 2022). First, we asked whether the 27 rapidly evolving ovariole-related SIGNALC genes in table 1 exhibited statistically significant differential expression between somatic and germ cells during TF formation (therein $P<0.01$, (Love, et al. 2014; Tarikere, et al. 2022)). Remarkably, as shown in table S4, we report that 25 of the 27 rapidly evolving SIGNALC ovariole-related genes showed up- or downregulation in the soma (versus germ cells; each cell type pooled across stages), or among the three different TF formation stages. Thus, this affirms that the SIGNALC genes in table 1 that experimentally affected ovariole numbers or functions using RNAi (Kumar, et al. 2020), and that showed signals of enhanced evolvability herein (table 1, table S3), also exhibited differential expression in the larval somatic ovary cells, based on an independent approach of bulk RNA-seq (Tarikere, et al. 2022). These two lines of evidence suggest that these genes are apt to have contributed towards the genetic basis of evolved ovariole number divergence.

Rapidly Evolving Genes are Highly Transcribed in the Larval Ovary Somatic Cells

We aimed to further identify any rapidly evolving genes that were highly differentially expressed in the larval ovarian soma during TF formation, and thus potentially involved in the evolution of ovariole number, using the BULKSG datasets. For this, we identified genes that were upregulated in the soma versus the germ cells, ranked them by $\log_2$fold upregulation, and in that subset, screened for genes that were rapidly evolving in the *melanogaster* subgroup as compared to the genome-wide values (see Methods, M0 $dN/dS>0.20$). The top ten genes matching these criteria are shown in table 2, with the highest $\log_2$fold values ranging from 5.1 to 10.0, and includes the branch dN/dS, branch-site positive selection tests for each species of the *melanogaster* subgroup and *tau* values (see Supplementary Text File S1 for analysis of genes highly upregulated in germ cells, table S5).

Remarkably, eight of the ten most highly upregulated and rapidly evolving somatic genes had extremely elevated *tau* values >0.90, and six had values above 0.94, indicating very narrow expression breadth (as compared to genome-wide values in fig. S2). This low pleiotropy may facilitate their rapid evolution, via neutral evolution, and/or by adaptive sequence evolution (Otto 2004; Larracuente, et al. 2008; Mank and Ellegren 2009). For the *D. sechellia* branch, five of the ten genes had the highest dN/dS in this species terminal branch, including *Ilp5 (Insulin-like peptide 5*, dN/dS=0.5843, discussed in Supplementary Text File S1) and four unnamed genes (CG identifiers only, *CG32581*, *CG31157*, *CG10232*, *CG30281*). Two of these, *CG31157* and *CG10232*, exhibited gene-wide positive selection with dN/dS values larger than 1, and the latter gene also had dN/dS >1 in *D. simulans* (table 2). Further, *CG31904* exhibited branch-site positive selection in *D. sechellia* (table 2). These patterns are consistent with a prevalent history of rapid protein evolution coupled with the ovariole number decline within the *D. sechellia* branch, as also observed for multiple SIGNALC genes (table 1). Further, three of the ten genes also showed branch-site positive selection in *D. melanogaster*, and one displayed this pattern in *D. erecta* (table 2), suggesting that many of these genes experienced a history of adaptive evolution across multiple lineages of the phylogeny.

Terminal Filament Cells and Sheath Cells Express Rapidly Evolving Genes

The SINGLEC dataset was based on sc-RNA seq data generated from the late third instar *D. melanogaster* ovary (Slaidina, et al. 2020) and includes expression data for all the cell types shown in fig. 1 (the germ cells (GC) and eight somatic cell types, namely the cap cells (CC), follicle stem cell precursors (FSCP), intermingled cells (IC), anterior sheath cells (SHa), migrating sheath cells (SHm), swarm cells (SW), anterior terminal filament cells (TFa), and posterior terminal filament cells (TFp)). Using hierarchical clustering of average standardized gene expression per gene, across all genes (fig. S3), we

found that the germ cells exhibited the most unique transcriptome of all studied cell types, and formed an outgroup to all somatic cells. Among the somatic cells, the two types of terminal filament (TF) cells, TFa and TFp, formed their own cluster, as did the two types of sheath (SH) cells, SHm and SHa; each of these clusters was separate from all other somatic cell types (fig. S3). The FSCP and SW cells had highly similar transcription profiles, as did the IC and CC cells. Thus, the TFs and SH cells had more distinctive transcriptomes than the other LL3 ovarian somatic cell types.

Rapidly evolving genes identified in both the BULKSG and SINGLEC datasets

To identify genes with roles in specific ovarian cell types that were putatively involved in interspecies ovariole number divergence, we first extracted those SINGLEC genes that were upregulated in one cell type relative to all others ($P<0.05$, analyzed in Seurat v. 2; genes could be upregulated in more than one cell type (Satija, et al. 2015; Slaidina, et al. 2020)), and that also had M0 dN/dS more than two fold above the genome-wide median (dN/dS>0.20) within the *melanogaster* subgroup. We then compared this SINGLEC gene set to the 30 most highly differentially expressed and rapidly evolving genes identified from the somatic larval ovary cells at three different stages of development for terminal filament formation (listed in table S6, extracted from BULKSG dataset) and determined whether any genes were upregulated in both datasets. We identified five genes that matched these criteria (table 3): *Drip*, *CG3713*, *MtnA*, *vkg*, and *Col4a1* (table 3). Among the nine somatic cell types, these genes were nearly exclusively upregulated in the TFs (TFa or TFp, or both) and/or the SHm cells. We note that *vkg* and *Col4a1* play roles in basement membrane formation (Yasothornsrikul, et al. 1997; Kiss, et al. 2019), and that SHm cells lay the membrane that separates the TFs for ovariole development (King 1970; Slaidina, et al. 2020). Given the crucial roles of these cell types in determining ovariole number (King, Aggarwal et al. 1968, Sarikaya and Extavour 2015), the rapid evolution of these five genes may partially underlie ovariole

number divergence between species (King, et al. 1968; Sarikaya and Extavour 2015) in the *melanogaster* subgroup (table 3).

In terms of molecular evolution per terminal species branch, the five genes in table 3 exhibited a striking propensity for adaptive evolution. Four the five genes showed a gene-wide level of positive selection (terminal branch dN/dS values >1) in at least one species branch (table 3). Moreover, *Drip*, *vkg* and *Col4a1* each exhibited branch-site positive selection in three different species branches ($P<0.05$), suggesting a profound history of adaptive changes across multiple lineages. In addition, McDonald and Kreitman (1991) tests also showed positive selection for *vkg* and *Col4a1* ($P<0.05$, table 3 Notes). All five genes exhibited *tau* values above 0.875 with *Drip* having a value of 0.979, suggesting especially high expression specificity (see Materials and Methods, fig. S2), which may facilitate the observed adaptive evolution of the protein sequences (Otto 2004; Mank and Ellegren 2009; Whittle, et al. 2021). In sum, these five genes were identified from two distinct expression datasets (Slaidina, et al. 2020; Tarikere, et al. 2022), were upregulated in two of the most crucial cell types for ovariole number determination namely TFs and SH cells (table S6, table 3), and exhibited rapid protein changes, positive selection, and narrow expression breadth (table 3). Thus, multiple lines of evidence point towards these genes as having a central role in the interspecies divergence of ovariole number.

Genes upregulated in TF and SH cells frequently display branch-site positive selection

We assessed the frequency of genes that exhibited branch-site positive selection ($P<0.05$) per species terminal branch for the rapidly evolving genes that were upregulated in each of the nine cell types in the SINGLEC dataset ($P<0.05$). The results for *D. simulans*, *D. sechellia* and *D. melanogaster* (a very closely related species group with substantial differences in ovariole numbers (fig. 2)), are shown in fig. 4, and for all five species in fig. S4. The genes with the highest percent branch-site positive selection were

those upregulated in the SH and TF cells (fig. 4; the TF and SH genes are listed in Table S7). Specifically, positive selection was most commonly observed for genes up-expressed in the SHm cells for the *D. sechellia* branch (45%), from the TFa (34.1%) and TFp (36.7%) cells in the *D. sechellia* branch, and for SHa cells in the *D. sechellia* (33.33%) and *D. simulans* (33.33%) branches (all values were statistically significantly higher than the genome-wide percentages of genes with branch-site positive selection per species, which were 5.4% for *D. simulans* and 12.0% for *D. sechellia*; Chi-square $P<0.05$, fig. 4). Thus, the most important somatic cell types for ovariole number determination (TF and SH cells) (King, et al. 1968; Godt and Laski 1995; Sarikaya, et al. 2012; Sarikaya and Extavour 2015; Slaidina, et al. 2020), are also those in which highly upregulated genes most commonly exhibited branch-site positive selection, particularly in *D. sechellia*.

The genes identified above as highly expressed in TF and SH cells, could also be highly expressed in additional cell types (Slaidina, et al. 2020). Indeed, on average we found that differentially expressed genes were upregulated in 1.9±0.02 cell types. Thus, for additional stringency we isolated the subset of rapidly evolving genes (with M0 $dN/dS>0.20$) that were upregulated in only one cell type. While most somatic cell types had very few genes matching this stringent criterion ($N\leq4$ per cell type), by pooling the two types of SH cells (SHa and/or SHm) and TF cells (TFa and/or TFb) we found 8 and 26 such genes in these cell types respectively (provided in table S7 Notes). We found that *D. simulans*, *D. sechellia* and *D. melanogaster* showed branch-site positive selection in 25.0%, 25.0% and 0% of these genes respectively for SH cells, and in 11.5%, 23.1%, and 7.7% of these genes for TF cells. These values were well above the genome-wide frequency for *D. sechellia* and *D. simulans* (although tests were conservative due to sample size, Chi-square P values for SH for *D. simulans* = 0.047 and TF for *D. sechellia* = 0.077 relative to the genome-wide values). In sum, interpreting the results in fig. 4 conservatively, we observe that upregulation of a gene in TF or SH cells is correlated with enhanced rates of positive selection in the *D.*

421 *sechellia* and/or *D. simulans* lineages, regardless of whether the genes were also upregulated in another
422 cell type (fig. 4; table S7).

423 While we focused on the three-species clade in fig. 4, the results for all five *melanogaster* subgroup
424 species are provided in fig. S4. Of particular note, those results showed that 45.5% of the genes that were
425 upregulated in the SHm cells also exhibited positive selection in the *D. yakuba* and in the *D. erecta*
426 terminal branches (similar to *D. sechellia* in fig. 4, table S7). This suggests a history of branch-site positive
427 selection for genes expressed in the SHm cells across outgroup branches of the phylogeny, potentially
428 partly contributing to the divergence in ovariole numbers or functions in the two outgroup species from
429 the three ingroup species (fig. 2).

430

431 *Functional predictions for upregulated TF and SH genes*

432 The studied molecular evolutionary parameters for all genes studied in fig. 4 that were upregulated
433 in SHa, SHm, TFa, and TFp are provided in table S7. Analysis of GO-predicted functions using DAVID
434 (Huang da, et al. 2009) showed that the genes expressed in SHa and SHm cells, such as *Jupiter* and *Timp*
435 (table S7), were preferentially involved in microtubule formation and basement membranes (Huang da, et
436 al. 2009), consistent with roles in TF formation (Slaidina, et al. 2020). The highly upregulated and rapidly
437 evolving TF genes in fig. 4 and table S7 were more than threefold more common than the SH cell genes,
438 and thus allowed us to perform functional clustering (Huang da, et al. 2009). As shown in table S8, the
439 TF genes were preferentially associated with extracellular matrix (20.5% and 23.3% of genes from TFa
440 and TFp respectively), basement membranes (6.8 and 10%), and 40% of genes from TFp were an integral
441 component of membranes.

442 Given that the TF and SH cells types in fig. 4 (and TF and SH cells are included in the larval ovary
443 somatic cells in table 2) have been experimentally shown to regulate the formation and number of

ovarioles in *D. melanogaster* larvae (King, et al. 1968; King 1970; Godt and Laski 1995; Dansereau and Lasko 2008; Sarikaya, et al. 2012; Sarikaya and Extavour 2015; Slaidina, et al. 2020), the genes that were both highly expressed in and/or required for these ovariole functions in these cells, and exhibited rapid sequence evolution and signals of adaptive evolution (fig. 4, table 2), have the potential to directly cause an interspecies shifts in ovariole numbers. In turn, it may also be the case that the protein sequence changes observed in some of these genes may be in response to evolved shifts in ovariole numbers (potentially mediated by other ovariole-involved genes identified herein), and thus that the adaptive changes that we report here reflect the physiological intracellular changes in TFs and SH cells needed to support ovariole number changes

Molecular Evolutionary Rates of Key Genes Predicts Ovariole Number

Finally, we conducted follow-up assessments of the main genes identified throughout our study that showed signs of high evolvability, positive selection, and involvement in *Drosophila* ovariole number divergence, to determine to what extent the molecular evolutionary characteristics of these genes were predictive of ovariole numbers in the context of *Drosophila* phylogeny. Specifically, for all genes identified from SIGNALC (N=27; table 1), from BULKSG (N=10; table 2) and from BULKSG and SINGLEC combined (N=5; table 3), we conducted a phylogenetic generalized least squares (PGLS) assessment of the relationship between ovariole number and the dN/dS values for the 41 of these 42 genes that were testable (*MtnA* was untestable due to infinity dN/dS (near zero dS, dN>0) in several branches; table 4; a summary of McDonald and Kreitman (1991) test values for all genes is shown in table S9). We found that 17 of the 41 testable genes (41.5%) showed a statistically significant relationship between ovariole number and dN/dS value (table 4; $P<0.05$, *CG3630* had $P<0.07$ and was noted in the list), indicating that dN/dS values of these genes can be a predictive factor for ovariole number per species. This further demonstrates the high effectiveness of utilizing protein sequence analysis to identify genes

467 putatively involved in the evolution of phenotypes, similar to suggestions for other diverse traits across
468 multiple taxa (Dorus, et al. 2004; Nadeau, et al. 2007; Ramm, et al. 2008; Wlasiuk and Nachman 2010;
469 Luke, et al. 2014; Corso, et al. 2016; Chebbo, et al. 2021).
470
471 **Supplementary Analysis of a Three-Species Clade of Hawaiian *Drosophila***
472 While we focused on the *melanogaster* subgroup for our core analyses, as a supplementary
473 assessment, we considered a three-species clade of Hawaiian *Drosophila* that matched our strict criteria
474 for study (very closely related species, genome-wide data, known (and variable) ovariole numbers). We
475 note, however, that these species are relatively distantly related to *D. melanogaster*, the species used to
476 identify ovariole-involved genes on the basis of function and/or expression (the SIGNALC, BULKSG and
477 SINGLESC datasets). Hawaiian *Drosophila* are paraphyletic to the *melanogaster* subgroup (Suvorov, et
478 al. 2022), and estimates of divergence time since the last common ancestor of extant species from the two
479 taxon groups exceed 60 Mya (Tamura, et al. 2004; Goldman-Huertas, et al. 2015). We chose the species
480 *D. sproati* (mean 65.6 ovarioles), *D. murphyi* (mean 41.6 ovarioles) and *D. grimshawi* (mean 47.8
481 ovarioles) for study, with a phylogeny of: ((*D. sproati*, *D. murphyi*), *D. grimshawi*) shown in fig. S5 (Kim,
482 et al. 2021; Suvorov, et al. 2022) (ovariole numbers from (Starmer, et al. 2003; Sarikaya, et al. 2019)).
483 For dN/dS analysis, we focused on the ovariole-involved SIGNALC genes identified in table 1, as these
484 signalling proteins are functionally confirmed to regulate ovariole number (Srivastava, et al. 2010; Kumar,
485 et al. 2020). Thus, among the studied gene sets (SIGNALC, BULKSG, SINGLEC), we considered them
486 the most appropriate for dN/dS analysis in a divergent group. We found that 21 of the 27 rapidly evolving
487 ovariole-related genes in table 1, which were identified from study of the *melanogaster* subgroup, had a
488 high confidence three-species orthologous gene set in the Hawaiian *Drosophila* clade (table S10). Our
489 evaluation of branch-dN/dS values revealed that ten of the 21 genes evolved especially rapidly, with

dN/dS>0.33 in at least one species terminal branch in the Hawaiian clade, which was more than two-fold higher than the genome-wide dN/dS values for the species under study (13 of 21 genes evolved rapidly using a criterion of 1.5 fold higher than the genome-wide medians; genome-wide dN/dS median values = 0.152, 0.164, and 0.160 for *D. murphyi*, *D. sproati* and *D. grimshawi* respectively; table S10; Supplementary Text File S1). Moreover, *D. sproati*, the ingroup species with highest ovariole number per female of all three Hawaiian species (fig. S5), had eight of the ten genes with dN/dS>0.33 (table S10). The ten most rapidly evolving genes included *upd2*, *CG2199*, *vn*, *elB*, *bun*, *CG3630*, *aPKC*, *H*, *Su(var)205* and *E(spl)m2-BFM*, six of which also exhibited branch-site positive selection in at least one branch. For *upd2*, we observed branch (dN/dS>1) and branch-site positive selection (P<0.05) for all three species branches (table S10), suggesting it may have a putative role in ovariole number divergence in all three species. Nonetheless, it is notable that in Table S10, eight of the 21 genes had branch-dN/dS below the aforementioned thresholds (were not 2-fold or 1.5- fold higher than the genome median) in all three Hawaiian species branches (table S10). This suggests that while these genes may be involved in ovariole functions in those taxa (as they are in *D. melanogaster* (Kumar, et al. 2020)), their protein sequence divergence may be less apt to shape interspecies shifts in ovariole numbers in these Hawaiian *Drosophila* species (table S10). Together, the data suggest that a substantial number of the rapidly evolving ovariole-involved genes in table 1, also evolved very rapidly in the Hawaiian clade, and thus may have possibly contributed to its interspecies divergence in ovariole numbers.

We also examined the Hawaiian *Drosophila* species orthologs of some of the rapidly and adaptively evolving genes in the *melanogaster* subgroup, which we identified from the SINGLEC transcription dataset (Slaidina, et al. 2020) shown in fig. 4 (and fig. S4, N values per cell type shown therein; the TF and SH cell genes are in table S7 that include certain genes from BULKSG in table 3). We hypothesized that for these genes, identified as candidate ovariole number regulators based on *D.*

513 *melanogaster* expression profiles alone, it might be harder to confidently assume conservation of function
514 in ovariole number regulation in a clade as distantly related as the Hawaiian *Drosophila* (Ranz, et al. 2003;
515 Whittle and Extavour 2019). We therefore adopted a prudent approach, based on evaluation of the rate of
516 high confidence ortholog detection in the Hawaiian group (see Methods and Results in Supplementary
517 Text File S1). As shown in fig. S6, we found that genes in the TF and SH cells (fig. 4, fig. S4), had the
518 fewest high confidence Hawaiian orthologous gene sets, as compared to genes highly expressed in the
519 other ovarian cell types (orthologs were defined as having an ortholog found in all three Hawaiian species,
520 and between *D. melanogaster-D. grimshawi* for gene identification). Specifically, genes upregulated in
521 the SHa cells and those in the TFp cells (fig. S6), each had 66.6% of genes with an orthologous Hawaiian
522 three-species orthologous gene set. In contrast, genes upregulated in CC had 85.7%, and FSCP and IC
523 each had 82.4% (fig.S6). We speculate that genes expressed in the TF and SH cells may have evolved at
524 a relatively higher rate (fig. 4, fig. S4, table S7) than those expressed in other ovarian cell types, making
525 orthologs more frequently unrecognizable between *D. melanogaster* and the Hawaiian clade and/or among
526 the three species in the Hawaiian clade (Tautz and Domazet-Loso 2011; Tautz, et al. 2013)(discussed
527 further in Results within in Supplementary Text File S1). This rapid evolution could potentially be due to
528 adaptive sequence changes associated with ovariole number divergence in the genus (fig. 4). It is also
529 possible that there has been a greater propensity of genes directly involved in ovariole formation (TF and
530 SH cells) to undergo gains and/or losses over evolutionary time (Tautz and Domazet-Loso 2011; Tautz,
531 et al. 2013), than genes involved in regulating the other ovarian cell types. While our central focus herein
532 was on the interspecies divergence of ovariole number and protein sequences of orthologous genes within
533 the very closely related *Drosophila melanogaster* subgroup (table 1, table 2, table 3, table 4, fig. 4), these
534 supplementary analyses in a Hawaiian clade provide insights into the dynamics potentially contributing
535 to ovariole number divergence over extended time scales.”.

Discussion

While insects exhibit a diverse number of ovarioles, including across two orders of magnitude in the genus *Drosophila* alone (Hodin and Riddiford 2000; Starmer, et al. 2003; Markow, et al. 2009; Sarikaya, et al. 2019; Church, et al. 2021), little has been known about the genetic basis of rapid interspecies divergence of this fundamental female reproductive trait. Here, we directly tackled this issue by comprehensively determining *a priori* genes with experimental and/or transcriptional evidence for roles in determining ovariole numbers or functions in *D. melanogaster* (Kumar, et al. 2020; Slaidina, et al. 2020; Tarikere, et al. 2022), and then assessing their molecular evolutionary characteristics within very closely related species in the *melanogaster* subgroup. The results revealed a highly evolvable set of ovariole-related genes that exhibited high gene-wide dN/dS and/or branch-site positive selection in patterns consistent with a role in the evolution of ovariole number divergence (table 1, table 2, table 3, table 4, table S7). Moreover, PGLS analyses supported a predictive relationship between ovariole number per species and dN/dS for many of the identified rapidly evolving ovariole-related genes (table 4). From these collective results, we propose that the rapid interspecies ovariole number divergence in *Drosophila* (fig. 2) has been facilitated by a group of highly evolvable genes with ovariole-related functions (42 identified and of focus herein, (Kumar, et al. 2020; Slaidina, et al. 2020; Tarikere, et al. 2022)) that exhibit a propensity for rapid evolution (gene-wide dN/dS) and adaptive protein sequence changes (table 1, table 2, table 3, table S7, fig. 4, fig. S4). This hypothesis is further supported by the fact that all of the ovariole-related genes revealed herein have been explicitly demonstrated to regulate ovariole number (Kumar, Blondel et al. 2020), and/or are highly and/or exclusively expressed in somatic ovarian cells whose behaviour determines ovariole number (King, et al. 1968; King 1970; Sarikaya, et al. 2012; Sarikaya, et al. 2019; Slaidina, et al. 2020; Tarikere, et al. 2022).

559

560 **Evolvability of Ovariole-Related Genes and *tau***

561 The evolvability, defined here as the propensity of traits or gene sequences to diverge (Wagner
562 and Zhang 2011; Cutter and Bundus 2020), including adaptive evolution, for the ovariole-related genes
563 identified herein for the *melanogaster* subgroup (table 1, table 2, table 3; and for the rapidly evolving
564 ovariole genes for Hawaiian *Drosophila*, table S10), may potentially reflect fitness advantages of the fixed
565 ovariole-related mutations, and/or may have been influenced by relaxed purifying selection. Previous
566 studies have found that genes with high values of *tau* (Yanai, et al. 2005), which suggests low pleiotropy
567 (Mank and Ellegren 2009; Meisel 2011; Dean and Mank 2016), may exhibit relaxed purifying selection,
568 thereby allowing both elevated neutral protein sequence changes (and thus elevated dN/dS), and greater
569 potential for adaptive evolution (Otto 2004; Larracuente, et al. 2008; Mank, et al. 2008; Mank and Ellegren
570 2009; Meisel 2011; Whittle, et al. 2021). Consistent with this pattern, we found that many of the rapidly
571 evolving ovariole-associated genes, including those with explicit evidence of adaptive evolution from
572 gene-wide dN/dS values larger than 1 or from branch-site positive selection tests ($P<0.05$), also exhibited
573 relatively high *tau* (for example, those with values >0.90, table 1, table 1, table 3). Thus, low pleiotropy
574 may have partly contributed to high evolvability, and enhanced adaptive potential. These events of
575 positive selection in the ovariole-related genes (table 1, table 2, table 3, fig. 4), may have arisen by natural
576 selection for adaption to changes in environment or oviposition substrates (Jagadeeshan and Singh 2007),
577 and/or may have often been driven by the widely-reported and dynamic sexual behaviors of *Drosophila*,
578 as described below.

579

580 **Putative Roles of Sexual Selection on Ovariole Number Evolution**

Sexual selection may contribute to the adaptive evolution of reproductive characteristics and genes in animals (Swanson and Vacquier 2002; Clark, et al. 2009), including in *Drosophila* (Civetta and Singh 1998; Swanson, et al. 2004; Proschel, et al. 2006). Thus, one possibility is that this phenomenon may shape the evolution of ovariole-related genes observed herein (table 1, table 2, table 3, fig. 4). Different species of *Drosophila* exhibit wide variation in their reproductive behaviors (Markow and O'Grady 2005), and examples of sexual selection in the genus include intrasexual selection from sperm competition (Singh, et al. 2002; Singh and Singh 2014) and male-male (Singh and Singh 2014) and female-female competition (Bath, et al. 2018). In addition, there is evidence of intersexual selection including female- and male-mate choice (Friberg and Arnqvist 2003; LeVasseur-Viens, et al. 2015). In the latter case, if males favor larger females, a choice that may correlate with female fecundity in species where body size correlates positively with ovariole number (Bonduriansky 2001; Byrne and Rice 2006; Sinclair, et al. 2021), then this could result in positive selection on amino acid changes favoring increased ovariole numbers. Moreover, *Drosophila* exhibits sexual antagonism, which could also potentially shape female (and male) reproductive characteristics and their underlying genes (Arnqvist 1995; Rice 1996; Swanson, et al. 2004; Innocenti and Morrow 2010). For example, in *D. melanogaster*, some male reproductive traits and behaviors (e.g. seminal fluid toxicity, aggressive male re-mating behaviors) may be harmful to female reproduction and/or survival (Civetta and Clark 2000; Chapman, et al. 2001; Sirot, et al. 2014). Some studies have suggested that this could prompt female adaptive responses, and give rise to adaptive changes in the *D. melanogaster* ovaries or eggs and in the protein sequences of genes expressed in the ovaries (Civetta and Clark 2000; Jagadeeshan and Singh 2005; Sirot, et al. 2014). If this phenomenon also occurs across other members of the *melanogaster* subgroup, it may contribute to positive selection on ovariole numbers and thus on ovariole genes observed here. Significantly, sexual selection may affect reproductive phenotypes and genes (Swanson and Vacquier 2002; Proschel, et al. 2006) in a polygenic manner (Lande

1981; Coyne and Charlesworth 1997; Singh, et al. 2001; Markow and O'Grady 2005; Singh and Singh 2014), which is relevant to ovariole number evolution as this is a highly polygenic trait (Coyne, et al. 1991; Wayne and McIntyre 2002; Bergland, et al. 2008; Green and Extavour 2012; Sarikaya and Extavour 2015; Lobell, et al. 2017; Kumar, et al. 2020).

Neutral Evolution and Ovariole Number

While we propose that our results could suggest an important role for adaptive evolution in ovariole-related genes in the interspecies divergence of ovariole numbers, it is worthwhile to consider the potential, and possibly complementary, roles of neutral evolution. Relaxed purifying selection in itself may lead to accelerated evolution and protein sequence changes (Kimura 1983; Mank and Ellegren 2009; Gossmann, et al. 2012), and to an elevated gene-wide dN/dS in a particular branch. Thus, it may be possible that some selectively neutral amino acids in ovariole-related genes were fixed via random genetic drift and affected ovariole numbers, possibly facilitated by low pleiotropy (high *tau*) (Fisher 1930; Meisel 2011; Assis, et al. 2012; Whittle, et al. 2021). Crucially however, such neutral (non-directional) changes would not be expected to yield the striking patterns we found for gene-wide dN/dS per species in ovariole-related genes and ovariole numbers (across species table 1, table 2, table 3), nor to give rise to the observed predictive relationships between dN/dS and ovariole numbers using PGLS (table 4). Moreover, our explicit evidence of adaptive evolution across many ovariole-related genes, by gene-wide dN/dS values larger than 1, branch-site positive selection analysis and McDonald and Kreitman (1991) tests ($P<0.05$, table 1, table 2, table 3, table S9, fig. 4, fig. S4), is unlikely to be explained by neutral evolution alone. Thus, the present data suggest that neutral evolution has not been the only or main driving factor shaping amino acid changes in ovariole-related genes in the *melanogaster* group, which we propose instead are best explained by a history of adaptive evolution.

627 Another factor in addition to narrow expression breadth (a factor that affects individual genes) that
628 could in theory lead to relaxed purifying selection on nonsynonymous mutations in ovariole genes is small
629 population size, which may affect entire genomes (Kimura 1962; Strasburg, et al. 2011; Gossmann, et al.
630 2012). As an example, under this scenario, relaxed selection may be expected to be more common in the
631 *D. sechellia* lineage (fig. 2), in which the extant species has been suggested to have a smaller population
632 size than other closely related *Drosophila* species such as *D. simulans* (Legrand, et al. 2009). Thus, we do
633 not exclude the possibility that certain gene-wide nonsynonymous changes (dN in dN/dS) in that species
634 branch may have contributed to its altered ovariole numbers, under an assumption that some slightly
635 deleterious mutations may behave as selectively neutral mutations (as effective population size (N_e) and
636 selection coefficient (s) may yield, $N_es<1$) and be fixed by random genetic drift (Strasburg, et al. 2011;
637 Gossmann, et al. 2012). However, as outlined above, the analyses showing affirmative branch-site positive
638 selection tests here and the findings of gene-wide dN/dS values larger than 1 each control for neutral
639 evolution (Zhang, et al. 2005; Yang 2007), and showed that positive selection was common in the *D.*
640 *sechellia* branch (table 1, table 2, table 4,table S7, and fig. 4). Furthermore, the results revealed a high
641 frequency of positive selection in genes upregulated in the TFs and SH cells in *D. sechellia* (fig. 4, table
642 S7), a pattern not explainable by neutral evolution (relaxed selection) due to population size. Collectively,
643 the evidence suggests that relaxed purifying selection, while potentially accelerating divergence rates of
644 some ovariole-related genes studied here (Duret and Mouchiroud 2000; Mank and Ellegren 2009; Meisel
645 2011; Whittle, et al. 2021), may have its most significant role in the evolvability of ovariole-related genes
646 (e.g., under high *tau*), enhancing the potential for adaptive evolution of protein sequences (Otto 2004;
647 Larracuente, et al. 2008; Mank and Ellegren 2009; Whittle, et al. 2021), and in that manner potentially
648 affecting interspecies ovariole number evolution.
649

Evolution of Multiple Developmental Processes via Rapid Divergence of Genes that Regulate Ovariole Number

Generating the right number of ovarioles for a given species relies on multiple developmental processes that begin during embryogenesis and are not completed until puparium formation. These include establishment of a specific number of somatic gonad precursor cells in the embryonic primordial gonad, proliferation at a specific rate and to a specific degree during larval stages, morphogenetic movements including intercalation and migration to establish terminal filaments, and extracellular matrix deposition to separate ovarioles from each other within the gonad (King 1970). Any of these developmental processes could in principle be the target of evolutionary change in interspecies ovariole number divergence. Indeed, we previously showed that evolution of different developmental mechanisms underlies convergent evolution of similar ovariole numbers between or within species (Green and Extavour 2012). Accordingly, we would expect that the genes underlying these evolutionary changes might play roles in multiple different developmental processes, and this prediction is supported by our findings herein. The genes that we have identified here as not only rapidly evolving in the *melanogaster* subgroup (table1, table 2, table 3), but also with molecular evolutionary rates that are highly predictive of lineage-specific ovariole numbers (table 4), have known functional roles in cell-cell signalling, cell proliferation, cell shape change, cell migration, and extracellular matrix composition and function (table 3, table S8; see gene descriptions in Supplementary Text File S1), including in but not limited to ovariole formation in *D. melanogaster*. Further, the distinct patterns of branch-site positive selection in different lineages, suggest that ovariole number evolution involved modification of distinct developmental processes in different lineages. For example, the rapid evolution of *Zyx*, *vkg*, *col4a1*, *Ilp5*, and *CG3630* in the lineage leading to *D. sechellia* (table 1,table 2, table 3) suggests that alteration of the TF morphogenesis program was an important mechanism through which this species evolved its unusually low ovariole number (relative both to the

673 other extant subgroup members and to its hypothesized last common ancestor (Green and Extavour 2012)).
674 In contrast, evolutionary changes in the JAK/STAT, Wnt, EGF and Notch signaling pathways may have
675 played a comparatively larger role in the evolution of more ovarioles in *D. simulans*, given the rapid
676 evolution of *Su(var)2*, *CKIIbeta*, *vn*, *Gug* and *E(spl)m2-BFM* along this branch (table 1, table S3).
677
678 **Future Directions**
679 The present study reveals a set of ovariole-involved genes, with established roles in ovariole
680 numbers and functions, whose protein sequence divergence is linked to ovariole number divergence in the
681 *Drosophila melanogaster* subgroup, based on a multi-layered analysis of branch-dN/dS, branch-site
682 analyses, *tau*, and PGLS. For many genes, the branch-dN/dS value was predictive of ovariole numbers
683 among species (table 4), consistent with an interdependent relationship. Further, our analyses of ovariole-
684 involved genes the in Hawaiian *Drosophila* clade suggests that protein divergence of ovariole-related
685 genes may shape ovariole number changes broadly across disparate clades of the *Drosophila* genus (table
686 S10, fig. S6). The molecular evolutionary approach used herein may provide valuable opportunities for
687 the discovery of genes and evolutionary processes involved in interspecies phenotype divergence,
688 particularly important for reproductive and fitness related traits (Dorus, et al. 2004; Nadeau, et al. 2007;
689 Ramm, et al. 2008; Wlasiuk and Nachman 2010; Luke, et al. 2014; Corso, et al. 2016; Chebbo, et al.
690 2021),which remains a central challenge in evolutionary developmental biology (Hoekstra and Coyne
691 2007; Cutter and Bundus 2020).
692 We suggest that future examinations of the genetic basis of interspecies divergence in ovariole
693 number and other related reproductive traits will be most fruitfully pursued along one or more of the
694 following major directions: First, assessments of protein sequence changes in ovariole-related genes
695 identified here at the population level using genome-wide association studies and mutational frequency

spectra (Akashi 1997; Whittle, et al. 2012; Lobell, et al. 2017), combined with McDonald-Kreitman tests (McDonald and Kreitman 1991; Murga-Moreno, et al. 2019), for multiple *Drosophila* species, will help discern evolutionary dynamics of these genes at the microevolutionary scale. Second, studies of expression divergence and functional divergence of genes in each species for the rapidly evolving ovariole-related genes identified here (table 1, table 2, table 3, table S7). Third, studies on the mating behaviors and sexual selection pressures, including male-mate-choice, female competition, and sexual antagonism, in species of the *melanogaster* subgroup (Bonduriansky 2001; Sirot, et al. 2014; Bath, et al. 2018; Veltsos, et al. 2022), will be valuable to revealing their possible links to ovariole numbers. Fourth, while we focused on the ovariole-related genes that had five-species orthologs in the *melanogaster* subgroup, ovariole number divergence may be also partly influenced by gene losses and gains in *Drosophila* lineages (Coyne and Hoekstra 2007; Tautz and Domazet-Loso 2011; Tautz, et al. 2013), as well as by genes that have diverged too rapidly to allow identification of orthologs (fig. S6) (Tautz and Domazet-Loso 2011; Tautz, et al. 2013), and thus those topics warrant further study. Finally, further research should include studies in the Hawaiian *Drosophila*, given our results suggest protein divergence of numerous ovariole-related genes may contribute to ovariole number changes in the three-species Hawaiian clade studied herein (table S10, fig. S5, fig. S6). The Hawaiian group is known for its wide phenotypic diversity in sexual characteristics, ranging from behaviours to ovariole numbers (Carson 1997; Singh and Singh 2014; Sarikaya, et al. 2019). Studies on the relationships between protein sequence changes and ovariole numbers in Hawaiian *Drosophila* will be facilitated by increased collection of whole genomes and transcriptomic data for the larval ovaries, including TFs and SH cells, and potentially by the use of expanding tools aimed to correlate gene and phenotype evolution (Kowalczyk, et al. 2019). Such research will help further decipher the genetic factors shaping the rapid evolution of ovariole numbers in the *Drosophila* genus, and thus in insects more broadly.

Materials and Methods

Identification of Rapidly-Evolving Ovariole-Related Genes for Follow-up Analyses

For the SIGNALC gene set, that was based on *D. melanogaster* RNAi data (Kumar, et al. 2020), we screened the 67 genes that directly affected ovariole numbers, named *hpo[RNAi]* Ovariole Number, 59 and 49 genes that affected egg laying, named *hpo[RNAi]* Egg Laying and Egg Laying *[wt]* and the 17 connector genes. For these four SIGNALC genes sets, we identified any genes with M0 dN/dS ≥1.5 higher than the genome-wide median. The cut-off was marginally lower than the BULKSG and SINGLEC because of the innate conserved nature of these signalling pathway genes, which are largely at least as old as animal divergence, in excess of 600 million years (Srivastava, et al. 2010; Kumar, et al. 2020). For the BULKSG dataset (Tarikere, et al. 2022), we screened for any differentially expressed genes that had M0 dN/dS≥0.20 in the *melanogaster* subgroup for further study. This represents a value ≥2.2 higher than the genome-wide median. With respect to the SINGLEC dataset (Slaidina, et al. 2020), for the genes with differential expression in one cell type relative to the others ($P<0.05$), we identified those with M0 dN/dS≥0.20, similar to the BULKSG dataset. The M0 dN/dS values for the five-species under study in the *melanogaster* subgroup were from FlyDivas (Stanley and Kulathinal 2016) that matched our own M0 dN/dS calculations in PAML (Yang 2007) (additional details on screening is available in Supplemental Text File 1).

Follow-up Assessments: dN/dS per Species Terminal Branch, Branch-Site Positive Selection, and *tau*

Determining dN/dS for each species terminal branch

We calculated the M1 free ratios dN/dS per species terminal branch using codeml package in PAML (Yang 2007), which allows a separate dN/dS value for each branch, using as input publicly available high confidence genome-wide five-species sequence alignments from FlyDivas, which has data for various species groups of *Drosophila* (Stanley and Kulathinal 2016). Codeml is based on maximum likelihood in deriving estimates of dN/dS values, and default parameters were used in the assessments (Yang 2007). Using the dN/dS for each of the five terminal species branches, we assessed associations with respect to species transitions in ovariole numbers (terminal species branch analysis), an approach that has proven effective for determining relationships between dN/dS values and phenotypes of interest (Dorus, et al. 2004; Nadeau, et al. 2007; Wlasiuk and Nachman 2010).

We assessed the distributions of dN/dS for all studied genes per species branch (fig. S1). To affirm the suitability of the obtained data to determine dN/dS in each individual species terminal branch, we examined the magnitude of dN and dS values. The vast majority of genes had dN and dS <1.5 per species terminal branch and thus were unsaturated: 99.95 and 99.5% of genes in *D. simulans* respectively had values below this threshold, and we found even higher percentages (up to 100%) for the four other species. Only gene branches that had dN or dS >0.001 were included for further assessment to ensure sufficient divergence for study (Cusack and Wolfe 2007; Whittle, et al. 2021). The minority of cases of a branch where dN was >0.001 and dS was at or near zero were denoted simply as "dN/dS>1" (e.g., 0.2% of all 9,232 genes studied in *D. melanogaster*, 2.2% in *D. simulans*), rather than infinity (see also other approaches to cases of dS near 0 and dN>0 (Wlasiuk and Nachman 2010) and were interpreted conservatively.

Branch-site positive selection analysis

763 Branch-site codon analyses was used to assess positive selection at specific codon sites for each
764 species terminal branch of the *melanogaster* subgroup (fig. 2) as described in the PAML manual (Yang
765 and Nielsen 2002; Zhang, et al. 2005; Yang 2007). For all aligned genes from the *melanogaster* subgroup
766 (N=9,237 alignments; note 9,232 had M0 values for study) (Stanley and Kulathinal 2016), including for
767 the identified rapidly evolving ovariole-related genes, one of the five *Drosophila* species was assigned as
768 the foreground branch in its own individual branch-site analysis. Thus, a separate branch-site analysis was
769 conducted for all studied genes for *D. simulans*, *D. sechellia*, *D. melanogaster*, *D. yakuba* and *D. erecta*.
770 For each gene, the maximum likelihood values were compared between a model with and without branch-
771 site positive selection (codeml Model=2, NSsites=2, with fix_omega=1 versus 0, and P value of Chi-
772 square for 2XΔlnL). P values <0.05 for 2XΔlnL for any gene were interpreted as evidence of positive
773 selection at one or more codon sites in that species branch. We studied the presence or absence of branch-
774 site positive selection within each gene, suggested by Zhang, et al. (2005), without including the post-hoc
775 option for BEB probability analysis per codon site that has low power (Zhang, et al. 2005). The frequency
776 of genes with branch-site positive selection in the ovariole-related gene sets under study were compared
777 to the genome-wide frequency per species branch. Multiple test corrections were not applied as this was
778 deemed overly conservative for our purposes of identification of ovariole-related genes with signals of
779 positive selection, and these results were combined with other multiple layers of analyses (branch dN/dS,
780 *tau*, and PGLS). The input tree for branch and branch-site analysis was an unrooted Newick phylogeny
781 (unrooted version of fig. 2) as required by PAML (Yang 2007).
782
783 *Expression specificity quantification using tau*
784 We used the index *tau* to measure expression specificity of the genes under study here (Yanai, et
785 al. 2005). For this, we accessed expression data from 59 tissue types and developmental stages from *D.*

melanogaster (30 developmental stages and 29 tissues, table S1). The data include gene expression levels (RPKM) across development for embryos (12 stages), larvae (6 stages), pupae (6 stages) and adults (3 stages of males/females), and for major tissue types of the adult males and females (including heads, gonads, and central nervous system). The expression data were from modEncode and included the RNA-seq datasets generated by Graveley, et al. (2011) (available at: https://flybase.org/commentaries/2013_05/rna-seq_bulk.html; downloaded March 2022; see also Supplementary Text File S1) which comprise among the widest scope of expression data available in insects (Li, et al. 2014). The *tau* value per gene was calculated as follows:

$$tau(\tau) = \frac{\sum_{i=1}^{n}(1-\hat{x}_i)}{n-1}\,;\, \hat{x}_i = x_i/max(x_i)$$

where *n*=number of tissues/stages studied, *i*= tissue/stage, x_i= expression level of gene in tissue/stage *i*, and max (x_i)= the expression level in the tissue/stage type with maximum expression (Yanai, et al. 2005).

Elevated values in one gene relative to another indicate greater expression specificity, such that most transcripts originate from few tissues/stages (see fig. S2 and Supplementary Text File S1 for an overview of the genome-wide *tau* values herein). Genes with *tau* values above 0.90 were considered highly specific in expression.

Phylogenetic Generalized Least Squares (PGLS) Analysis

PGLS was assessed for ovariole number (dependent parameter) with respect to branch-dN/dS (independent parameter) using the five terminal species branches of the *melanogaster* subgroup (fig. 2). PGLS was conducted using the Comparative analysis of phylogenetics and evolution (Caper) package available in R (R-Core-Team 2022) (https://cran.r-project.org/web/packages/caper/index.html). The

809 covariance matrix of species relationships was obtained under the assumption of Brownian motion using
810 the vcv function in caper. Under a five-species tree, any genes showing P<0.05 suggest a strong
811 relationship between ovariole number and dN/dS, sufficient to be detected under this sample size. In turn,
812 P>0.05 does not necessarily preclude a relationship, which may be inferred from our combined analysis
813 of dN/dS, positive selection analysis, and *tau*. The phylogenetic tree used for the covariance matrix in
814 PGLS is shown in fig. 2.
815
816 **McDonald-Kreitman Tests**
817 We conducted McDonald and Kreitman (1991) tests for genes of interest, using The Integrative
818 McDonald and Kreitman test (iMKT) database (Murga-Moreno, et al. 2019). For these tests, we examined
819 the Raleigh NC and Zambia populations, and the interspecies divergence was conducted using *D.*
820 *melanogaster-D. simulans* contrasts (Murga-Moreno, et al. 2019). Thus, this analysis tests positive
821 selection since divergence of the *D. melanogaster-D. simulans* branches only.
822
823 ***Drosophila* Phylogeny**
824 To obtain the phylogeny for the five-species *melanogaster* subgroup in fig. 2, we used aligned
825 sequence data from DrosoPhyla (Finet, et al. 2021) that contains a pre-screened dataset of 17 genes across
826 704 species of *Drosophilidae* (which were screened for quality, sufficient divergence, and phylogenetic
827 informativeness). We extracted the concatenated aligned sequences for *D. simulans*, *D. sechellia*, *D.*
828 *melanogaster*, *D. yakuba* and *D. erecta*, included *D. ananassae* as an outgroup as a reference (for the
829 phylogeny construction), and removed all gaps and any sites with unknown nucleotides, yielding a total
830 of 9,235 nucleotide sites. Using MEGA11 (Tamura, et al. 2021), we generated a maximum likelihood
831 (ML) phylogenetic tree, including the tree lengths, based on the default parameters. We also obtained a

tree using the Neighbor-Joining (NJ) Method, with nearly identical results. The relative relationships of the species in the obtained trees matched those previously observed for these five species (Obbard, et al. 2012; Finet, et al. 2021).

Hierarchical Clustering of Expression in the SINGLEC Dataset

The relationships in gene expression across the nine different cell types of the *D. melanogaster* LL3 ovary (fig. 1A) from the SINGLEC dataset (Slaidina, et al. 2020) were assessed using hierarchical clustering under the average linkage method applied to the average standardized expression values per gene for all genes with nonzero expression (determined in Suerat v2, see Slaidina, et al. (2020)). The analysis was conducted in the Morpheus program (https://software.broadinstitute.org/morpheus).

Gene Ontology

To study inferred gene functions and the clustering of genes by inferred function we used the program DAVID (Huang da, et al. 2009), which provides inferred gene function data for *D. melanogaster* using the FlyBase gene identifiers (Gramates, et al. 2022).

Supplementary Analyses of a three-species Hawaiian Clade

We followed up on our main assessments of the *melanogaster* subgroup, with a supplementary evaluation of ovariole numbers and ovariole-related gene dN/dS in a three-species clade from the distantly related Hawaiian *Drosophila*, that included *D. sproati*, *D. murphyi* and *D. grimshawi*. The methods applied for CDS extraction, ortholog identification, gene alignments and dN/dS analyses for that assessment are described in Supplementary Text File S1.

Data Availability

All data used in the present study are publicly available as described in Materials and Methods and Supplementary Text File S1.

Acknowledgements

The authors thank members of the Extavour lab for valuable discussions. The experimental and transcriptome data generated by cited research of Dr. Tarun Kumar, Dr. Leo Blondel, Dr. Shreeharsha Tarikere and Dr. Guillem Ylla, that allowed pre-screening of genes for ovariole functions is appreciated, as well as by the authors of the sc-RNA seq datasets cited in the Materials and Methods. The authors also appreciate valuable comments by the anonymous Reviewers that helped improve the manuscript.

Author Contributions

CAW and CGE conceived the study, wrote the manuscript and conducted analysis.

References

Akashi H. 1997. Codon bias evolution in *Drosophila*. Population genetics of mutation-selection drift. Gene 205:269-278.

Arnqvist G, Rock, L. 1995. Sexual conflict and arms races between the sexes: a morphological adaptation for control of mating in a female insect. Proceedings of the Royal Society. Section B: Biology:261123–261127.

Assis R, Zhou Q, Bachtrog D. 2012. Sex-biased transcriptome evolution in Drosophila. Genome Biology and Evolution 4:1189-1200.

Bath E, Morimoto J, Wigby S. 2018. The developmental environment modulates mating-induced aggression and fighting success in adult female *Drosophila*. Functional Ecology 32:2542-2552.

Bergland AO, Genissel A, Nuzhdin SV, Tatar M. 2008. Quantitative trait loci affecting phenotypic plasticity and the allometric relationship of ovariole number and thorax length in *Drosophila melanogaster*. Genetics 180:567-582.

Bielawski JP, Yang Z. 2005. Statistical Methods in Molecular Evolution. In Maximum Likelihood Methods for Detecting Adaptive Protein Evolution:103-124.

Bonduriansky R. 2001. The evolution of male mate choice in insects: a synthesis of ideas and evidence. Biological Reviews of the Cambridge Philosophical Society 76:305-339.

Borges R, Fonseca J, Gomes C, Johnson WE, O'Brien SJ, Zhang G, Gilbert MTP, Jarvis ED, Antunes A. 2019. Avian Binocularity and Adaptation to Nocturnal Environments: Genomic Insights from a Highly Derived Visual Phenotype. Genome Biology and Evolution 11:2244-2255.

Bromham L, Rambaut A, Harvey PH. 1996. Determinants of rate variation in mammalian DNA sequence evolution. Journal of Molecular Evolution 43:610-621.

Bubnell JE, Ulbing CKS, Fernandez Begne P, Aquadro CF. 2022. Functional Divergence of the bag-of-marbles Gene in the *Drosophila* melanogaster Species Group. Mol Biol Evol 39.

Buschiazzo E, Ritland C, Bohlmann J, Ritland K. 2012. Slow but not low: genomic comparisons reveal slower evolutionary rate and higher dN/dS in conifers compared to angiosperms. BMC Evolutionary Biology 12:8.

Byrne PG, Rice WR. 2006. Evidence for adaptive male mate choice in the fruit fly *Drosophila melanogaster*. Proceedings: Biological Sciences 273:917-922.

Carson HL. 1997. The Wilhelmine E. Key 1996 Invitational Lecture. Sexual selection: a driver of genetic change in Hawaiian *Drosophila*. Journal of Heredity 88:343-352.

Castillo-Davis CI, Bedford TB, Hartl DL. 2004. Accelerated rates of intron gain/loss and protein evolution in duplicate genes in human and mouse malaria parasites. Molecular Biology and Evolution 21:1422-1427.

Chapman T, Herndon LA, Heifetz Y, Partridge L, Wolfner MF. 2001. The Acp26Aa seminal fluid protein is a modulator of early egg hatchability in *Drosophila melanogaster*. Proc Biol Sci 268:1647–1654.

Chebbo S, Josway S, Belote JM, Manier MK. 2021. A putative novel role for Eip74EF in male reproduction in promoting sperm elongation at the cost of male fecundity. J Exp Zool B Mol Dev Evol 336:620-628.

Church SH, de Medeiros BAS, Donoughe S, Marquez Reyes NL, Extavour CG. 2021. Repeated loss of variation in insect ovary morphology highlights the role of development in life-history evolution. Proceedings: Biological Sciences 288:20210150.

Civetta A, Clark AG. 2000. Correlated effects of sperm competition and postmating female mortality. Proceedings of the National Academy of Sciences of the United States of America 97:13162-13165.

Civetta A, Singh RS. 1998. Sex-related genes, directional sexual selection, and speciation. Molecular Biology and Evolution 15:901-909.

Clark IB, Jarman AP, Finnegan DJ. 2007. Live imaging of *Drosophila* gonad formation reveals roles for Six4 in regulating germline and somatic cell migration. BMC Developmental Biology 7:52.

Clark NL, Gasper J, Sekino M, Springer SA, Aquadro CF, Swanson WJ. 2009. Coevolution of interacting fertilization proteins. PLoS Genetics 5:e1000570.

Corso J, Mundy NI, Fagundes NJ, de Freitas TR. 2016. Evolution of dark colour in toucans (Ramphastidae): a case of molecular adaptation? Journal of Evolutionary Biology 29:2530-2538.

Coyne JA, Charlesworth B. 1997. Genetics of a pheromonal difference affecting sexual isolation between *Drosophila mauritiana* and *D. sechellia*. Genetics 145:1015-1030.

Coyne JA, Hoekstra HE. 2007. Evolution of protein expression: new genes for a new diet. Current Biology 17:R1014-1016.

Coyne JA, Rux J, David JR. 1991. Genetics of morphological differences and hybrid sterility between *Drosophila sechellia* and its relatives. Genetical Research 57:113-122.

Cui Q, Purisima EO, Wang E. 2009. Protein evolution on a human signaling network. BMC Systems Biology 3:21.

Cusack BP, Wolfe KH. 2007. Not born equal: increased rate asymmetry in relocated and retrotransposed rodent gene duplicates. Mol Biol Evol 24:679-686.

Cutter AD. 2008. Divergence times in Caenorhabditis and Drosophila inferred from direct estimates of the neutral mutation rate. Molecular Biology and Evolution 25:778-786.

Cutter AD, Bundus JD. 2020. Speciation and the developmental alarm clock. Elife 9.

Dansereau DA, Lasko P. 2008. The development of germline stem cells in *Drosophila*. Methods in Molecular Biology 450:3-26.

Dean R, Mank JE. 2016. Tissue Specificity and Sex-Specific Regulatory Variation Permit the Evolution of Sex-Biased Gene Expression. The American Naturalist 188:E74-E84.

Dorus S, Evans PD, Wyckoff GJ, Choi SS, Lahn BT. 2004. Rate of molecular evolution of the seminal protein gene SEMG2 correlates with levels of female promiscuity. Nature Genetics 36:1326-1329.

Duret L, Mouchiroud D. 2000. Determinants of substitution rates in mammalian genes: expression pattern affects selection intensity but not mutation rate. Molecular Biology and Evolution 17:68-74.

Eliazer S, Buszczak M. 2011. Finding a niche: studies from the *Drosophila* ovary. Stem Cell Research & Therapy 2:45.

Extavour CG, Akam ME. 2003. Mechanisms of germ cell specification across the metazoans: epigenesis and preformation. Development 130:5869-5884.

Felsenstein J. 1985. Phylogenies and the Comparative Method. American Naturalist 125:1-15.

Finet C, Kassner VA, Carvalho AB, Chung H, Day JP, Day S, Delaney EK, De Re FC, Dufour HD, Dupim E, et al. 2021. DrosoPhyla: Resources for Drosophilid Phylogeny and Systematics. Genome Biology and Evolution 13.

Fisher RA. 1930. The genetical theory of natural selection: Oxford: Clarendon Press.

Friberg U, Arnqvist G. 2003. Fitness effects of female mate choice: preferred males are detrimental for *Drosophila melanogaster* females. Journal of Evolutionary Biology 16:797-811.

Gilboa L, Lehmann R. 2006. Soma-germline interactions coordinate homeostasis and growth in the *Drosophila* gonad. Nature 443:97-100.

Godt D, Laski FA. 1995. Mechanisms of cell rearrangement and cell recruitment in *Drosophila* ovary morphogenesis and the requirement of bric a brac. Development 121:173-187.

Goldman-Huertas B, Mitchell RF, Lapoint RT, Faucher CP, Hildebrand JG, Whiteman NK. 2015. Evolution of herbivory in Drosophilidae linked to loss of behaviors, antennal responses, odorant receptors, and ancestral diet. Proceedings of the National Academy of Sciences of the United States of America 112:3026-3031.

Gossmann TI, Keightley PD, Eyre-Walker A. 2012. The effect of variation in the effective population size on the rate of adaptive molecular evolution in eukaryotes. Genome Biology and Evolution 4:658-667.

Gramates LS, Agapite J, Attrill H, Calvi BR, Crosby MA, Dos Santos G, Goodman JL, Goutte-Gattat D, Jenkins VK, Kaufman T, et al. 2022. FlyBase: a guided tour of highlighted features. Genetics 220.

Graveley BR, Brooks AN, Carlson JW, Duff MO, Landolin JM, Yang L, Artieri CG, van Baren MJ, Boley N, Booth BW, et al. 2011. The developmental transcriptome of *Drosophila melanogaster*. Nature 471:473-479.

Green DA, 2nd, Extavour CG. 2012. Convergent evolution of a reproductive trait through distinct developmental mechanisms in *Drosophila*. Developmental Biology 372:120-130.

Haerty W, Jagadeeshan S, Kulathinal RJ, Wong A, Ram KR, Sirot LK, Levesque L, Artieri CG, Wolfner MF, Civetta A, et al. 2007. Evolution in the fast lane: rapidly evolving sex-related genes in Drosophila. Genetics 177:1321-1335.

Ho WW, Smith SD. 2016. Molecular evolution of anthocyanin pigmentation genes following losses of flower color. BMC Evolutionary Biology 16:98.

Hodin J, Riddiford LM. 2000. Different mechanisms underlie phenotypic plasticity and interspecific variation for a reproductive character in drosophilids (Insecta: Diptera). Evolution 54:1638-1653.

Hoekstra HE, Coyne JA. 2007. The locus of evolution: evo devo and the genetics of adaptation. Evolution 61:995-1016.

Huang da W, Sherman BT, Lempicki RA. 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nature protocols 4:44-57.

Innocenti P, Morrow EH. 2010. The sexually antagonistic genes of *Drosophila melanogaster*. PLoS Biology 8:e1000335.

Jagadeeshan S, Singh RS. 2007. Rapid evolution of outer egg membrane proteins in the *Drosophila melanogaster* subgroup: a case of ecologically driven evolution of female reproductive traits. Mol Biol Evol 24:929-938.

Jagadeeshan S, Singh RS. 2005. Rapidly evolving genes of Drosophila: differing levels of selective pressure in testis, ovary, and head tissues between sibling species. Molecular Biology and Evolution 22:1793-1801.

Kambysellis MP, Heed WB. 1971. Studies of Oogenesis in Natural Populations of Drosophilidae. I. Relation of Ovarian Development and Ecological Habitats of the Hawaiian Species. American Society of Naturalists 105:31-49.

Kambysellis MP, Ho KF, Craddock EM, Piano F, Parisi M, Cohen J. 1995. Pattern of ecological shifts in the diversification of Hawaiian *Drosophila* inferred from a molecular phylogeny. Current Biology 5:1129-1139.

Kaneshiro KY, Boake CR. 1987. Sexual selection and speciation: Issues raised by Hawaiian *Drosophila*. Trends in Ecology & Evolution 2:207-212.

Kang L, Settlage R, McMahon W, Michalak K, Tae H, Garner HR, Stacy EA, Price DK, Michalak P. 2016. Genomic Signatures of Speciation in Sympatric and Allopatric Hawaiian Picture-Winged *Drosophila*. Genome Biology and Evolution 8:1482-1488.

Kim BY, Wang JR, Miller DE, Barmina O, Delaney E, Thompson A, Comeault AA, Peede D, D'Agostino ER, Pelaez J, et al. 2021. Highly contiguous assemblies of 101 drosophilid genomes. Elife 10.

Kim PM, Korbel JO, Gerstein MB. 2007. Positive selection at the protein network periphery: evaluation in terms of structural constraints and cellular context. Proceedings of the National Academy of Sciences of the United States of America 104:20274-20279.

Kimura M. 1983. The neutral theory of molecular evolution. Cambridge: Cambridge University Press.

Kimura M. 1989. The neutral theory of molecular evolution and the world view of the neutralists. Genome 31:24-31.

Kimura M. 1962. On the probability of fixation of mutant genes in a population. Genetics 47:713-719.

King RC. 1970. Ovarian Development in *Drosophila melanogaster*: New York: Academic Press.

King RC, Aggarwal SK, Aggarwal U. 1968. The development of the female *Drosophila* reproductive system. Journal of Morphology 124:143-166.

Kiss AA, Somlyai-Popovics N, Kiss M, Boldogkői Z, Csiszár K, Mink M. 2019. Type IV Collagen Is Essential for Proper Function of Integrin-Mediated Adhesion in Drosophila Muscle Fibers. International Journal of Molecular Sciences 20.

Kong HG, Kim HH, Chung JH, Jun J, Lee S, Kim HM, Jeon S, Park SG, Bhak J, Ryu CM. 2019. The *Galleria mellonella* Hologenome Supports Microbiota-Independent Metabolism of Long-Chain Hydrocarbon Beeswax. Cell Reports 26:2451-2464 e2455.

Kowalczyk A, Meyer WK, Partha R, Mao W, Clark NL, Chikina M. 2019. RERconverge: an R package for associating evolutionary rates with convergent traits. Bioinformatics 35:4815-4817.

Kumar T, Blondel L, Extavour CG. 2020. Topology-driven analysis of protein-protein interaction networks detects functional genetic sub-networks regulating reproductive capacity. Elife 9:e54082.

LaBella AL, Opulente DA, Steenwyk JL, Hittinger CT, Rokas A. 2021. Signatures of optimal codon usage in metabolic genes inform budding yeast ecology. PLoS Biology 19:e3001185.

Lande R. 1981. Models of speciation by sexual selection on polygenic traits. Proceedings of the National Academy of Sciences of the United States of America 78:3721-3725.

Larracuente AM, Sackton TB, Greenberg AJ, Wong A, Singh ND, Sturgill D, Zhang Y, Oliver B, Clark AG. 2008. Evolution of protein-coding genes in *Drosophila*. Trends in Genetics 24:114-123.

Lebo DPV, McCall K. 2021. Murder on the Ovarian Express: A Tale of Non-Autonomous Cell Death in the *Drosophila* Ovary. Cells 10.

Legrand D, Tenaillon MI, Matyot P, Gerlach J, Lachaise D, Cariou ML. 2009. Species-wide genetic variation and demographic history of *Drosophila sechellia*, a species lacking population structure. Genetics 182:1197-1206.

LeVasseur-Viens H, Polak M, Moehring AJ. 2015. No evidence for external genital morphology affecting cryptic female choice and reproductive isolation in *Drosophila*. Evolution 69:1797-1807.

Li JJ, Huang H, Bickel PJ, Brenner SE. 2014. Comparison of *D. melanogaster* and *C. elegans* developmental stages, tissues, and cells by modENCODE RNA-seq data. Genome Research 24:1086-1101.

Li MA, Alls JD, Avancini RM, Koo K, Godt D. 2003. The large Maf factor Traffic Jam controls gonad morphogenesis in *Drosophila*. Nature Cell Biology 5:994-1000.

Lobell AS, Kaspari RR, Serrano Negron YL, Harbison ST. 2017. The Genetic Architecture of Ovariole Number in *Drosophila melanogaster*: Genes with Major, Quantitative, and Pleiotropic Effects. G3 (Bethesda) 7:2391-2403.

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology 15:550.

Luke L, Vicens A, Tourmente M, Roldan ER. 2014. Evolution of protamine genes and changes in sperm head phenotype in rodents. Biology of Reproduction 90:67.

Lupold S, Manier MK, Puniamoorthy N, Schoff C, Starmer WT, Luepold SH, Belote JM, Pitnick S. 2016. How sexual selection can drive the evolution of costly sperm ornamentation. Nature 533:535-538.

Macagno AL, Beckers OM, Moczek AP. 2015. Differentiation of ovarian development and the evolution of fecundity in rapidly diverging exotic beetle populations. Journal of Experimental Zoology. Part A: Ecological Genetics and Physiology 323:679-688.

Mank JE, Ellegren H. 2009. Are sex-biased genes more dispensable? Biology letters 5:409-412.

Mank JE, Hultin-Rosenberg L, Zwahlen M, Ellegren H. 2008. Pleiotropic constraint hampers the resolution of sexual antagonism in vertebrate gene expression. American Naturalist 171:35-43.

Markow TA. 2002. Perspective: female remating, operational sex ratio, and the arena of sexual selection in *Drosophila* species. Evolution 56:1725-1734.

Markow TA, Beall S, Matzkin LM. 2009. Egg size, embryonic development time and ovoviviparity in *Drosophila* species. Journal of Evolutionary Biology 22:430-434.

Markow TA, O'Grady PM. 2005. Evolutionary genetics of reproductive behavior in *Drosophila*: connecting the dots. Annual Review of Genetics 39:263-291.

Masalia RR, Bewick AJ, Burke JM. 2017. Connectivity in gene coexpression networks negatively correlates with rates of molecular evolution in flowering plants. PloS One 12:e0182289.

McDonald JH, Kreitman M. 1991. Adaptive protein evolution at the Adh locus in Drosophila. Nature 351:652-654.

Meisel RP. 2011. Towards a more nuanced understanding of the relationship between sex-biased gene expression and rates of protein-coding sequence evolution. Molecular Biology and Evolution 28:1893-1900.

Mensch J, Serra F, Lavagnino NJ, Dopazo H, Hasson E. 2013. Positive selection in nucleoporins challenges constraints on early expressed genes in *Drosophila* development. Genome Biology and Evolution 5:2231-2241.

Miller PB, Obrik-Uloho OT, Phan MH, Medrano CL, Renier JS, Thayer JL, Wiessner G, Bloch Qazi MC. 2014. The song of the old mother: reproductive senescence in female *Drosophila*. Fly (Austin) 8:127-139.

Mitterboeck TF, Liu S, Adamowicz SJ, Fu J, Zhang R, Song W, Meusemann K, Zhou X. 2017. Positive and relaxed selection associated with flight evolution and loss in insect transcriptomes. Gigascience 6:1-14.

Montague JR, Mangan RL, Starmer WT. 1981. Reproductive allocation in the Hawaiian Drosophilidae: egg size and number. American Naturalist 118:865-871.

Montgomery SH, Capellini I, Venditti C, Barton RA, Mundy NI. 2011. Adaptive evolution of four microcephaly genes and the evolution of brain size in anthropoid primates. Mol Biol Evol 28:625-638.

Munds RA, Titus CL, Moreira LAA, Eggert LS, Blomquist GE. 2021. Examining the molecular basis of coat color in a nocturnal primate family (*Lorisidae*). Ecology and Evolution 11:4442-4459.

Murga-Moreno J, Coronado-Zamora M, Hervas S, Casillas S, Barbadilla A. 2019. iMKT: the integrative McDonald and Kreitman test. Nucleic Acids Research 47:W283-W288.

Nadeau NJ, Burke T, Mundy NI. 2007. Evolution of an avian pigmentation gene correlates with a measure of sexual selection. Proceedings: Biological Sciences 274:1807-1813.

Obbard DJ, Maclennan J, Kim KW, Rambaut A, O'Grady PM, Jiggins FM. 2012. Estimating divergence dates and substitution rates in the *Drosophila* phylogeny. Mol Biol Evol 29:3459-3473.

Otto SP. 2004. Two steps forward, one step back: the pleiotropic effects of favoured alleles. Proceedings: Biological Sciences 271:705-714.

Proschel M, Zhang Z, Parsch J. 2006. Widespread adaptive evolution of Drosophila genes with sex-biased expression. Genetics 174:893-900.

R-Core-Team. 2022. R: A language and environment for statistical computing. . Vienna, Austria: R Foundation for Statistical Computing.

Ramm SA, Oliver PL, Ponting CP, Stockley P, Emes RD. 2008. Sexual selection and the adaptive evolution of mammalian ejaculate proteins. Mol Biol Evol 25:207-219.

Ranz JM, Castillo-Davis CI, Meiklejohn CD, Hartl DL. 2003. Sex-dependent gene expression and evolution of the *Drosophila* transcriptome. Science 300:1742-1745.

Rice WR. 1996. Sexually antagonistic male adaptation triggered by experimental arrest of female evolution. Nature 381:232-234.

Sahut-Barnola I, Dastugue B, Couderc JL. 1996. Terminal filament cell organization in the larval ovary of *Drosophila melanogaster*: ultrastructural observations and pattern of divisions. Roux's Archives of Developmental Biology 205:356-363.

Sarikaya DP, Belay AA, Ahuja A, Dorta A, Green DA, Extavour CG. 2012. The roles of cell size and cell number in determining ovariole number in Drosophila. Developmental Biology 363:279-289.

Sarikaya DP, Church SH, Lagomarsino LP, Magnacca KN, Montgomery SL, Price DK, Kaneshiro KY, Extavour CG. 2019. Reproductive Capacity Evolves in Response to Ecology through Common Changes in Cell Number in Hawaiian Drosophila. Current Biology 29:1877-1884.e1876.

Sarikaya DP, Extavour CG. 2015. The Hippo Pathway Regulates Homeostatic Growth of Stem Cell Niche Precursors in the Drosophila Ovary. PLoS Genetics 11:e1004962.

Satija R, Farrell JA, Gennert D, Schier AF, Regev A. 2015. Spatial reconstruction of single-cell gene expression data. Nature Biotechnology 33:495-502.

Sinclair CS, Lisa SF, Pischedda A. 2021. Does sexual experience affect the strength of male mate choice for high-quality females in *Drosophila melanogaster*? Ecology and Evolution 11:16981-16992.

Singh A, Singh BN. 2014. Role of sexual selection in speciation in Drosophila. Genetica 142:23-41.

Singh SR, Singh BN, Colorado Adoption P. 2001. Female remating in *Drosophila ananassae*: bidirectional selection for remating speed. Behavior Genetics 31:361-370.

Singh SR, Singh BN, Hoenigsberg HF. 2002. Female remating, sperm competition and sexual selection in *Drosophila*. Genetics and Molecular Research 1:178-215.

Sirot LK, Wong A, Chapman T, Wolfner MF. 2014. Sexual conflict and seminal fluid proteins: a dynamic landscape of sexual interactions. Cold Spring Harbor Perspectives in Biology 7:a017533.

Slaidina M, Banisch TU, Gupta S, Lehmann R. 2020. A single-cell atlas of the developing Drosophila ovary identifies follicle stem cell progenitors. Genes Dev 34:239-249.

Slaidina M, Gupta S, Banisch TU, Lehmann R. 2021. A single-cell atlas reveals unanticipated cell type complexity in *Drosophila* ovaries. Genome Research 31:1938-1951.

Song X, Zhu CH, Doan C, Xie T. 2002. Germline stem cells anchored by adherens junctions in the *Drosophila* ovary niches. Science 296:1855-1857.

Srivastava M, Simakov O, Chapman J, Fahey B, Gauthier ME, Mitros T, Richards GS, Conaco C, Dacre M, Hellsten U, et al. 2010. The Amphimedon queenslandica genome and the evolution of animal complexity. Nature 466:720-726.

Stanley CE, Jr., Kulathinal RJ. 2016. flyDIVaS: A Comparative Genomics Resource for Drosophila Divergence and Selection. G3 (Bethesda) 6:2355-2363.

Starmer W, Polak M, Pitnick S, McEvey S, Barke rJ, Wolf L. 2003. Phylogenetic, Geographical, and Temporal Analysis of Female Reproductive Trade-Offs in Drosophilidae. Evolutionary Biology 33:139-171.

Strasburg JL, Kane NC, Raduski AR, Bonin A, Michelmore R, Rieseberg LH. 2011. Effective population size is positively correlated with levels of adaptive divergence among annual sunflowers. Mol Biol Evol 28:1569-1580.

Suvorov A, Kim BY, Wang J, Armstrong EE, Peede D, D'Agostino ERR, Price DK, Waddell P, Lang M, Courtier-Orgogozo V, et al. 2022. Widespread introgression across a phylogeny of 155 *Drosophila* genomes. Current Biology 32:111-123 e115.

Swanson WJ, Vacquier VD. 2002. The rapid evolution of reproductive proteins. Nature Reviews: Genetics 3:137-144.

Swanson WJ, Wong A, Wolfner MF, Aquadro CF. 2004. Evolutionary expressed sequence tag analysis of *Drosophila* female reproductive tracts identifies genes subjected to positive selection. Genetics 168:1457-1465.

Symonds MRE, Blomberg SP. 2014. "Modern Phylogenetic Comparative Methods and Their Application in Evolutionary Biology". In: A Primer on Phylogenetic Generalised Least Squares Publisher: Springer-Verlag (Editor: L. Z. Garamszegi) Berlin Heidelberg, pp.105-130.

Tamura K, Stecher G, Kumar S. 2021. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. Mol Biol Evol 38:3022-3027.

Tamura K, Subramanian S, Kumar S. 2004. Temporal patterns of fruit fly (*Drosophila*) evolution revealed by mutation clocks. Molecular Biology and Evolution 21:36-44.

Tarikere S, Ylla G, Extavour CG. 2022. Distinct gene expression dynamics in germ line and somatic tissue during ovariole morphogenesis in *Drosophila melanogaster*. G3 (Bethesda) 12: DOI: 10.1093/g1093journal/jkab1305

Tautz D, Domazet-Loso T. 2011. The evolutionary origin of orphan genes. Nature Reviews Genetics 12:692-702.

Tautz D, Neme R, Domazet-Loso T. 2013. Evolutionary Origin of Orphan Genes. eLS.

Thomas JA, Welch JJ, Lanfear R, Bromham L. 2010. A generation time effect on the rate of molecular evolution in invertebrates. Mol Biol Evol 27:1173-1180.

Treangen TJ, Rocha EP. 2011. Horizontal transfer, not duplication, drives the expansion of protein families in prokaryotes. PLoS Genetics 7:e1001284.

Veltsos P, Porcelli D, Fang Y, Cossins AR, Ritchie MG, Snook RR. 2022. Experimental sexual selection reveals rapid evolutionary divergence in sex-specific transcriptomes and their interactions following mating. Molecular Ecology 31:3374-3388.

Wagner GP, Zhang J. 2011. The pleiotropic structure of the genotype-phenotype map: the evolvability of complex organisms. Nature Reviews: Genetics 12:204-213.

Wayne ML, McIntyre LM. 2002. Combining mapping and arraying: An approach to candidate gene identification. Proceedings of the National Academy of Sciences of the United States of America 99:14903-14906.

Whittle CA, Extavour CG. 2019. Selection shapes turnover and magnitude of sex-biased expression in *Drosophila* gonads. BMC Evolutionary Biology 19:60 (doi: 10.1186/s12862-12019-11377-12864).

Whittle CA, Johnston MO. 2003. Broad-scale analysis contradicts the theory that generation time affects molecular evolutionary rates in plants. Journal of Molecular Evolution 56:223-233.

Whittle CA, Kulkarni A, Extavour CG. 2021. Evolutionary dynamics of sex-biased genes expressed in cricket brains and gonads. Journal of Evolutionary Biology 34:1188-1211.

Whittle CA, Sun Y, Johannesson H. 2012. Genome-wide selection on codon usage at the population level in the fungal model organism *Neurospora crassa*. Mol Biol Evol 29:1975-1986.

Wigby S, Brown NC, Allen SE, Misra S, Sitnik JL, Sepil I, Clark AG, Wolfner MF. 2020. The *Drosophila* seminal proteome and its role in postcopulatory sexual selection. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences 375:20200072.

Wlasiuk G, Nachman MW. 2010. Promiscuity and the rate of molecular evolution at primate immunity genes. Evolution 64:2204-2220.

Yanai I, Benjamin H, Shmoish M, Chalifa-Caspi V, Shklar M, Ophir R, Bar-Even A, Horn-Saban S, Safran M, Domany E, et al. 2005. Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. Bioinformatics 21:650-659.

Yang Z. 1998. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. Mol Biol Evol 15:568-573.

Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Molecular Biology and Evolution 24:1586-1591.

Yang Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. Computer Applications in the Biosciences 13:555-556.

Yang Z, Nielsen R. 2002. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. Mol Biol Evol 19:908-917.

Yasothornsrikul S, Davis WJ, Cramer G, Kimbrell DA, Dearolf CR. 1997. viking: identification and characterization of a second type IV collagen in *Drosophila*. Gene 198:17-25.

Zhang J, Nielsen R, Yang Z. 2005. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. Molecular Biology and Evolution 22:2472-2479.

Table 1. The gene-wide dN/dS per species branch values for each of the 27 signalling or connector genes (determined to be evolving rapidly in table S2) in the five terminal species branches in the *melanogaster* subgroup of *Drosophila*. Branch-site positive selection (BR-S pos. sel,) analysis and cases with P<0.05 are shown by species name (Dsim = *D. simulans*, Dsec = *D. sechellia*, Dmel = *D. melanogaster*, Dyak = *D. yakuba* and Dere = *D. erecta*). The ovariole number/egg laying phenotypic categories defined in the RNAi experiments from (Kumar, et al. 2020) are shown here as: H-ON for *hpo[RNAi]* Ovariole Number, H-EL for *hpo[RNAi]* Egg Laying, and EL for the Egg Laying *[wt]*, and genes designated in that study as "connector genes" with observed phenotypes (on ovariole number or egg laying) are also shown.

Fbgn ID	CG No.	Name	Symbol	Branch dN/dS					BR-S pos. sel. P<0.05)	Gene phenotypic category in (Kumar, et al. 2020)
				Dsim	Dsec	Dmel	Dyak	Dere		
FBgn0011274	*CG6794*	*Dorsal-related immunity factor*	*Dif*	0.0001	0.5981	0.0001	**0.7233**	0.4146	Dyak	H-EL
FBgn0014020	*CG8416*	*Rho1*	*Rho1*	-	0.0001	-	0.0001	-		H-ON,H-EL,EL
FBgn0003612	*CG8068*	*Suppressor of variegation 2-10*	*Su(var)2-10*	**0.5723**	0.0001	0.5482	0.0662	0.0122	Dsim, Dmel	H-ON,H-EL,EL
FBgn0026379	*CG5671*	*Phosphatase and tensin homolog*	*Pten*	0.0001	0.1773	0.3122	0.1278	**0.5944**	Dere	H-ON
FBgn0000259	*CG15224*	*Casein kinase II beta subunit*	*CkIIbeta*	**0.6516**	0.2694	0.0001	0.0001	0.0001	Dsim	H-ON,H-EL,EL
FBgn0035213	*CG2199*	*CG2199*	*CG2199*	**1.0905**	0.404	0.3582	0.328	0.2765		Connector
FBgn0011642	*CG32018*	*Zyxin*	*Zyx*	0.3100	**>1**	0.2877	0.2668	0.3222		H-EL
FBgn0262614	*CG43140*	*polychaetoid*	*pyd*	0.0165	0.0341	**0.4745**	0.0168	0.0969	Dmel	H-ON
FBgn0036974	*CG5605*	*eukaryotic translation release factor 1*	*eRF1*	0.0001	0.1445	**0.3901**	0.0001	0.0697	Dmel	H-ON,H-EL,EL
FBgn0003984	*CG10491*	*vein*	*vn*	**0.4712**	0.2069	0.0841	0.2802	0.1511	Dyak	H-ON
FBgn0004858	*CG4220*	*elbow B*	*elB*	0.0001	**0.6159**	0.0297	0.066	0.0617	Dsec	H-ON
FBgn0010825	*CG6964*	*Grunge*	*Gug*	**0.5467**	0.3674	0.0539	0.0469	0.0416	Dsim, Dsec	H-ON,H-EL,EL
FBgn0002174	*CG5504*	*CG5504*	*CG5504*	0.2527	**0.3135**	0.0423	0.0812	0.0937	Dsec	H-ON
FBgn0037218	*CG1107*	*auxilin*	*aux*	0.0648	0.1874	0.1639	**0.2738**	0.2413	Dere	H-EL,EL
FBgn0259176	*CG42281*	*bunched*	*bun*	0.0661	0.2569	0.1009	0.1814	**0.2716**	Dere	H-ON
FBgn0023540	*CG3630*	*CG3630*	*CG3630*	0.3585	**0.5938**	0.1247	0.239	0.1129		Connector
FBgn0261854	*CG42783*	*atypical protein kinase C*	*aPKC*	**0.1931**	0.0126	0.0001	0.0001	0.0855	Dsim	H-EL
FBgn0001169	*CG5460*	*Hairless*	*H*	0.1982	0.1646	0.1585	**0.2220**	0.1746		H-ON
FBgn0024291	*CG5216*	*Sirtuin 1*	*Sirt1*	0.0001	0.1876	**0.2589**	0.1113	0.071	Dmel	H-EL,EL
FBgn0030904	*CG5988*	*unpaired 2*	*upd2*	-	**0.4168**	0.0347	0.0793	0.1667	Dsec, Dere	H-ON,H-EL,EL
FBgn0020496	*CG7583*	*C-terminal Binding Protein*	*CtBP*	**0.5440**	0.0001	0.0697	0.0001	0.1103	Dsim	H-ON,H-EL,EL
FBgn0003607	*CG8409*	*Suppressor of variegation 205*	*Su(var)205*	0.1102	0.0001	**0.2107**	0.058	0.0634		Connector
FBgn0261592	*CG10944*	*Ribosomal protein S6*	*RpS6*	0.0001	**0.3052**	0.0179	0.0001	0.0001	Dsec	H-ON,H-EL,EL

FBgn0020386	*CG1210*	*Phosphoinositide-dependent kinase 1*	*Pdk1*	0.3558	**0.4589**	0.0996	0.0624	0.0490	Dsim, Dsec	H-ON
FBgn0002592	*CG6104*	*Enhancer of split m2, Bearded family member*	*E(spl)m2-BFM*	**0.5554**	0.2217	0.0197	0.1984	0.1469	Dyak	H-ON
FBgn0032006	*CG8222*	*PDGF- and VEGF-receptor related*	*Pvr*	0.0077	**0.2500**	0.0535	0.1614	0.2063		H-EL
FBgn0045035	*CG6535*	*telomere fusion*	*tefu*	0.0868	**0.1908**	0.1222	0.1051	0.1411		H-OV,EL

Notes: a value of ">1" indicates that dN/dS>1 and that PAML indicates the value of infinity, where dN>0.001 and typically dS are approaching zero, and thus is simply denoted as dN/dS>1, inferring positive selection. BR-S Pos.Sel.= branch-site positive selection. "-" indicates the dN and dS were each <0.001 and thus had too low divergence to determine dN/dS. The species branch per gene with the highest dN/dS is in **bold**. The connector gene *Paris* (FBgn0031610) was rapidly evolving in Dmel-Dsim but lacked high confidence orthologs in all five species(table S2). Genes that showed positive selection using McDonald and Kreitman (1991)-tests of Dmel-Dsim included FBgn0026379 (*Pten*), FBgn0004858 (*elB*),FBgn0010825 (*Gug*), FBgn0261854 (*aPKC*), FBgn0032006 (*Pvr*). One gene, *Zyx*, was not available for MK tests in the database (Murga-Moreno, et al. 2019).

Table 2. Genes that were highly upregulated in the larval ovary somatic cells relative to germ cells when pooled across three larval stages (Tarikere, et al. 2022) and that exhibited rapid protein sequence divergence in the *melanogaster* subgroup (M0 dN/dS>0.20). The dN/dS per species terminal branch, branch-site positive selection (P<0.05) and *tau* values are shown for each gene. The genes with the top 10 log_2 fold change values matching these criteria are shown.

Fbgn ID	**Log_2 fold change**	**Gene name**	**Gene symbol**	**M0 dN/dS**	**Branch dN/dS**					**BR-S pos. sel. P<0.05**	***tau***
					Dsim	Dsec	Dmel	Dyak	Dere		
FBgn0052581	10.012	*CG32581*	*CG32581*	0.3052	0.1647	**0.6899**	0.5668	0.2455	0.1672	Dmel	0.7378
FBgn0051157	9.389	*CG31157*	*CG31157*	0.2962	0.1163	**1.3228**	0.1967	0.3405	0.1777		0.9010
FBgn0039108	7.526	*CG10232*	*CG10232*	0.7202	>1	**2.0881**	0.4894	0.6514	0.5914	Dere	0.9260
FBgn0039598	7.217	*aquarius*	*aqrs*	0.2305	0.1183	0.1097	0.1265	**0.3029**	0.1972		0.9946
FBgn0260479	5.373	*CG31904*	*CG31904*	0.3038	0.0001	0.4808	**0.6401**	0.0556	0.1363	Dsec, Dmel	0.9466
FBgn0044048	5.343	*Insulin-like peptide 5*	*Ilp5*	0.3776	0.2932	**0.5843**	0.0001	0.4907	0.5501		0.8485
FBgn0031900	5.308	*CG13786*	*CG13786*	0.2487	0.1709	0.2778	0.2748	0.2362	**0.3271**		0.9483
FBgn0050281	5.216	*CG30281*	*CG30281*	0.2155	0.4796	**0.613**	0.1951	0.1806	0.249		0.9820
FBgn0031646	5.146	*snustorr snarlik*	*snsl*	0.2672	0.1571	0.0635	0.1566	0.2317	**0.585**		0.9408
FBgn0051815	5.070	*CG31815*	*CG31815*	0.3745	0.1725	0.3185	0.3421	**0.4465**	0.3695	Dmel	0.9519

Notes: The species branch per gene with the highest dN/dS is in **bold**. A name for FBgn0052581 as *suppression of retinal degeneration disease 1 upon overexpression 2* (*sordd2*) has been recently added/proposed at FlyBase. One gene, *CG10232*, showed positive selection using McDonald and Kreitman (1991)-tests of Dmel-Dsim.

Table 3. Genes with rapid divergence (M0 dN/dS>0.20) and that were highly upregulated at one stage of the larval ovary somatic cells (versus the others; three stages early, mid, late, among the top 30 most upregulated genes, table S5) in Dmel using BULKSG data (Tarikere, et al. 2022) and that also exhibited upregulation in at least one cell type (versus all others) using SINGLEC data among the nine studied LL3 ovary cell types (Slaidina, et al. 2020). Shown are the dN/dS per species branch, the presence of branch-site positive selection (P<0.05), the *tau* values and an example of key functionality as described in DAVID (Huang da, et al. 2009). "Stage up" indicates the larval ovary stage where the gene was upregulated (P<0.05). SINGLEC up indicates the cell type(s) with upregulation.

Fbgn ID	**Gene**	**BULKSG Up**		**SINGLEC Up (Seurat P<0.05)**	**M0 dN/dS**	**Branch dN/dS**					**Branch-site positive selection P<0.05**					***tau***	**Example of key function**
		Stage up	**log$_2$fold change**			**Dsim**	**Dsec**	**Dmel**	**Dyak**	**Dere**	**Dsim**	**Dsec**	**Dmel**	**Dyak**	**Dere**		
FBgn0015872	*Drip*	Late	7.798	TFa, TFb	0.2734	0.6248	**>1**	0.4140	0.3750	0.1312	yes		yes	yes		0.9786	Membrane
FBgn0040343	*CG3713*	Late	5.951	TFa	0.2636	**>1**	0.0001	0.8226	0.1501	0.3225						0.9146	Uncharacterized
FBgn0002868	*MtnA*	Early	4.380	TFp, CC	0.6883	-	**>1**	**>1**	**>1**	0.1265						0.8753	Response to metal ion
FBgn0016075	*vkg*	Late	4.200	TFa, TFp, SHm, CC	0.3860	0.3038	**0.7557**	0.2965	0.3510	0.5572		yes		yes	yes	0.9037	Basement membrane
FBgn0000299	*Col4a1*	Late	3.875	TFa, TFp, SHm, CC	0.4065	0.3034	**1.2519**	0.2072	0.4728	0.6879		yes		yes	yes	0.8887	Basement membrane

Notes: The cell types with upregulation are shown by the following abbreviations TFa=terminal filaments anterior, TFp=terminal filaments posterior, SHm=sheath cells migrating, CC=cap cells. "-" indicates the dN and dS were each <0.001 and thus have too little divergence to be able to determine dN/dS. The species branch with the highest dN/dS is in **bold**. Genes that showed positive selection using McDonald and Kreitman (1991) tests of Dmel-Dsim included FBgn0016075 (*vkg*) and FBgn0000299 *(Col4a1*). The full gene name is *Metallothionein A for MtnA, viking for vkg and Collagen type IV alpha 1 for Col4a1 and Drip and CG3713* are named as shown.

Table 4. PGLS analysis of the relationship between ovariole number and dN/dS for genes putatively involved in ovariole number evolution from tables 1, 3 and 4 (42 genes total). The 17 genes that showed a relationship using PGLS are shown (P<0.05), and includes the intercept, the slope, and the predicted ovariole numbers using the model. In addition, the dataset that each gene was identified from and the table it was presented in are provided.

FBgn ID	Symbol	Dataset	Table with Gene	PGLS P-value	Intercept	Slope	Predicted Ovariole No. Under PGLS Model				
							Dsim	Dsec	Dmel	Dyak	Dere
FBgn0011274	*Dif*	SIGNALC	Table 1	0.0189	38.0298	-25.8738	38.03	22.55	38.03	19.32	27.30
FBgn0003612	*Su(var)2-10*	SIGNALC	Table 1	0.0115	22.0048	28.8688	38.53	22.01	37.83	23.92	22.36
FBgn0011642	*Zyx*	SIGNALC	Table 1	0.0205	35.9756	-14.2358	31.56	14.62	31.88	32.18	31.39
FBgn0004858	*elB*	SIGNALC	Table 1	0.0170	32.7685	-28.2679	32.77	15.36	31.93	30.90	31.02
FBgn0259176	*bun*	SIGNALC	Table 1	0.0316	43.5717	-83.9484	38.02	22.01	35.10	28.34	20.77
FBgn0023540	*CG3630*[a]	SIGNALC	Table 1	0.0689	41.6675	-46.7838	24.90	13.89	35.83	30.49	36.39
FBgn0030904	*upd2*	SIGNALC	Table 1	0.0143	34.3707	-42.2274	34.37	16.77	32.91	31.02	27.33
FBgn0003607	*Su(var)205*	SIGNALC	Table 1	0.0092	18.0230	120.0788	31.26	18.04	43.32	24.99	25.64
FBgn0261592	*RpS6*	SIGNALC	Table 1	0.0261	31.9450	-55.4869	31.94	15.01	30.95	31.94	31.94
FBgn0032006	*Pvr*	SIGNALC	Table 1	0.0162	38.4848	-71.3851	37.94	20.64	34.67	26.96	23.76
FBgn0045035	*tefu*	SIGNALC	Table 1	0.0520	49.0222	-158.1686	35.29	18.84	29.69	32.40	26.70
FBgn0051157	*CG31157*	BULKSG	Table 2	0.0175	34.6446	-14.4238	32.97	15.56	31.81	29.73	32.08
FBgn0044048	*Ilp5*	BULKSG	Table 2	0.0225	44.4030	-40.6225	32.49	20.67	44.40	24.47	22.06
FBgn0015872	*Drip*	BULKSG & SINGLEC	Table 3	0.0474	38.3688	-16.6128	27.99	13.45	31.49	32.14	36.19
FBgn0040343	*CG3713*	BULKSG & SINGLEC	Table 3	0.0470	22.8789	10.9136	39.25	22.88	31.86	24.52	26.40
FBgn0016075	*vkg*	BULKSG & SINGLEC	Table 3	0.0171	45.3114	-37.1647	34.02	17.23	34.29	32.27	24.60
FBgn0000299	*Col4a1*	BULKSG & SINGLEC	Table 3	0.0053	39.2022	-18.3941	33.62	16.17	35.39	30.51	26.55

Notes: [a], included as close to cut-off and P=0.069. The phylogeny is in fig. 2 and branch lengths used for PGLS analyses are in Materials and Methods.

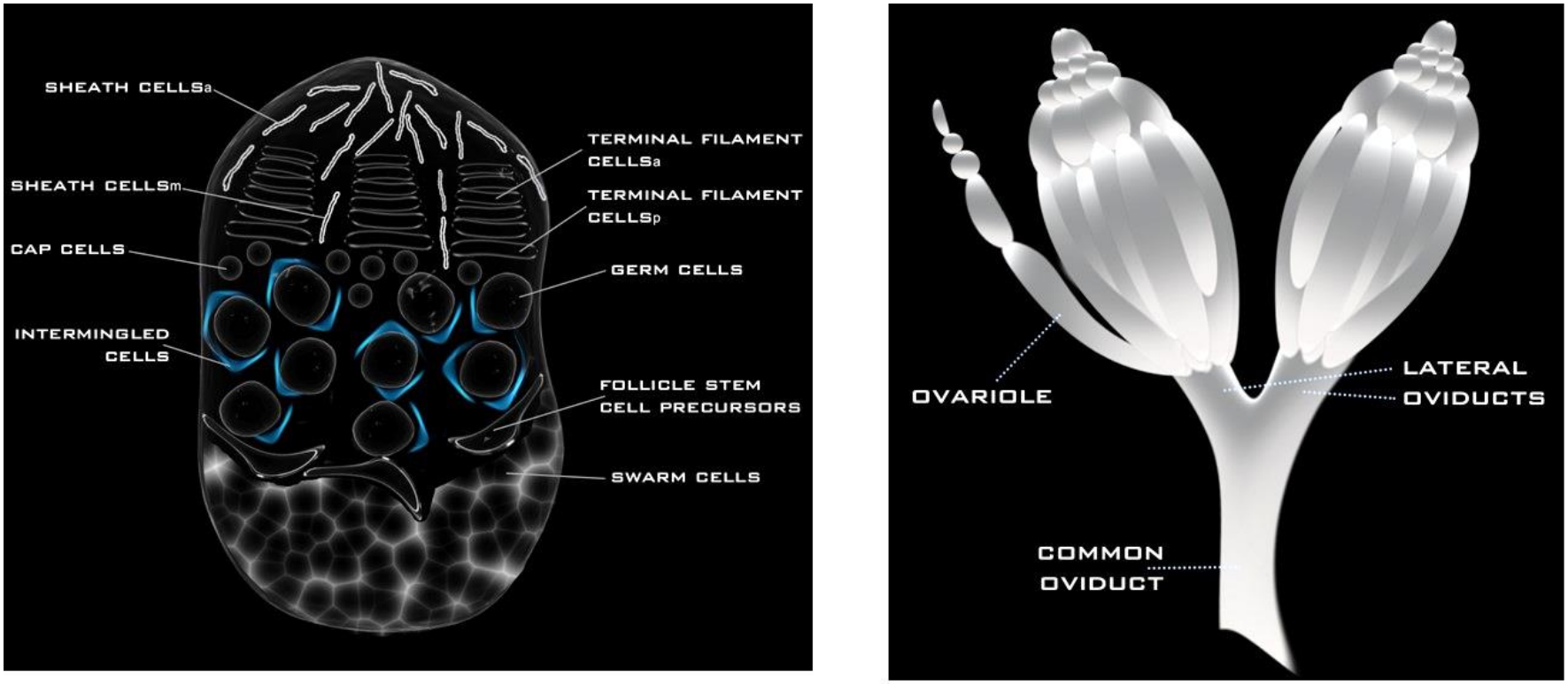


Figure 1. A schematic diagram of A) the late third instar larval ovary with its germ cells and various somatic cell types; and B) an external view of an adult ovary showing the ovarioles in each of the two ovaries that converge to the common oviduct in *D. melanogaster*. The relative cell positioning of cells in panel A is as denoted by Slaidina, et al. (2020). For orientation, anterior is up in both panels.

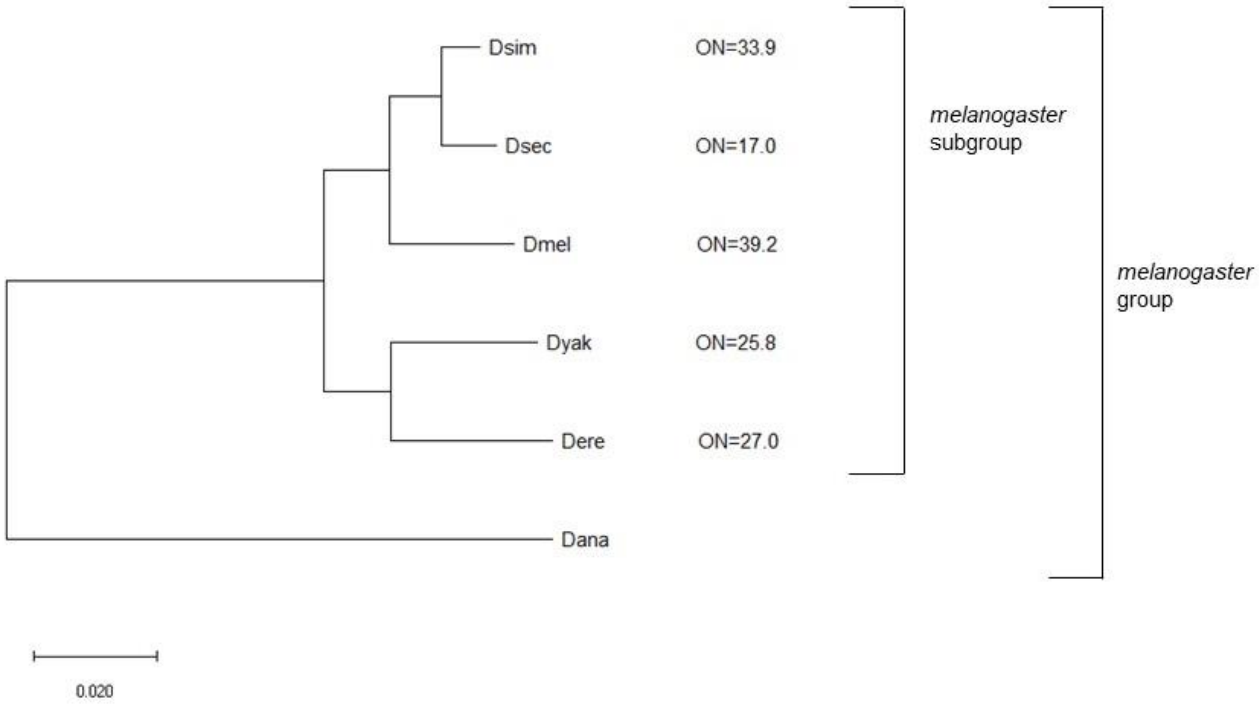
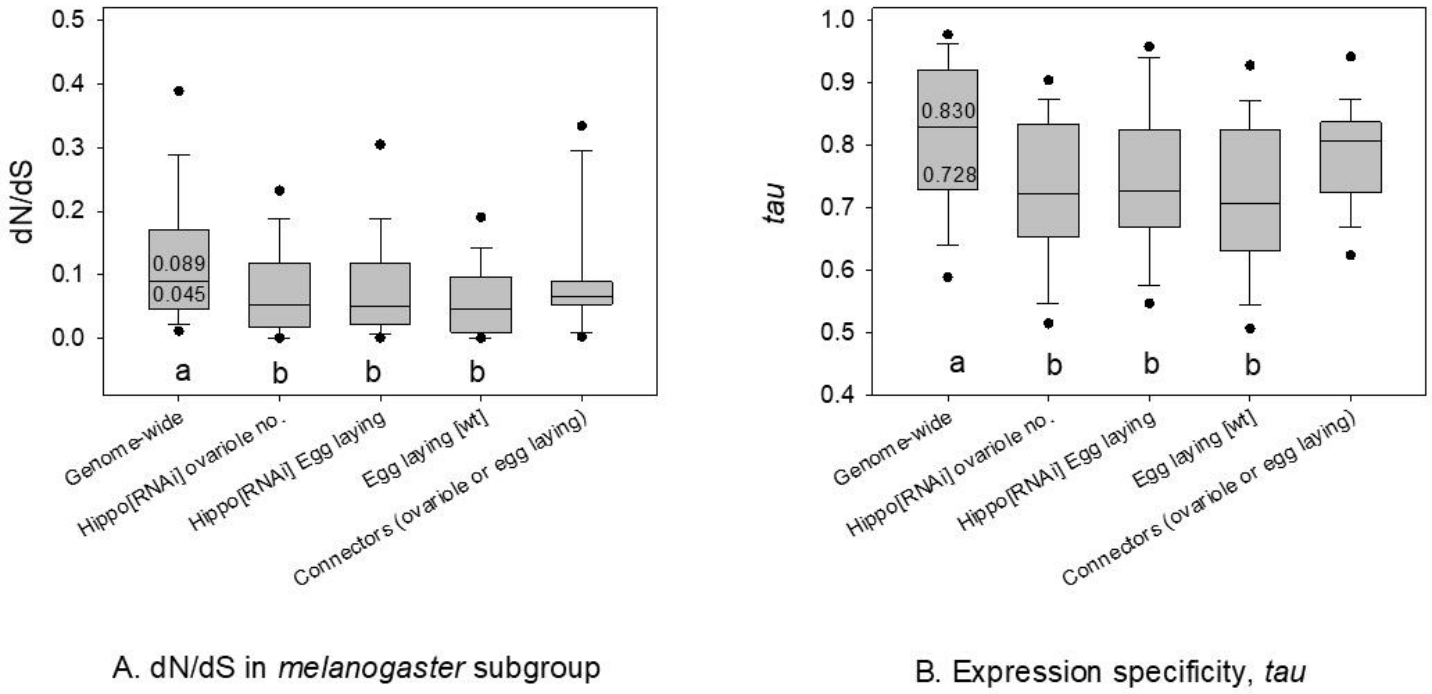


Figure 2. The phylogeny showing the five-species *melanogaster* subgroup under study that was based on a Maximum Likelihood tree generated in MEGA v. 11 (Tamura, et al. 2021) and DNA sequence data from DrosoPhyla (Finet, et al. 2021). The five species of the *melanogaster* subgroup are shown. The relatively distantly related *D. ananassae* (Dana) was used as an outgroup for tree construction. Ovariole numbers (ON) are shown and are for two ovaries per female and are from the following sources: *D. melanogaster* (Dmel), *D. sechellia* (Dsec), and *D. yakuba* (Dyak) (Hodin and Riddiford 2000), *D. simulans* (Dsim) (averaged, (Hodin and Riddiford 2000; Starmer, et al. 2003) and *D. erecta* (Dere) (Markow, et al. 2009) (see respective articles for variation). All nodes had 100/100 bootstrap support.



A. dN/dS in *melanogaster* subgroup

B. Expression specificity, *tau*

Figure. 3. Box plots of A) M0 dN/dS of genes with five-species orthologs in the *melanogaster* subgroup for each of four groups of signalling/connector genes that affected ovariole/egg numbers using RNAi in *D. melanogaster* (Kumar, et al. 2020) and for the genome-wide values; and B) *tau* for all genes in each of the four groups of ovariole number/egg laying affecting genes and the genome-wide values. Different letters (a. b) below bars indicate a statistically significant difference (MWU-tests $P<0.05$) between the genome-wide values and each group of genes. The median and 25th percentiles are shown for dN/dS and *tau* as reference points for the genome-wide values (that is, across all 9,232 genes with known dN/dS and five-species orthologs).

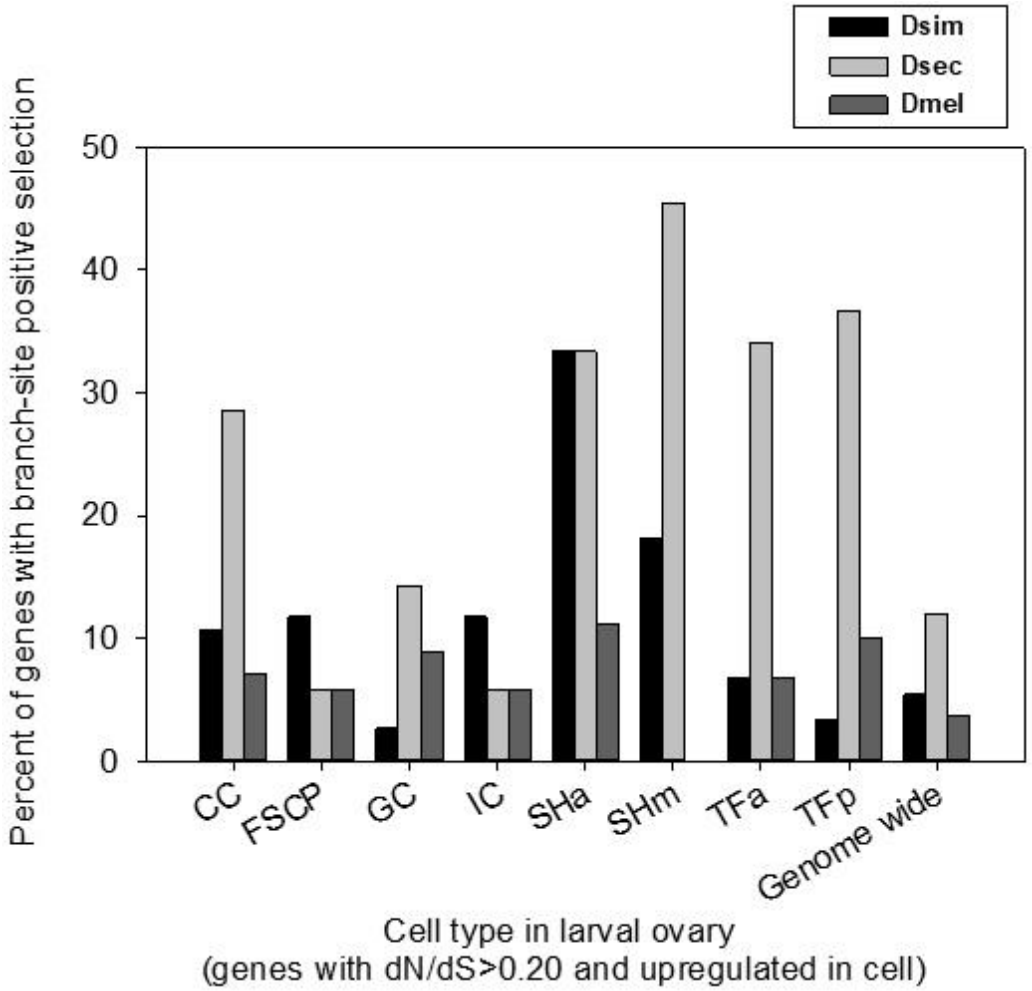


Figure 4. The percentage of the genes that were both upregulated in a particular cell type and rapidly evolving in the *melanogaster* subgroup (M0 dN/dS>0.20) that exhibited branch-site positive selection in the *D. simulans* (Dsim), *D. sechellia* (Dsec), and *D. melanogaster* (Dmel) branches (P<0.05). The number of genes per category were as follows: cap cells (CC: 28), follicle stem cell precursors (FSCP: 17), germ cells (GC: 112), intermingled cells (IC: 17), anterior sheath cells (SHa: 9), migrating sheath cells (SHm: 11), anterior terminal filament cells (TFa: 44), posterior terminal filament cells (TFp: 30). Swarm cells (SW) cells were excluded as too few genes were rapidly evolving for study (SW: 4). Note that a gene could be upregulated in more than one cell type. The genome-wide values are for all genes with five-species orthologs in the *melanogaster* subgroup.

Supplementary Material

Gene protein sequence evolution can predict the rapid divergence of ovariole numbers in *Drosophila*

Carrie A. Whittle, Cassandra G. Extavour

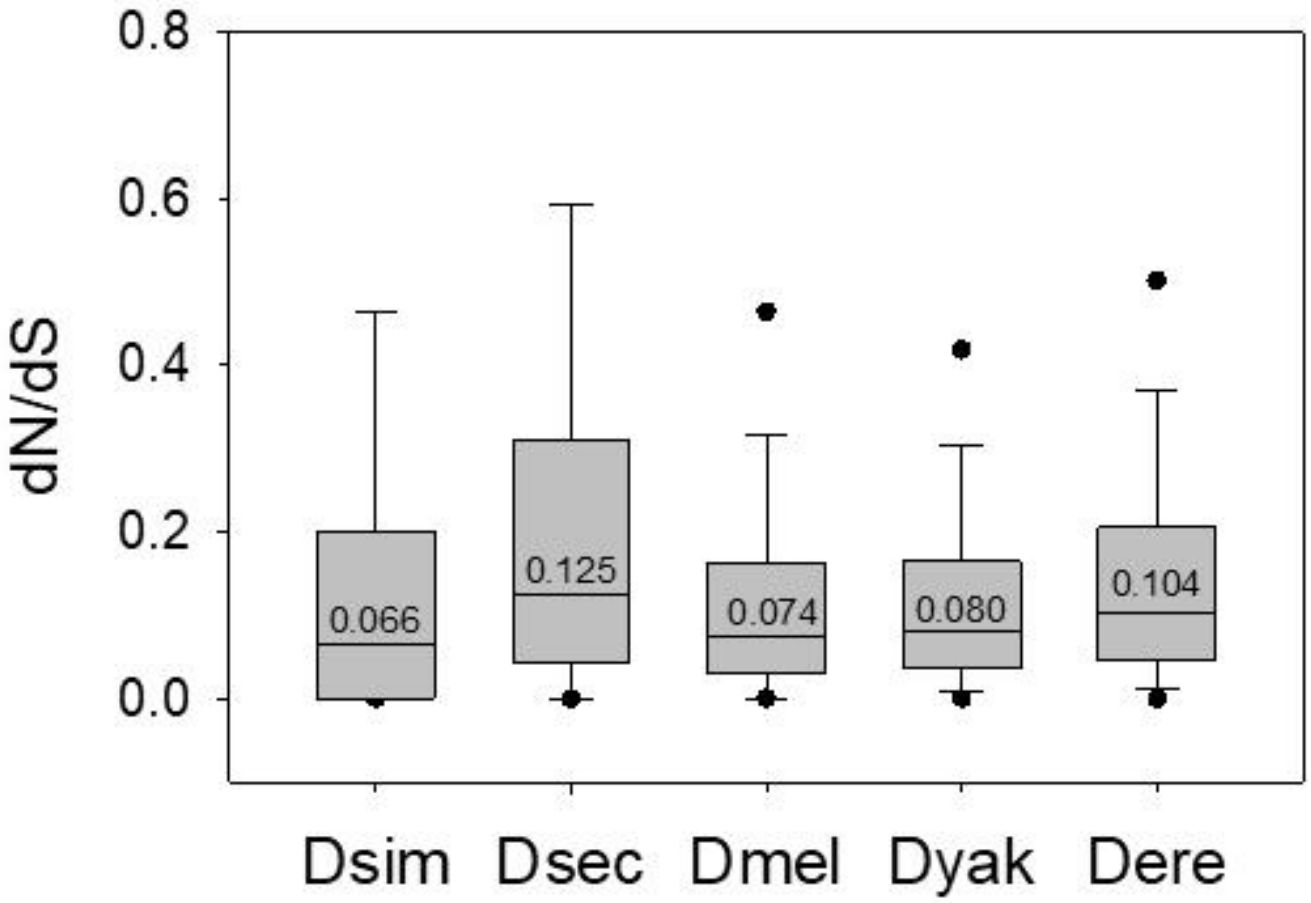


Figure S1. Box plots of the distribution of the free-ratio dN/dS values in each species terminal branch for all genes per genome with five-species orthologs in the *melanogaster* subgroup (Dsim = *Drosophila simulans*; Dsec = *D. sechellia*, Dmel = *D. melanogaster*, Dyak = *D. yakuba*, Dere = *D. erecta*). The median dN/dS value is shown within each box. Outliers are excluded for Dsim and Dsec as they were outside the range of the Y axis.

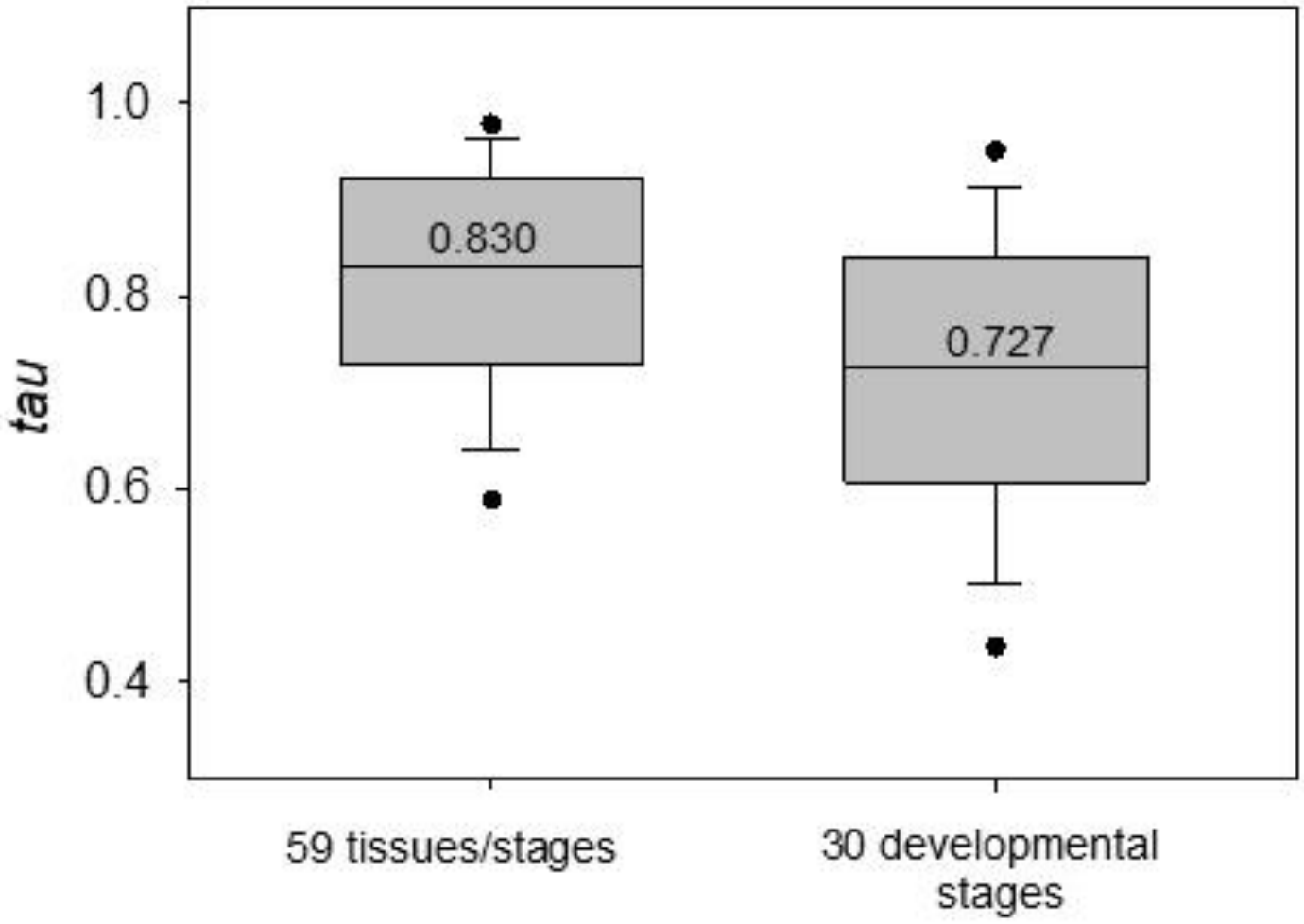


Figure S2. Box plots of the *tau* values across all studied genes in *D. melanogaster* using expression data across 59 tissues and developmental stages (combined) and solely using the data from 30 developmental stages. Values are for all genes with orthologs among the five studied species in the *melanogaster* subgroup. Median values are shown within bars. Tissues and stages are described in table S1.

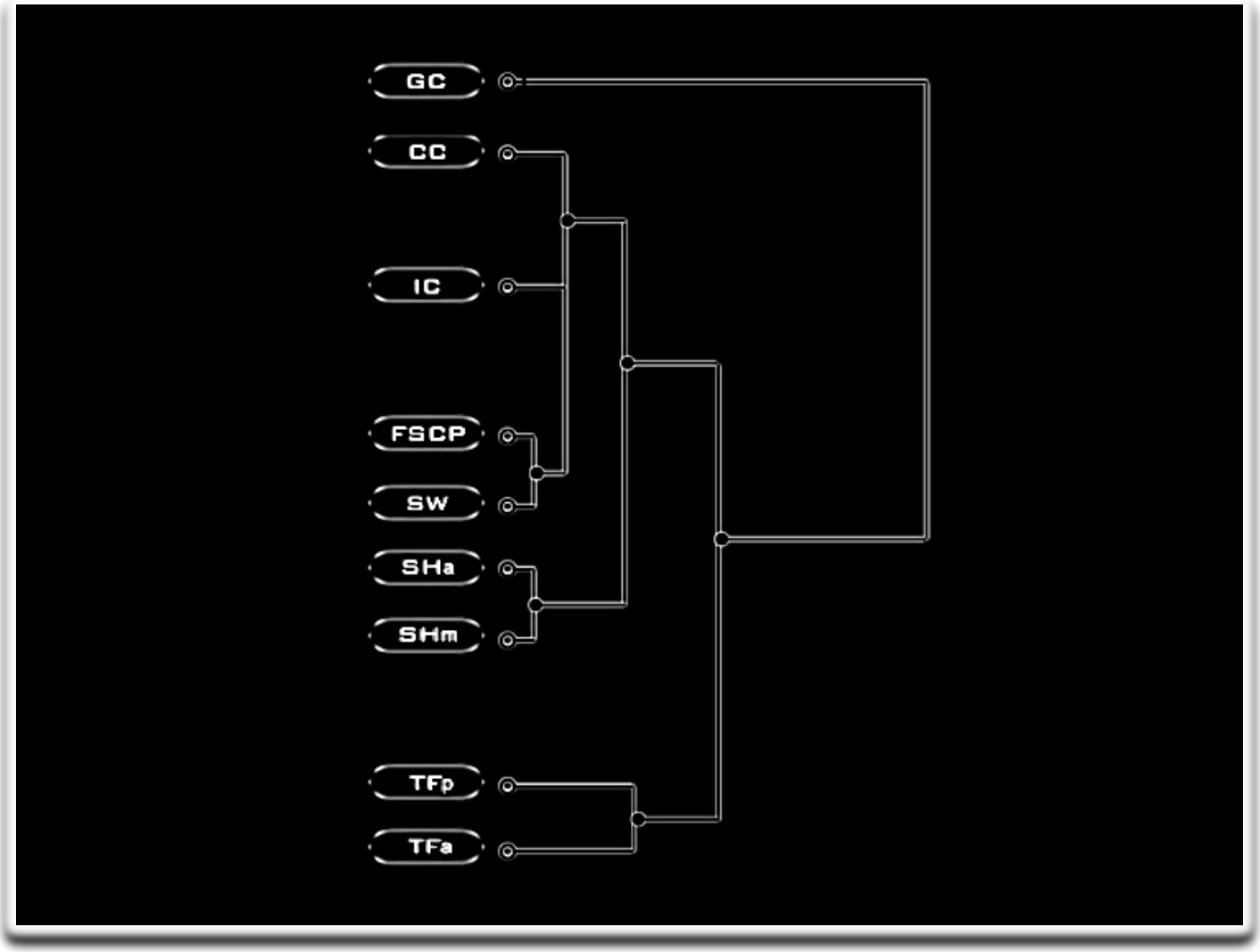


Figure S3. Clustering of the average standardized expression of all genes in the *D. melanogaster* genome (with non-zero expression) in the various cell types obtained from the study of LL3 larval ovaries in Slaidina, et al. (2020). The expression was clustered using hierarchical clustering and average linkage in the Morpheus program (see Materials and Methods). CC=cap cells, FSCP=follicle stem cell precursors, GC=germ cells, IC=intermingled cells, SHa= anterior sheath cells, SHm= migrating sheath cells, SW=swarm cells, TFa= anterior terminal filament cells, TFp= posterior terminal filament cells.

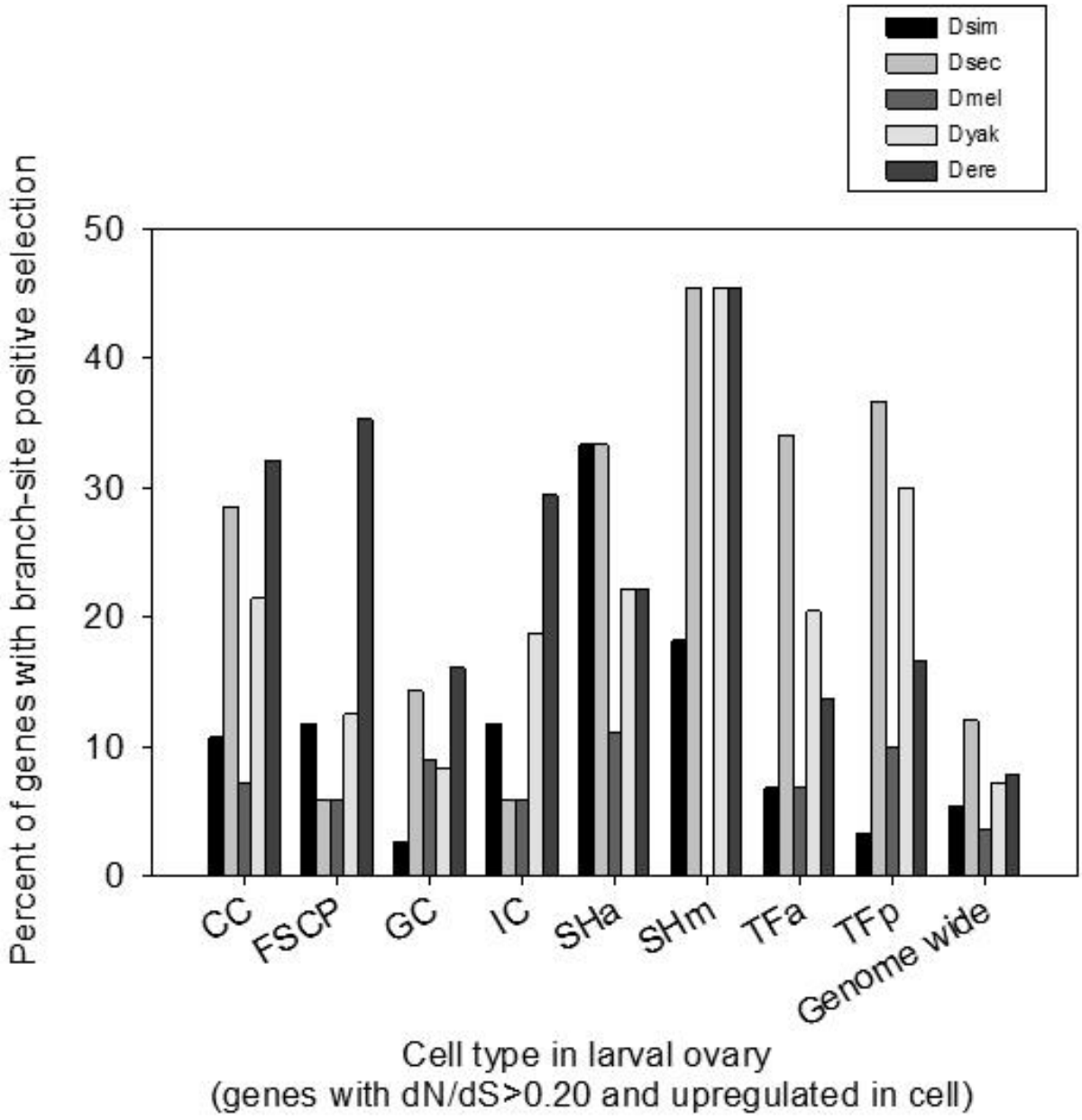


Figure S4. The percentage of the genes upregulated in a particular cell type using *D. melanogaster* sc-RNA seq(Slaidina, et al. 2020) and rapidly evolving in the *melanogaster* subgroup (M0 dN/dS>0.20) that exhibited branch-site positive selection in the *D. simulans* (Dsim), *D. sechellia (*Dsec), *D. melanogaster* (Dmel), *D. yakuba* (Dyak) or *D. erecta* (Dere) branches. The number of genes per category were as follows: CC (28), FSCP (17), GC (112), IC (17), SHa (9), SHm (11), SW (4), TFa (44), TFp (30). SW cells were excluded as too few genes were rapidly evolving for study. Note that a gene could be upregulated in more than one cell type. The genome-wide values are for all genes with orthologs in the *melanogaster* subgroup.

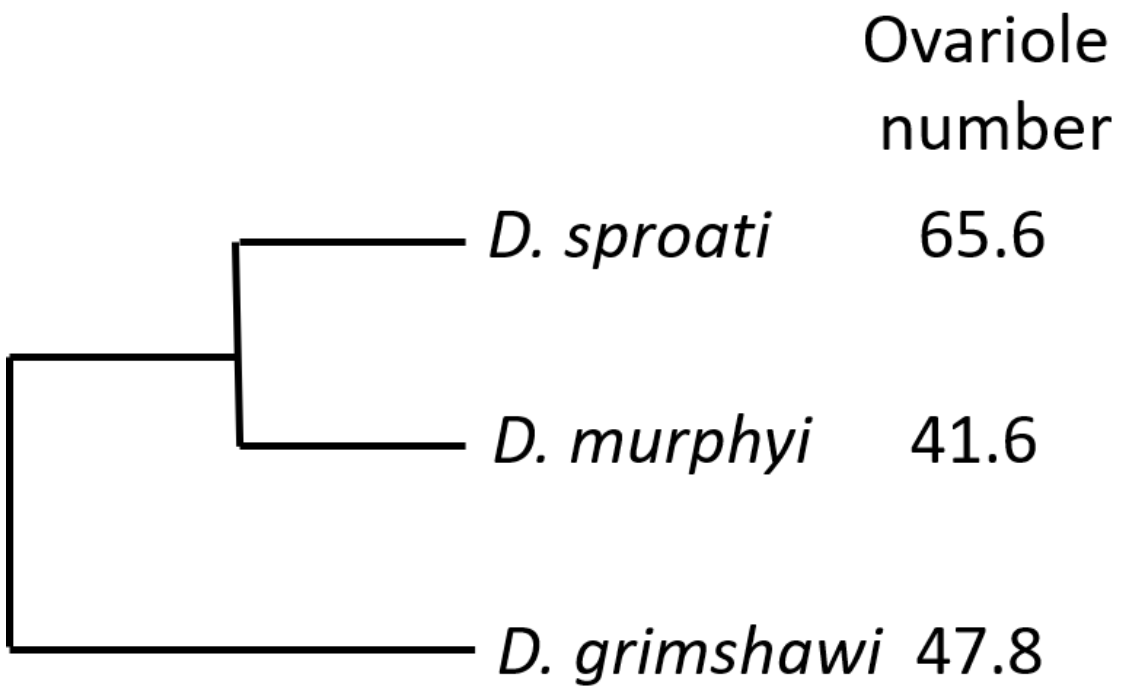


Figure S5. The phylogenetic relationship of the three Hawaiian *Drosophila* species under study and their mean ovariole numbers per female (phylogeny from (Kim, et al. 2021; Suvorov, et al. 2022), ovariole numbers from (Starmer, et al. 2003; Sarikaya, et al. 2019).

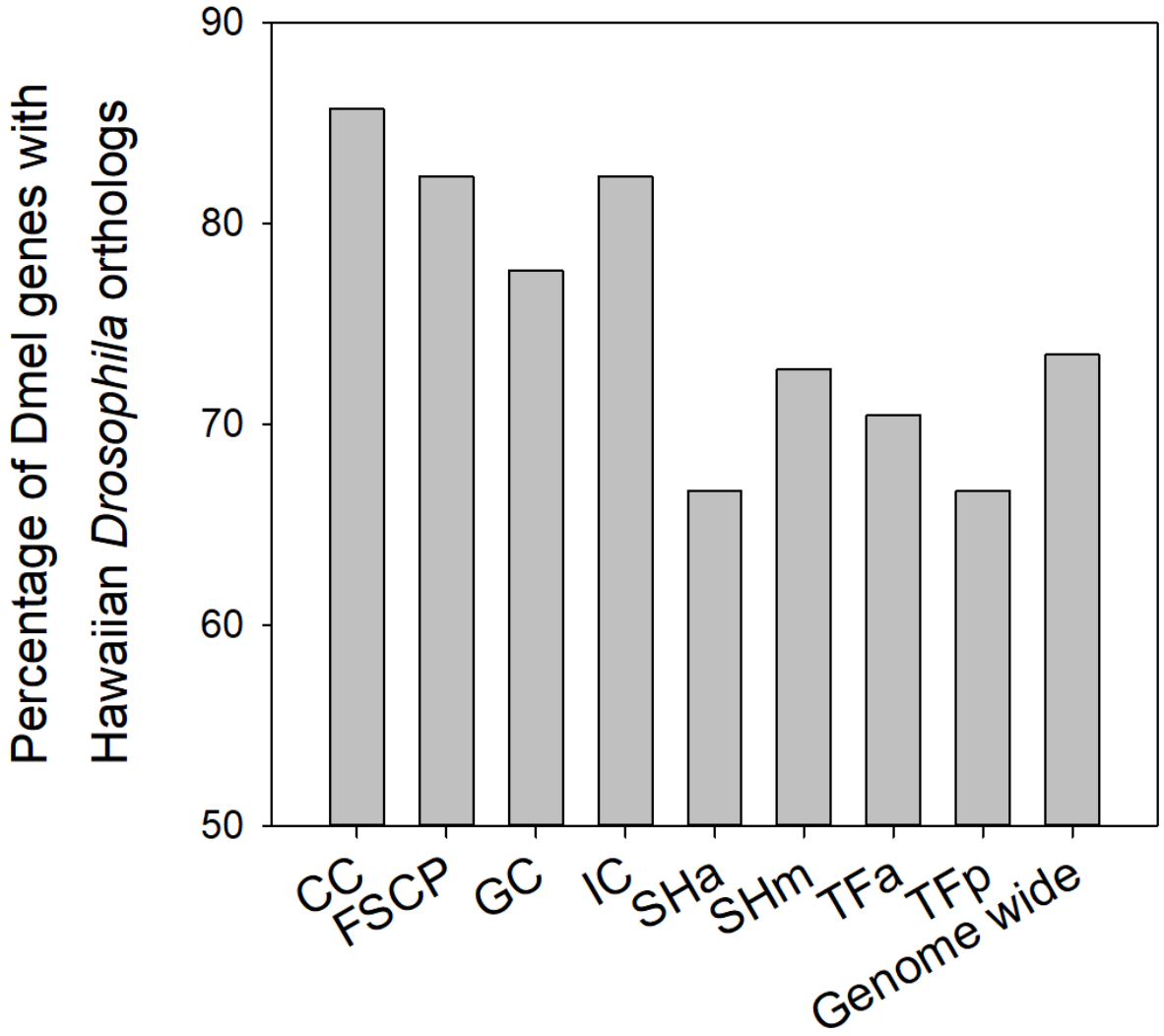


Figure S6. The percent of studied *D. melanogaster* sc-RNA seq genes upregulated and rapidly evolving in *D. melanogaster* that had three-species orthologs in Hawaiian *Drosophila*. N values in *D. melanogaster* are in Fig. 4. To be defined as having an orthologous gene set, the *D. melanogaster* gene had to have a match to the annotated genome *D. grimshawi*, and the three Hawaiian species had to all have high confidence orthologs to each other.

Table S1. The 30 developmental stages and 29 tissue types used for expression breadth analysis, or *tau*. The RPKM data is from ModEncode (Li, et al. 2014) and is available at FlyBase (Gramates, et al. 2022). Downloaded in April 2022.

Developmental stages	**Tissue types**
em0-2hr	A_MateF_1d_head
em2-4hr	A_MateF_4d_ovary
em4-6hr	A_MateM_1d_head
em6-8hr	A_VirF_1d_head
em8-10hr	A_VirF_4d_head
em10-12hr	A_MateF_20d_head
em12-14hr	A_MateF_4d_head
em14-16hr	A_MateM_20d_head
em16-18hr	A_MateM_4d_acc_gland
em18-20hr	A_MateM_4d_head
em20-22hr	A_MateM_4d_testis
em22-24hr	A_1d_carcass
L1	A_1d_dig_sys
L2	A_20d_carcass
L3_12hr	A_20d_dig_sys
L3_PS1-2	A_4d_carcass
L3_PS3-6	A_4d_dig_sys
L3_PS7-9	P8_CNS
WPP	L3_CNS
P5	L3_Wand_carcass
P6	L3_Wand_dig_sys
P8	L3_Wand_fat
P9-10	L3_Wand_imag_disc
P15	L3_Wand_saliv
AdF_Ecl_1days	A_VirF_20d_head
AdF_Ecl_5days	A_VirF_4d_ovary
AdF_Ecl_30days	WPP_fat
AdM_Ecl_1days	WPP_saliv
AdM_Ecl_5days	P8_fat
AdM_Ecl_30days	

Table S2. The 28 genes (27 with five-species orthologs for further follow-up study) that were identified herein as having a high dN/dS value (≥1.5 fold higher than genome-wide median) based on assessment of the gene sets that affected ovariole numbers and/or egg laying in *D. melanogaster* based on RNAi in a *hpo* loss of function genetic background from Kumar, et al. (2020). The observed phenotypes were named as *hpo[RNAi]* Ovariole number, *hpo[RNAi]* Egg laying and Egg laying *[wt]*, or as "connectors". Results are shown for dN/dS obtained for *D. melanogaster-D. simulans* (Dmel-Dsim) (genome-wide median=0.086), the *melanogaster* subgroup (median=0.089), and the *melanogaster* group (0.062) using data available from (Stanley and Kulathinal 2016) and the M0 model of PAML that provides a single dN/dS per alignment (Yang 2007). For completeness, all details for genes that belong to more than one of the four phenotype gene sets described above, and those with dN/dS in more than one of the three sets of *Drosophila* species are shown. Genes are listed in descending order with respect to the M0 dN/dS of the highest value per gene. Occasional genes near the cut-offs were retained for analysis. These genes were subjected to follow-up molecular evolutionary analyses.

Fbgn ID	CG number	Name	Symbol	dN/dS	Species set for dN/dS	Gene phenotype observed/Status
FBgn0011274	CG6794	*Dorsal-related immunity factor*	*Dif*	0.475	*melanogaster* subgroup	*hpo[RNAi]* Egg laying
FBgn0011274	CG6794	*Dorsal-related immunity factor*	*Dif*	0.311	*melanogaster* group	*hpo[RNAi]* Egg laying
FBgn0014020	CG8416	*Rho1*	*Rho1*	0.418	Dmel-Dsim	*hpo[RNAi]* Ovariole number
FBgn0014020	CG8416	*Rho1*	*Rho1*	0.418	Dmel-Dsim	*hpo[RNAi]* Egg laying
FBgn0014020	CG8416	*Rho1*	*Rho1*	0.418	Dmel-Dsim	Egg laying [wt]
FBgn0003612	CG8068	*Suppressor of variegation 2-10*	*Su(var)2-10*	0.387	Dmel-Dsim	*hpo[RNAi]* Ovariole number
FBgn0003612	CG8068	*Suppressor of variegation 2-10*	*Su(var)2-10*	0.387	Dmel-Dsim	*hpo[RNAi]* Egg laying
FBgn0003612	CG8068	*Suppressor of variegation 2-10*	*Su(var)2-10*	0.387	Dmel-Dsim	Egg laying [wt]
FBgn0026379	CG5671	*Phosphatase and tensin homolog*	*Pten*	0.360	Dmel-Dsim	*hpo[RNAi]* Ovariole number
FBgn0026379	CG5671	*Phosphatase and tensin homolog*	*Pten*	0.289	*melanogaster* subgroup	*hpo[RNAi]* Ovariole number
FBgn0026379	CG5671	*Phosphatase and tensin homolog*	*Pten*	0.232	*melanogaster* group	*hpo[RNAi]* Ovariole number
FBgn0000259	CG15224	*Casein kinase II beta subunit*	*CkIIbeta*	0.359	Dmel-Dsim	*hpo[RNAi]* Ovariole number
FBgn0000259	CG15224	*Casein kinase II beta subunit*	*CkIIbeta*	0.359	Dmel-Dsim	*hpo[RNAi]* Egg laying
FBgn0000259	CG15224	*Casein kinase II beta subunit*	*CkIIbeta*	0.359	Dmel-Dsim	Egg laying [wt]
FBgn0000259	CG15224	*Casein kinase II beta subunit*	*CkIIbeta*	0.195	*melanogaster* subgroup	*hpo[RNAi]* Ovariole number
FBgn0000259	CG15224	*Casein kinase II beta subunit*	*CkIIbeta*	0.195	*melanogaster* subgroup	*hpo[RNAi]* Egg laying

FBgn0000259	CG15224	*Casein kinase II beta subunit*	*CkIIbeta*	0.195	*melanogaster* subgroup	Egg laying [wt]
FBgn0035213	CG2199	-	*CG2199*	0.355	Dmel-Dsim	Connector
FBgn0035213	CG2199	-	*CG2199*	0.334	*melanogaster* subgroup	Connector
FBgn0011642	CG32018	*Zyxin*	*Zyx*	0.317	*melanogaster* subgroup	*hpo[RNAi]* Egg laying
FBgn0011642	CG32018	*Zyxin*	*Zyx*	0.290	Dmel-Dsim	*hpo[RNAi]* Egg laying
FBgn0011642	CG32018	*Zyxin*	*Zyx*	0.127	*melanogaster* group	*hpo[RNAi]* Egg laying
FBgn0262614	CG43140	*polychaetoid*	*pyd*	0.316	Dmel-Dsim	*hpo[RNAi]* Ovariole number
FBgn0031610	CG15436	*Parkin Interacting Substrate*	*Paris*	0.290	Dmel-Dsim	Connector
FBgn0036974	CG5605	*eukaryotic translation release factor 1*	*eRF1*	0.267	Dmel-Dsim	*hpo[RNAi]* Ovariole number
FBgn0036974	CG5605	*eukaryotic translation release factor 1*	*eRF1*	0.267	Dmel-Dsim	*hpo[RNAi]* Egg laying
FBgn0036974	CG5605	*eukaryotic translation release factor 1*	*eRF1*	0.267	Dmel-Dsim	Egg laying [wt]
FBgn0003984	CG10491	*vein*	*vn*	0.258	*melanogaster* subgroup	*hpo[RNAi]* Ovariole number
FBgn0003984	CG10491	*vein*	*vn*	0.221	*melanogaster* group	*hpo[RNAi]* Ovariole number
FBgn0004858	CG4220	*elbow B*	*elB*	0.211	*melanogaster* subgroup	*hpo[RNAi]* Ovariole number
FBgn0004858	CG4220	*elbow B*	*elB*	0.138	*melanogaster* group	*hpo[RNAi]* Ovariole number
FBgn0010825	CG6964	*Grunge*	*Gug*	0.211	Dmel-Dsim	*hpo[RNAi]* Ovariole number
FBgn0010825	CG6964	*Grunge*	*Gug*	0.211	Dmel-Dsim	*hpo[RNAi]* Egg laying
FBgn0010825	CG6964	*Grunge*	*Gug*	0.211	Dmel-Dsim	Egg laying [wt]
FBgn0002174	CG5504	-	*CG5504*	0.191	*melanogaster* subgroup	*hpo[RNAi]* Ovariole number
FBgn0002174	CG5504	-	*CG5504*	0.124	*melanogaster* group	*hpo[RNAi]* Ovariole number
FBgn0037218	CG1107	*auxilin*	*aux*	0.190	*melanogaster* subgroup	*hpo[RNAi]* Egg laying

FBgn0037218	CG1107	*auxilin*	*aux*	0.190	*melanogaster* subgroup	Egg laying [wt]
FBgn0037218	CG1107	*auxilin*	*aux*	0.157	Dmel-Dsim	*hpo[RNAi]* Egg laying
FBgn0037218	CG1107	*auxilin*	*aux*	0.157	Dmel-Dsim	Egg laying [wt]
FBgn0037218	CG1107	*auxilin*	*aux*	0.099	*melanogaster* group	*hpo[RNAi]* Egg laying
FBgn0037218	CG1107	*auxilin*	*aux*	0.099	*melanogaster* group	Egg laying [wt]
FBgn0259176	CG42281	*bunched*	*bun*	0.168	*melanogaster* subgroup	*hpo[RNAi]* Ovariole number
FBgn0259176	CG42281	*bunched*	*bun*	0.136	Dmel-Dsim	*hpo[RNAi]* Ovariole number
FBgn0023540	CG3630	-	*CG3630*	0.159	Dmel-Dsim	Connector
FBgn0023540	CG3630	-	*CG3630*	0.138	*melanogaster* subgroup	Connector
FBgn0023540	CG3630	-	*CG3630*	0.098	*melanogaster* group	Connector
FBgn0261854	CG42783	*atypical protein kinase C*	*aPKC*	0.184	*melanogaster* subgroup	*hpo[RNAi]* Egg laying
FBgn0261854	CG42783	*atypical protein kinase C*	*aPKC*	0.098	*melanogaster* group	*hpo[RNAi]* Egg laying
FBgn0001169	CG5460	*Hairless*	*H*	0.155	*melanogaster* subgroup	*hpo[RNAi]* Ovariole number
FBgn0001169	CG5460	*Hairless*	*H*	0.149	Dmel-Dsim	*hpo[RNAi]* Ovariole number
FBgn0001169	CG5460	*Hairless*	*H*	0.148	*melanogaster* group	*hpo[RNAi]* Ovariole number
FBgn0024291	CG5216	*Sirtuin 1*	*Sirt1*	0.153	Dmel-Dsim	*hpo[RNAi]* Egg laying
FBgn0024291	CG5216	*Sirtuin 1*	*Sirt1*	0.153	Dmel-Dsim	Egg laying [wt]
FBgn0030904	CG5988	*unpaired 2*	*upd2*	0.152	*melanogaster* subgroup	*hpo[RNAi]* Ovariole number
FBgn0030904	CG5988	*unpaired 2*	*upd2*	0.152	*melanogaster* subgroup	*hpo[RNAi]* Egg laying
FBgn0030904	CG5988	*unpaired 2*	*upd2*	0.152	*melanogaster* subgroup	Egg laying [wt]
FBgn0030904	CG5988	*unpaired 2*	*upd2*	0.114	*melanogaster* group	*hpo[RNAi]* Ovariole number
FBgn0030904	CG5988	*unpaired 2*	*upd2*	0.114	*melanogaster* group	*hpo[RNAi]* Egg laying
FBgn0030904	CG5988	*unpaired 2*	*upd2*	0.114	*melanogaster* group	Egg laying [wt]
FBgn0020496	CG7583	*C-terminal Binding Protein*	*CtBP*	0.148	Dmel-Dsim	*hpo[RNAi]* Ovariole number

FBgn0020496	CG7583	*C-terminal Binding Protein*	*CtBP*	0.148	Dmel-Dsim	*hpo[RNAi]* Egg laying
FBgn0020496	CG7583	*C-terminal Binding Protein*	*CtBP*	0.148	Dmel-Dsim	Egg laying [wt]
FBgn0003607	CG8409	*Suppressor of variegation 205*	*Su(var)205*	0.144	Dmel-Dsim	Connector
FBgn0261592	CG10944	*Ribosomal protein S6*	*RpS6*	0.139	*melanogaster* subgroup	*hpo[RNAi]* Ovariole number
FBgn0261592	CG10944	*Ribosomal protein S6*	*RpS6*	0.139	*melanogaster* subgroup	*hpo[RNAi]* Egg laying
FBgn0261592	CG10944	*Ribosomal protein S6*	*RpS6*	0.139	*melanogaster* subgroup	Egg laying [wt]
FBgn0020386	CG1210	*Phosphoinositide-dependent kinase 1*	*Pdk1*	0.138	Dmel-Dsim	*hpo[RNAi]* Ovariole number
FBgn0002592	CG6104	*Enhancer of split m2, Bearded family member*	*E(spl)m2-BFM*	0.133	*melanogaster* subgroup	*hpo[RNAi]* Ovariole number
FBgn0002592	CG6104	*Enhancer of split m2, Bearded family member*	*E(spl)m2-BFM*	0.110	*melanogaster* group	*hpo[RNAi]* Ovariole number
FBgn0032006	CG8222	*PDGF- and VEGF-receptor related*	*Pvr*	0.129	*melanogaster* subgroup	*hpo[RNAi]* Egg laying
FBgn0032006	CG8222	*PDGF- and VEGF-receptor related*	*Pvr*	0.117	*melanogaster* group	*hpo[RNAi]* Egg laying
FBgn0045035	CG6535	*telomere fusion*	*tefu*	0.107	*melanogaster* group	*hpo[RNAi]* Ovariole number
FBgn0045035	CG6535	*telomere fusion*	*tefu*	0.107	*melanogaster* group	Egg laying [wt]

Table S3. The 27 rapidly evolving signalling and connector genes under study. For each gene, the associated pathways, *tau* and key functional terms at FlyBase (Gramates, et al. 2017) are shown. From calculation of *tau*, the $\hat{x}$ for adult virgin ovaries is shown as an example of this parameter (values of 1 underlined). Genes are listed in descending order from the highest M0 dN/dS values from PAML (Table S2). Pathways are from Kumar, et al. (2020)

Fbgn ID	**Gene Symbol**	**Pathways**	***Tau* 59-All**	$\hat{x}$ **Ovary virgin**	**Key functional terms at FlyBase**
FBgn0011274	*Dif*	Toll	0.717	0.027	Bacterial response
FBgn0014020	*Rho1*	JNK, EGF, Wnt, TGFβ	0.669	0.266	Actin, cytoskeleton
FBgn0003612	*Su(var)2-10*	JAK/STAT	0.786	<u>1.000</u>	Chromosome structure and function
FBgn0026379	*Pten*	mTOR, FOXO	0.663	<u>1.000</u>	Controlling cytoskeletal rearrangements.
FBgn0000259	*CkIIbeta*	Wnt	0.825	0.356	Functions in oogenesis, neurogenesis
FBgn0035213	*CG2199*	Connector	0.808	<u>1.000</u>	Unknown, transcriptional regulation
FBgn0011642	*Zyx*	Hippo	0.647	0.220	Actin cytoskeleton regulator
FBgn0262614	*pyd*	JNK	0.714	0.233	Cytoskeleton
FBgn0036974	*eRF1*	SHH	0.567	0.564	Termination of nascent peptide synthesis
FBgn0003984	*vn*	EGF, Hippo	0.740	0.091	Growth and patterning of tissues
FBgn0004858	*elB*	Notch	0.739	0.042	Development, cell proliferation
FBgn0010825	*Gug*	Hippo	0.706	0.886	Normal developmental patterning
FBgn0002174	*CG5504*	SHH	0.843	0.129	Tumor suppressor, larvae
FBgn0037218	*aux*	Notch	0.674	<u>1.000</u>	Sperm individualization and neuron death
FBgn0259176	*bun*	TGF B	0.606	0.541	Eye development and oogenesis
FBgn0023540	*CG3630*	Connector	0.856	0.097	Actin binding activity, cytoskeleton, Rho activator
FBgn0261854	*aPKC*	Hippo	0.679	0.395	Neuroblast proliferation and self-renewal
FBgn0001169	*H*	Notch	0.829	0.800	Imaginal development
FBgn0024291	*Sirt1*	FOXO	0.671	0.922	Gene silencing, apoptosis, development
FBgn0030904	*upd2*	Hippo, JAK/STAT	0.962	0.000	Development
FBgn0020496	*CtBP*	Wnt, Notch	0.670	0.584	Embryonic segmentation

FBgn0003607	*Su(var)205*	Connector	0.831	0.403	Gene repression, epigenetic repression
FBgn0261592	*RpS6*	mTOR	0.658	0.211	Component of small (40S) ribosomal subunit.
FBgn0020386	*Pdk1*	mTOR, FOXO	0.565	0.857	Embryo development, inhibits apoptosis, spermatogenesis
FBgn0002592	*E(spl)m2-BFM*	Notch	0.871	0.000	Development
FBgn0032006	*Pvr*	VEGF	0.768	0.034	Cell migration regulation
FBgn0045035	*tefu*	FOXO	0.722	0.889	Development, reproduction

Notes: The *Paris* connector gene had *tau*=0.798, $\hat{x}$=1.0 and functions in the negative regulation of transcription.

Table S4. The 27 rapidly evolving SIGNALC genes identified from Kumar, et al. (2020) and their expression status in the soma and germ cells in the larval ovary (each pooled across stages), and among somatic cells at the early, mid and late stages (DeSeq2 P<0.01 (Tarikere, et al. 2022)). Note that if a gene is designated as upregulated in the germ cells, this automatically indicates it is downregulated in soma. A total of 25 of 27 genes showed differential expression using at least one of these comparisons.

Fbgn ID	**Gene symbol**	**Upregulation observed** (Tarikere, et al. 2022)	
		Somatic versus germ cells	Somatic cells-stage
FBgn0011274	*Dif*	Soma	Early
FBgn0014020	*Rho1*	Soma	-
FBgn0003612	*Su(var)2-10*	-	Late
FBgn0026379	*Pten*	-	Late
FBgn0000259	*CkIIbeta*	-	Late
FBgn0035213	*CG2199*	-	-
FBgn0011642	*Zyx*	-	Late
FBgn0262614	*Pyd*	Soma	Late
FBgn0036974	*eRF1*	Soma	-
FBgn0003984	*vn*	Soma	-
FBgn0004858	*elB*	Soma	Early
FBgn0010825	*Gug*	Soma	-
FBgn0002174	*CG5504*	-	Early
FBgn0037218	*aux*	-	-
FBgn0259176	*bun*	Soma	-
FBgn0023540	*CG3630*	Germ	-
FBgn0261854	*aPKC*	Soma	-
FBgn0001169	*H*	-	Late
FBgn0024291	*Sirt1*	-	Late
FBgn0030904	*upd2*	Soma	-
FBgn0020496	*CtBP*	-	Late
FBgn0003607	*Su(var)205*	Soma	-
FBgn0261592	*RpS6*	-	Early
FBgn0020386	*Pdk1*	-	Late
FBgn0002592	*E(spl)m2-BFM*	Soma	-
FBgn0032006	*Pvr*	Soma	Late
FBgn0045035	*Tefu*	Germ	-

Notes: the gene *Paris* (FBgn0031610), that was rapidly evolving in Dmel-Dsim but lacked high confidence orthologs in all five species was reported as upregulated in the germ cells relative to the soma cells in (Tarikere, et al. 2022).

Table S5. Genes that were highly upregulated in germ cells of the *D. melanogaster* larval ovary (relative to somatic cells) pooled across three larval stages (Tarikere, et al. 2022) and that exhibited rapid divergence in the *melanogaster* subgroup (M0 dN/dS>0.20). Free-ratio dN/dS per branch, branch-site positive selection and *tau* values are shown for each gene. The genes with the top 10 log_2 fold change values matching these criteria are shown. Upregulation in germ cells implies downregulation in the somatic cells.

Fbgn ID	Log_2 Fold Change	Gene name	Gene Symbol	M0 dN/dS	Free-ratio dN/dS					Branch-site positive selection P<0.05					*tau*
					Dsim	Dsec	Dmel	Dyak	Dere	Dsim	Dsec	Dmel	Dyak	Dere	
FBgn0031946	27.531	*CG7164*	*CG7164*	0.2652	0.0001	0.2922	0.2698	0.269	0.5347						0.9678
FBgn0031620	10.754	*CG11929*	*CG11929*	0.4818	0.8017	0.3598	0.4851	0.3439	0.5503					yes	0.9407
FBgn0032375	10.728	*CG14932*	*CG14932*	0.6195	>1	0.7898	2.0139	0.2495	0.4939		yes	yes			0.9462
FBgn0047199	10.422	*CG31517*	*CG31517*	1.4448	-	>1	>1	>1	0.5721						0.8702
FBgn0034838	10.363	*CG12782*	*CG12782*	0.2148	0.4158	0.3893	0.1321	0.2894	0.2932						0.9753
FBgn0051475	10.354	*CG31475*	*CG31475*	0.2820	0.1872	0.4101	0.0408	0.1468	0.0676		yes		yes		0.8798
FBgn0034839	10.218	*CG13540*	*CG13540*	0.6728	0.3081	>1	0.7091	102.6503	0.3557				yes		0.9307
FBgn0051619	10.150	*no long nerve cord*	*nolo*	0.2090	0.327	1.0443	0.2099	0.2221	0.1103	yes			yes		0.9218
FBgn0003009	9.910	*orientation disruptor*	*ord*	0.2084	0.0285	0.2861	0.1592	0.2163	0.2876		yes			yes	0.9138
FBgn0039343	9.846	*CG5111*	*CG5111*	0.2013	0.138	0.1722	0.1472	0.1256	0.1282						0.9522

Notes: "-" indicates the dN and dS were each <0.001 and thus too low divergence to determine dN/dS.

Table S6. The genes that were highly upregulated in the somatic cells of the *D. melanogaster* larval ovary in one of three stages of development (early, mid or late) as defined by Tarikere, et al. (2022) and that had elevated M0 dN/dS values in the *melanogaster* subgroup (>0.20). All differentially expressed genes with M0 dN/dS>0.20 were ranked according to log_2 fold change and the 30 genes with the greatest degree of upregulation at one stage of the soma are shown. Also shown is the free ratio branch dN/dS for each species, the presence of branch-site positive selection (indicated by "yes" (P<0.05)), and the *tau* value. "Stage" indicates the developmental stage in which the gene displayed upregulation.

Fbgn ID	Log_2 fold change	Stage	Gene name	Symbol	M0 dN/dS	dN/dS per branch					Branch-site positive selection					*tau*
						Dsim	Dsec	Dmel	Dyak	Dere	Dsim	Dsec	Dmel	Dyak	Dere	
FBgn0052057	8.698	Late	*defective proboscis extension response 10*	*dpr10*	0.2104	0.0001	0.6537	0.0983	0.1774	0.1875		yes			yes	0.8784
FBgn0039099	8.607	Late	*Gamma-interferon-inducible lysosomal thiol reductase 2*	*GILT2*	0.2453	0.7806	1.5581	0.0437	0.4597	0.1061						0.9546
FBgn0051296	8.161	Early	*CG31296*	*CG31296*	0.3219	0.3233	0.2084	0.2412	0.3172	0.5849					yes	0.9893
FBgn0015872	7.798	Late	*drip*	*Drip*	0.2734	0.6248	>1	0.4140	0.3750	0.1312	yes		yes	yes		0.9786
FBgn0003651	7.644	Late	*seven up*	*svp*	0.2482	0.0001	0.3980	0.3406	0.0001	0.0001				yes		0.7428
FBgn0085222	7.440	Mid	*CG34193*	*CG34193*	0.2104	0.0001	0.2387	0.4706	0.0001	0.3696						0.9757
FBgn0260943	6.235	Late	*RNA-binding protein 6*	*Rbp6*	0.5015	0.5253	0.0001	0.5512	0.0001	0.0001	yes		yes			0.9251
FBgn0040343	5.951	Late	*CG3713*	*CG3713*	0.2636	>1	0.0001	0.8226	0.1501	0.3225						0.9146
FBgn0033076	5.417	Late	*CG15233*	*CG15233*	0.6252	0.1922	0.6592	1.2160	1.0832	0.7722		yes		yes		0.9498
FBgn0035426	5.001	Late	*CG12078*	*CG12078*	0.2513	0.6469	0.2561	0.2993	0.0911	0.1641						0.9517
FBgn0260479	4.938	Mid	*CG31904*	*CG31904*	0.3038	0.0001	0.4808	0.6401	0.0556	0.1363		yes	yes			0.9466
FBgn0052259	4.863	Late	*CG32259*	*CG32259*	0.9261	>1	1.1050	0.8184	0.5433	1.6762				yes	yes	0.9540
FBgn0026315	4.804	Early	*UDP-glycosyltransferase family 35 member A1*	*Ugt35A1*	0.2279	0.1670	0.2725	0.2604	0.1125	0.3850					yes	0.7528
FBgn0030776	4.620	Mid	*CG4653*	*CG4653*	0.2657	0.1363	0.2587	0.3934	0.1863	0.3557			yes			0.8962
FBgn0032780	4.397	Late	*CG13085*	*CG13085*	0.2262	0.2711	0.4829	0.1077	0.1360	0.3106						0.8017
FBgn0002868	4.380	Early	*Metallothionein A*	*MtnA*	0.6883	-	>1	>1	>1	0.1265						0.8753
FBgn0052246	4.351	Late	*CG32246*	*CG32246*	0.6842	2.9026	1.4628	0.4046	0.3813	0.8883		yes			yes	0.9569
FBgn0033541	4.345	Late	*CG12934*	*CG12934*	0.4946	0.0939	0.4220	0.1443	0.6220	0.7280					yes	0.9633
FBgn0087012	4.305	Early	*5-hydroxytryptamine (serotonin) receptor 2A*	*5-HT2A*	0.3593	0.4309	0.4999	0.3961	0.5155	0.5550					yes	0.8491
FBgn0032780	4.243	Mid	*CG13085*	*CG13085*	0.2262	0.2711	0.4829	0.1077	0.1360	0.3106						0.8017
FBgn0035917	4.230	Late	*Z band alternatively spliced PDZ-motif protein 66*	*Zasp66*	0.2118	0.0001	0.4500	0.0646	0.0340	0.0811						0.8459
FBgn0016075	4.200	Late	*viking*	*vkg*	0.3860	0.3038	0.7557	0.2965	0.3510	0.5572		yes		yes	yes	0.9037
FBgn0004607	4.199	Late	*Zn finger homeodomain 2*	*zfh2*	0.2410	0.1792	0.4571	0.2484	0.2128	0.2496		yes				0.7816
FBgn0039679	4.130	Mid	*pickpocket 19*	*ppk19*	0.2142	0.2244	0.1435	0.3729	0.1705	0.3416						0.9724
FBgn0039938	4.025	Late	*Sox102F*	*Sox102F*	0.2506	0.1676	0.3726	0.2315	0.1600	0.3848						0.7231
FBgn0039679	3.933	Late	*pickpocket 19*	*ppk19*	0.2142	0.2244	0.1435	0.3729	0.1705	0.3416						0.9724
FBgn0000299	3.875	Late	*Collagen type IV alpha 1*	*Col4a1*	0.4065	0.3034	1.2519	0.2072	0.4728	0.6879		yes		yes	yes	0.8887
FBgn0003380	3.700	Early	*Shaker*	*Sh*	0.2416	0.4047	0.0001	0.0001	0.0001	0.0975	yes					0.8968
FBgn0046874	3.519	Late	*PFTAIRE-interacting factor 1B*	*Pif1B*	0.3485	0.4171	0.3360	0.2116	0.3971	0.2675						0.6881
FBgn0030776	3.331	Late	*CG4653*	*CG4653*	0.2657	0.1363	0.2587	0.3934	0.1863	0.3557			yes			0.8962

Notes: "-" indicates the dN and dS were each <0.001 and thus too low divergence to determine dN/dS. "yes" indicates P<0.05 in branch-site selection test.

Table S7. Rapidly evolving genes (M0 dN/dS>0.20 in the *melanogaster* subgroup) that were statistically significantly upregulated in one cell type of the LL3 ovary relative to all other cell types from the Slaidina et al. (2020) sc-RNA seq data (P<0.05 using Seurat program v.2 (where log=$\log_e$ in v2 (Satija, et al. 2015)). The genes listed include all those with M0 dN/dS>0.20 and with upregulation in the cell type provided. Genes could be upregulated in more than one cell type and thus may be listed more than once (Slaidina, et al. 2020). The subset of genes that were exclusively upregulated in the terminal filament cells (TF) or sheath cells (SH) cells are listed under Notes. The results are shown for the SHa= anterior sheath cells, SHm= migrating sheath cells, TFa= anterior terminal filament cells, TFp= posterior terminal filament cells.

Fbgn	Gene name	Symbol	Avg_logFC	Cell type	M0 dN/dS	dN/dS per branch					Branch-site positive selection					*tau*
						Dsim	Dsec	Dmel	Dyak	Dere	Dsim	Dsec	Dmel	Dyak	Dere	
Anterior sheath cells (SHa)																
FBgn0051363	*Jupiter*	*Jupiter*	0.9101	SHa	0.3236	0.0001	0.3714	>1	0.3439	0.2218				yes		0.8004
FBgn0025879	*Tissue inhibitor of metalloproteases*	*Timp*	0.6981	SHa	0.3563	-	0.5309	0.1187	0.152	0.5264		yes				0.7205
FBgn0033438	*Matrix metalloproteinase 2*	*Mmp2*	0.3692	SHa	0.3464	0.0328	0.3251	0.3121	0.2215	0.528					yes	0.8539
FBgn0014133	*bifocal*	*bif*	0.3673	SHa	0.3763	0.2948	0.7923	0.1719	0.1024	0.151		yes				0.7583
FBgn0033724	*CG8501*	*CG8501*	0.3626	SHa	0.4481	>1	0.3405	0.1569	0.782	0.7183	yes					0.8952
FBgn0261552	*pasilla*	*ps*	0.3091	SHa	0.3513	0.1772	1.0481	0.1473	0.1014	0.0519	yes			yes	yes	0.7835
FBgn0260768	*CG42566*	*CG42566*	0.2971	SHa	0.2277	-	0.3063	0.5437	0.0807	0.3225			yes			0.8882
FBgn0000719	*folded gastrulation*	*fog*	0.2724	SHa	0.3551	0.754	0.6134	0.5948	0.4491	0.2274						0.6477
FBgn0003507	*serpent*	*srp*	0.2635	SHa	0.2258	0.3777	0.2928	0.1872	0.1613	0.2537	yes	yes				0.8608
Migrating sheath cells (SHm)																
FBgn0025879	*Tissue inhibitor of metalloproteases*	*Timp*	0.9614	SHm	0.3563	-	0.5309	0.1187	0.152	0.5264		yes				0.7205
FBgn0051363	*Jupiter*	*Jupiter*	0.9497	SHm	0.3236	0.0001	0.3714	>1	0.3439	0.2218				yes		0.8004
FBgn0033724	*CG8501*	*CG8501*	0.5135	SHm	0.4481	>1	0.3405	0.1569	0.782	0.7183	yes					0.8952
FBgn0027932	*A kinase anchor protein 200*	*Akap200*	0.4686	SHm	0.3478	0.3369	0.2055	0.5401	0.3198	0.5119						0.8698
FBgn0000299	*Collagen type IV alpha 1*	*Col4a1*	0.4654	SHm	0.4065	0.3034	1.2519	0.2072	0.4728	0.6879		yes		yes	yes	0.8887
FBgn0036101	*Ninjurin A*	*NijA*	0.3684	SHm	0.2904	0.1572	0.4455	0.0821	0.4177	0.1882		yes		yes		0.8700
FBgn0014133	*bifocal*	*bif*	0.3325	SHm	0.3763	0.2948	0.7923	0.1719	0.1024	0.151		yes				0.7583
FBgn0261552	*pasilla*	*ps*	0.3178	SHm	0.3513	0.1772	1.0481	0.1473	0.1014	0.0519	yes			yes	yes	0.7835
FBgn0016075	*viking*	*vkg*	0.3121	SHm	0.3860	0.3038	0.7557	0.2965	0.351	0.5572		yes		yes	yes	0.9037
FBgn0033438	*Matrix metalloproteinase 2*	*Mmp2*	0.3011	SHm	0.3464	0.0328	0.3251	0.3121	0.2215	0.528					yes	0.8539
FBgn0051955	*CG31955*	*CG31955*	0.2604	SHm	0.6646	5.8793	0.2682	0.7278	0.5471	0.7476					yes	0.8951
Anterior terminal filament cells (TFa)																
FBgn0015229	*gliolectin*	*glec*	1.7677	TFa	0.5567	0.14	0.1492	0.8271	0.4095	0.6083						0.6855
FBgn0034199	*Growth-blocking peptide 1*	*Gbp1*	1.6379	TFa	0.4162	>1	0.4375	0.0001	1.1754	0.3524				yes		0.9484
FBgn0025879	*Tissue inhibitor of metalloproteases*	*Timp*	1.4819	TFa	0.3563	-	0.5309	0.1187	0.152	0.5264		yes				0.7205
FBgn0034198	*CG11400*	*CG11400*	1.4270	TFa	0.3592	0.0001	0.188	0.3328	1.3207	0.355						0.7854
FBgn0001253	*Ecdysone-inducible gene E1*	*ImpE1*	1.3992	TFa	0.2267	0.2315	0.5900	0.2456	0.1932	0.2401		yes				0.9160
FBgn0034638	*CG10433*	*CG10433*	1.0711	TFa	0.2948	0.0001	0.8546	0.0001	0.1363	0.0362						0.8591
FBgn0262563	*CG43103*	*CG43103*	1.0423	TFa	0.2694	0.0001	0.3256	0.1094	0.4067	0.4623				yes		0.7215
FBgn0033134	*Tetraspanin 42El*	*Tsp42El*	0.9933	TFa	0.2052	0.0001	0.133	0.1603	0.1166	0.39						0.7592
FBgn0001227	*Heat shock gene 67Ba*	*Hsp67Ba*	0.9834	TFa	0.2828	0.0001	0.7352	0.1914	0.2109	0.4008		yes			yes	0.9249

FBgn0261822	*Basigin*	*Bsg*	0.9380	TFa	0.3411	0.1543	0.4645	0.0416	0.7968	0.157		yes		yes		0.6604
FBgn0038071	*Dpp target gene*	*Dtg*	0.8901	TFa	0.3946	0.2878	0.3957	0.503	0.4808	0.506						0.8793
FBgn0036101	*Ninjurin A*	*NijA*	0.8596	TFa	0.2904	0.1572	0.4455	0.0821	0.4177	0.1882		yes		yes		0.8700
FBgn0020503	*Cytoplasmic linker protein 190*	*CLIP-190*	0.7221	TFa	0.2379	0.1466	0.4152	0.2074	0.2021	0.1592		yes				0.6998
FBgn0015872	*Drip*	*Drip*	0.6456	TFa	0.2734	0.6248	>1	0.414	0.375	0.1312	yes		yes	yes		0.9786
FBgn0003435	*smooth*	*sm*	0.6382	TFa	0.3992	>1	0.6092	0.0309	0.0001	0.5679		yes				0.8599
FBgn0031968	*CG7231*	*CG7231*	0.6366	TFa	0.2445	0.3018	0.6806	0.1559	0.1826	0.3635		yes				0.8268
FBgn0034398	*CG15098*	*CG15098*	0.5951	TFa	0.4849	>6	0.743	0.8262	0.3037	0.6975					yes	0.7146
FBgn0032949	*Lamp1*	*Lamp1*	0.5880	TFa	0.3451	0.0001	0.4433	0.5723	0.3839	0.4147						0.8346
FBgn0038476	*kugelkern*	*kuk*	0.5606	TFa	0.2970	0.1612	0.3325	0.3852	0.1806	0.3661						0.9247
FBgn0000299	*Collagen type IV alpha 1*	*Col4a1*	0.5519	TFa	0.4065	0.3034	1.2519	0.2072	0.4728	0.6879		yes		yes	yes	0.8887
FBgn0033889	*CG6701*	*CG6701*	0.5396	TFa	0.3091	0.245	1.0488	0.3838	0.288	0.4045		yes				0.6936
FBgn0030174	*CG15312*	*CG15312*	0.5228	TFa	0.3686	0.0001	0.6218	0.1324	0.5161	0.1388		yes				0.8095
FBgn0038682	*CG5835*	*CG5835*	0.4978	TFa	0.3324	0.0384	0.1801	0.6696	0.2774	0.5286						0.8715
FBgn0030237	*CG15209*	*CG15209*	0.4724	TFa	0.2491	0.2992	0.2840	0.2506	0.1882	1.1764						0.7312
FBgn0051361	*defective proboscis extension response 17*	*dpr17*	0.4614	TFa	0.3000	0.2126	0.2577	0.255	0.6819	0.3233						0.8514
FBgn0035020	*CG13585*	*CG13585*	0.4568	TFa	0.2637	0.2087	>1	0.1471	0.0931	0.188						0.9061
FBgn0086686	*lethal (3) L1231*	*l(3)L1231*	0.4452	TFa	0.2198	0.1294	0.4058	0.2666	0.1919	0.2303						0.6575
FBgn0040343	*CG3713*	*CG3713*	0.4337	TFa	0.2636	>1	0.0001	0.8226	0.1501	0.3225						0.9146
FBgn0085407	*PDGF- and VEGF-related factor 3*	*Pvf3*	0.4201	TFa	0.2440	0.0245	0.3884	0.0525	0.1055	0.1217						0.9222
FBgn0016075	*viking*	*vkg*	0.4074	TFa	0.3860	0.3038	0.7557	0.2965	0.351	0.5572		yes		yes	yes	0.9037
FBgn0039914	*maverick*	*mav*	0.3770	TFa	0.2504	0.3964	0.532	0.5972	0.2124	0.4124						0.6731
FBgn0038498	*beaten path IIa*	*beat-IIa*	0.3759	TFa	0.2430	0.1713	0.0001	0.2514	0.3582	0.2106						0.8736
FBgn0034200	*Growth-blocking peptide 2*	*Gbp2*	0.3751	TFa	0.3622	0.3128	0.2299	0.3764	0.294	0.3957						0.8323
FBgn0027598	*CIN85 and CD2AP related*	*cindr*	0.3680	TFa	0.3297	0.2759	0.6744	0.0758	0.091	0.0772	yes	yes				0.6435
FBgn0035452	*CG10359*	*CG10359*	0.3640	TFa	0.2203	0.2327	0.132	0.3775	0.2047	0.3199						0.8757
FBgn0085432	*pangolin*	*pan*	0.3497	TFa	0.3218	0.2243	0.5643	0.1367	0.126	0.165	yes					0.8586
FBgn0010415	*Syndecan*	*Sdc*	0.3348	TFa	0.3006	0.6124	0.6265	0.2136	0.2055	0.5697		yes			yes	0.8616
FBgn0045842	*yuri gagarin*	*yuri*	0.3305	TFa	0.3037	0.2003	0.3102	0.4189	0.255	0.2499						0.9364
FBgn0037350	*Diphthamide biosynthesis 4*	*CG2911*	0.3238	TFa	0.2303	0.0001	0.4093	0.5159	0.2446	0.1918				yes		0.9143
FBgn0259740	*CG42394*	*CG42394*	0.3003	TFa	0.3499	0.0001	0.0001	2.6837	0.0001	0.2091						0.7401
FBgn0052803	*Lipid droplet assembly factor 1*	*CG32803*	0.2780	TFa	0.2664	0.0001	0.4282	0.8583	0.311	0.1959			yes			0.6557
FBgn0033683	*CG18343*	*CG18343*	0.2762	TFa	1.0011	1.2209	0.4346	4.108	0.3598	1.0348			yes	yes		0.5615
FBgn0026620	*transforming acidic coiled-coil protein*	*tacc*	0.2659	TFa	0.3878	0.2304	0.5105	0.3432	0.4609	0.435		yes			yes	0.5372
FBgn0034724	*babos*	*babos*	0.2632	TFa	0.2450	0.7171	0.0834	0.2045	0.2332	0.3787						0.7684
Posterior terminal filament cells (TFp)																
FBgn0034199	*Growth-blocking peptide 1*	*Gbp1*	1.6762	TFp	0.4162	>1	0.4375	0.0001	1.1754	0.3524				yes		0.9484
FBgn0000299	*Collagen type IV alpha 1*	*Col4a1*	1.4987	TFp	0.4065	0.3034	1.2519	0.2072	0.4728	0.6879		yes		yes	yes	0.8887
FBgn0015229	*gliolectin*	*glec*	1.4901	TFp	0.5567	0.14	0.1492	0.8271	0.4095	0.6083						0.6855
FBgn0001253	*Ecdysone-inducible gene E1*	*ImpE1*	1.3009	TFp	0.2267	0.2315	0.5900	0.2456	0.1932	0.2401		yes				0.9160

FBgn0025879	*Tissue inhibitor of metalloproteases*	*Timp*	1.1705	TFp	0.3563	-	0.5309	0.1187	0.152	0.5264		yes				0.7205
FBgn0262563	*CG43103*	*CG43103*	1.1400	TFp	0.2694	0.0001	0.3256	0.1094	0.4067	0.4623				yes		0.7215
FBgn0002868	*Metallothionein A*	*MtnA*	1.0706	TFp	0.6883	-	>1	>1	>1	0.1265						0.8753
FBgn0033134	*Tetraspanin 42El*	*Tsp42El*	0.9692	TFp	0.2052	0.0001	0.133	0.1603	0.1166	0.39						0.7592
FBgn0016075	*viking*	*vkg*	0.9189	TFp	0.3860	0.3038	0.7557	0.2965	0.351	0.5572		yes		yes	yes	0.9037
FBgn0034198	*CG11400*	*CG11400*	0.8837	TFp	0.3592	0.0001	0.188	0.3328	1.3207	0.355						0.7854
FBgn0001227	*Heat shock gene 67Ba*	*Hsp67Ba*	0.8687	TFp	0.2828	0.0001	0.7352	0.1914	0.2109	0.4008		yes			yes	0.9249
FBgn0034638	*CG10433*	*CG10433*	0.7609	TFp	0.2948	0.0001	0.8546	0.0001	0.1363	0.0362						0.8591
FBgn0015872	*Drip*	*Drip*	0.7607	TFp	0.2734	0.6248	>1	0.414	0.375	0.1312	yes		yes	yes		0.9786
FBgn0036101	*Ninjurin A*	*NijA*	0.7144	TFp	0.2904	0.1572	0.4455	0.0821	0.4177	0.1882		yes		yes		0.8700
FBgn0259740	*CG42394*	*CG42394*	0.6697	TFp	0.3499	0.0001	0.0001	2.6837	0.0001	0.2091						0.7401
FBgn0031968	*CG7231*	*CG7231*	0.5292	TFp	0.2445	0.3018	0.6806	0.1559	0.1826	0.3635		yes				0.8268
FBgn0035020	*CG13585*	*CG13585*	0.4914	TFp	0.2637	0.2087	>1	0.1471	0.0931	0.188						0.9061
FBgn0052803	*Lipid droplet assembly factor 1*	*CG32803*	0.4869	TFp	0.2664	0.0001	0.4282	0.8583	0.311	0.1959			yes			0.6557
FBgn0003435	*smooth*	*sm*	0.4832	TFp	0.3992	>1	0.6092	0.0309	0.0001	0.5679		yes				0.8599
FBgn0038476	*kugelkern*	*kuk*	0.4056	TFp	0.2970	0.1612	0.3325	0.3852	0.1806	0.3661						0.9247
FBgn0010415	*Syndecan*	*Sdc*	0.4054	TFp	0.3006	0.6124	0.6265	0.2136	0.2055	0.5697		yes			yes	0.8616
FBgn0261822	*Basigin*	*Bsg*	0.3741	TFp	0.3411	0.1543	0.4645	0.0416	0.7968	0.157		yes		yes		0.6604
FBgn0033683	*CG18343*	*CG18343*	0.3315	TFp	1.0011	1.2209	0.4346	4.108	0.3598	1.0348			yes	yes		0.5615
FBgn0034200	*Growth-blocking peptide 2*	*Gbp2*	0.2992	TFp	0.3622	0.3128	0.2299	0.3764	0.294	0.3957						0.8323
FBgn0038498	*beaten path IIa*	*beat-IIa*	0.2919	TFp	0.2430	0.1713	0.0001	0.2514	0.3582	0.2106						0.8736
FBgn0035452	*CG10359*	*CG10359*	0.2855	TFp	0.2203	0.2327	0.132	0.3775	0.2047	0.3199						0.8757
FBgn0034398	*CG15098*	*CG15098*	0.2681	TFp	0.4849	>1	0.743	0.8262	0.3037	0.6975					yes	0.7146
FBgn0030174	*CG15312*	*CG15312*	0.2649	TFp	0.3686	0.0001	0.6218	0.1324	0.5161	0.1388		yes				0.8095
FBgn0038071	*Dpp target gene*	*Dtg*	0.2564	TFp	0.3946	0.2878	0.3957	0.503	0.4808	0.506						0.8793
FBgn0043841	*virus-induced RNA 1*	*vir-1*	0.2541	TFp	0.2419	0.1295	0.111	0.1677	0.4729	0.3974				yes		0.7401

Notes: Genes that were exclusively upregulated in the SH cells: FBgn0051363, FBgn0033438, FBgn0014133, FBgn0033724, FBgn0000719, FBgn0003507, FBgn0027932, FBgn0051955. Genes that were exclusively upregulated in the TFs: FBgn0001253, FBgn0003435, FBgn0015872, FBgn0020503, FBgn0026620, FBgn0027598, FBgn0030174, FBgn0030237, FBgn0033134, FBgn0033683, FBgn0034198, FBgn0034200, FBgn0034638, FBgn0037350, FBgn0038071, FBgn0038476, FBgn0038498, FBgn0038682, FBgn0039914, FBgn0040343, FBgn0045842, FBgn0051361, FBgn0085407,FBgn0085432, FBgn0086686, FBgn0262563. Branch-site positive selection in *D. simulans*, *D. sechellia* and *D. melanogaster* using the subset of genes exclusively upregulated in SH cells was 25.0, 25.0 and 0% of studied genes respectively and for TFs was 11.5, 23.1, and 7.7% respectively. “-“ indicates the dN and dS were each <0.001 and thus too low divergence to determine dN/dS.

Table S8. The top GO functional classifications of rapidly evolving genes that were upregulated in TFa and TFp cell types. Classifications were determined using DAVID (Huang da, et al. 2009) for all genes per category. P-values indicate the statistical significance of preferential classification of the gene into each GO group. Genes could match more than one GO group.

Functional class	**Percent of genes matching**	**P-value**
Terminal filament cells anterior (a)		
basement membrane organization	6.8	9.10E-04
cytokine activity	6.8	1.00E-03
extracellular space	20.5	2.80E-03
cell adhesion	6.8	1.50E-02
growth factor activity	4.5	4.30E-02
extracellular matrix structural constituent	4.5	5.80E-02
Terminal filament cells posterior (p)		
basement membrane organization	10	3.70E-04
extracellular space	23.3	4.10E-03
cell adhesion	10	6.50E-03
collagen type IV trimer	6.7	6.80E-03
cytokine activity	6.7	2.90E-02
extracellular matrix structural constituent	6.7	3.60E-02
integral component of membrane	40	3.80E-02
axon guidance	10	3.90E-02
extracellular matrix	10	4.20E-02
	6.7	4.40E-02

Table S9. The results for the McDonald and Kreitman (1991)tests of positive selection for the 42 genes identified in Tables 1, 2, and 3. Only those genes with P<0.05 are shown. The *D. melanogaster* populations examined are the Raleigh North Carolina (RAL) and the Zambia population. The P-values are shown as MK-P. Alpha (α) indicates the proportion of substitutions apt to have experienced positive selection. Analysis was conducted using the integrative McDonald Kreitman test (Murga-Moreno, et al. 2019) that uses *D. simulans* as the comparison species (Dmel-Dsim). Note that iMKT may not necessarily have the exact same alignment per gene as FlyDivas (Stanley and Kulathinal 2016), which was used for all our core analyses, but the alignments from both analyses are highly similar and representative per gene for this supplementary assessment. Data are available publicly for each dataset (Stanley and Kulathinal 2016; Murga-Moreno, et al. 2019). MK-P indicates the P-value.

		RAL population		**Zambia population**	
Fbg ID	Gene	α	MK-P	α	MK-P
FBgn0026379	*Pten*	0.880	0,012	0.771	0.030
FBgn0004858	*elB*	0.756	0.057	0.703	0.001
FBgn0010825	*Gug*	0.530	0.010	0.598	0
FBgn0261854	*aPKC*	0.969	0	0.944	0
FBgn0032006	*Pvr*	0.558	0.020	0.525	0.014
FBgn0039108	*CG10232*	0.632	0.031	0.669	0.002
FBgn0016075	*vkg*	0.681	0	0.769	0
FBgn0000299	*Col4a1*	0.711	0,001	0,757	0

Table S10. The branch dN/dS and branch-site analysis Hawaiian *Drosophila* (*D. murphyi*, *D. sproati* and *D. grimshawi*) for orthologs to the SIGNALC genes identified in table 1 that exhibited signs of involvement in ovariole number evolution (from analysis of the *melanogaster* subgroup, and (Kumar, et al. 2020)). Genes having a dN/dS value of >1 at the gene-wide level and those with values above 0.330 in at least one species branch are each shown (that is double the median of the genome with the highest median dN/dS, *D. sproati*). Species that exhibited positive selection using branch-site (BR-S) analysis in PAML (P<0.05) are indicated by the abbreviated species name.

Fbgn	**Gene name**	**Symbol**	**Hawaiian branch-dN/dS**			**BR-S positive selection**		
			D. murphyi	*D. sproati*	*D. grimshawi*			
Genes with dN/dS>0.330 in at least one branch								
FBgn0030904	*unpaired 2*	*upd2*	**2.7852**	1.6904	1.1397	Dmur	Dspr	Dgri
FBgn0035213	*CG2199*	*CG2199*	**0.637**	0.5455	0.5134			
FBgn0003984	*vein*	*vn*	0.1162	**0.3489**	**0.3484**			Dgri
FBgn0004858	*elbow B*	*elB*	0.0738	**0.4486**	0.0744		Dspr	
FBgn0259176	*bunched*	*bun*	0.5683	0.3779	**0.7306**	Dmur		Dgri
FBgn0023540	*CG3630*	*CG3630*	**0.5018**	0.2986	0.3435			
FBgn0261854	*atypical protein kinase C*	*aPKC*	0.1756	0	**0.5173**			Dgri
FBgn0001169	*Hairless*	*H*	0.3588	**0.8827**	0.4621			Dgri
FBgn0003607	*Suppressor of variegation 205*	*Su(var)205*	0.4541	**0.4620**	0.1688			
FBgn0002592	*Enhancer of split m2, Bearded family member*	*E(spl)m2-BFM*	NA	**0.3478**	0.2298			
Genes with all branches dN/dS≤0.330								
FBgn0014020	*Rho1*	*Rho1*	NA	NA	0			
FBgn0003612	*Suppressor of variegation 2-10*	*Su(var)2-10*	0.0001	0.0001	0.2512			Dgri
FBgn0000259	*Casein kinase II beta subunit*	*CkIIbeta*	0.0001	NA	0.0001			
FBgn0262614	*polychaetoid*	*pyd*	0.191	0.0652	0.0777	Dmur		
FBgn0036974	*eukaryotic translation release factor 1*	*eRF1*	NA	0	0.0001			
FBgn0002174	*CG5504*	*CG5504*	0.1637	0.1607	0.2089			Dgri
FBgn0037218	*auxilin*	*aux*	0.2491	0.1765	0.1276			

FBgn0024291	*Sirtuin 1*	*Sirt1*	0.2621	0.294	0.2123
FBgn0020496	*C-terminal Binding Protein*	*CtBP*	0	0.0001	0.0001
FBgn0261592	*Ribosomal protein S6*	*RpS6*	0.0001	0.0001	0.0274
FBgn0045035	*telomere fusion*	*tefu*	0.2727	0.1711	0.2456
No Orthologs					
FBgn0011274	*Dorsal-related immunity factor*	*Dif*	No Orth-HD		
FBgn0026379	*Phosphatase and tensin homolog*	*Pten*	No Orth-HD		
FBgn0011642	*Zyxin*	*Zyx*	No Orth-HD		
FBgn0010825	*Grunge*	*Gug*	No Orth-HD		
FBgn0020386	*Phosphoinositide-dependent kinase 1*	*Pdk1*	No Orth-HD		
FBgn0032006	*PDGF- and VEGF-receptor related*	*Pvr*	No Orth-HD		

Notes: No Orth-HD=no high confidence reciprocal BLASTX orthologs identified among the three Hawaiian *Drosophila* (HD). NA indicates the branch did not meet the criteria that dN or dS>0.001 for dN/dS analysis. Ten genes had dN/dS>2 fold higher (value>0.330) than the genome-wide medians in at least one species branch (the branch with the highest dN/dS value is in bold). Three additional genes, *aux*, *Sirt1*, and *tefu* had an elevated value when using a criterion of dN/dS >1.5 fold higher (value>0.246) than the genome-wide medians.

SUPPLEMENTARY TEXT FILE S1

Supplementary Methods for the *melanogaster* subgroup

Identification of Rapidly Evolving Ovariole-Related Genes

With respect to the SIGNALC dataset (Kumar, et al. 2020), for each of four gene sets identified therein as having roles in regulating egg-laying and/or ovariole numbers, namely *hpo[RNAi]* Ovariole Number, *hpo[RNAi]* Egg Laying, Egg Laying [*wt*], and the so-called "Connectors" (genes not included in the initial screen but identified as candidates by network topology analysis), we identified the subset of genes that had orthologs in all five species of the *melanogaster* subgroup and exhibited a markedly elevated M0 dN/dS value relative to the genome-wide values (Yang 2007) using data from FlyDivas (Stanley and Kulathinal 2016). Given that signalling genes are typically components of complex pathways and have pleiotropic roles in a wide range of cells or tissues (Cui, et al. 2009; Kumar, et al. 2020), they may be expected to typically, but, not always (Cui, et al. 2009; Darfour-Oduro, et al. 2016), have been subjected to strong purifying selection and thus evolve slowly in their protein sequence (Cui, et al. 2009; Mank and Ellegren 2009; Meisel 2011; Assis, et al. 2012; Darfour-Oduro, et al. 2016; Masalia, et al. 2017; Whittle, et al. 2021). We therefore defined rapidly evolving ovariole-related genes as those having a value of ≥1.5 fold above the genome-wide median M0 dN/dS in the *melanogaster* subgroup (genome-wide median M0 dN/dS=0.0887, N=9,232). The cut-off was 1.5 (M0 dN/dS value ≥0.133) rather than higher, as for the other studied datasets (see below, BULKSG and SINGLEC), given the innate conserved nature of these genes. In addition, for this one dataset, the SIGNALC, to identify the widest scope possible of genes with a propensity for rapid evolution for follow-up study, we also included any genes with ≥1.5 fold higher M0 dN/dS in the *D. simulans*-*D. melanogaster* pair (≥1.5 fold higher than the median=0.857, N=10,765 aligned genes) and the six-species *melanogaster* group (which includes *D. ananassae* as a outgroup, fig. 2; genome-wide median=0.062, N=8,696 genes) (Stanley and Kulathinal 2016). These genes were added to the gene list for our follow-up study within the *melanogaster* subgroup (fig. 2).

For the BULKSG transcription dataset (Tarikere, et al. 2022), we identified and extracted all genes that were statistically significantly upregulated (using DeSeq2 $P<0.01$ criterion in that assessment (Love, et al. 2014)) in the *D. melanogaster* pooled larval ovary somatic cells and those upregulated in the pooled germ cells. Upregulated in germ cells automatically indicates downregulation in soma cells, and thus both up- and downregulated genes in soma were studied. We also extracted for study those genes that were upregulated at a single one of the larval stages studied (early, mid, or late) for the pooled ovary somatic cells; these stages correspond to different stages of TF formation (Tarikere, et al. 2022). For these gene sets, we then identified those genes with M0 dN/dS>0.20 in the *melanogaster* subgroup, which represent a value ≥2.2 fold higher than the genome-wide median. The cut-off was higher for the BULKSG than SIGNALC dataset, given that we examined all genes in the genome that were differentially expressed in the former dataset, which are apt to be less conserved as a group than the SIGNALC genes. The genes matching these criteria (dN/dS>0.20, and up- or downregulated in soma cells), were ranked based on degree of differential expression, and subjected to follow-up analysis in the *melanogaster* subgroup.

With respect to the SINGLEC dataset (Slaidina, et al. 2020), genes were extracted for further study that were statistically significantly differentially transcribed in a specific cell type of the LL3 larval ovary relative to all other cell types based on sc-RNA seq (using average standardized expression and P-values from Seurat v.2; some genes were upregulated in more than one cell type based on these criteria (Slaidina, et al. 2020)). The nine cell types studied (shown in fig. 1) included the germ cells (GC) and eight somatic cell types, namely the cap cells (CC), follicle stem cell precursors (FSCP), intermingled cells (IC), anterior sheath cells (SHa), migrating sheath cells (SHm), swarm cells (SW), anterior terminal filaments (TFa), and posterior terminal filaments (TFp), with a particular focus on the two types of SH cells and the two types of TF cells, which are thought to largely shape species-specific ovariole numbers and function (King, et al. 1968; Sarikaya, et al. 2012; Sarikaya and Extavour 2015; Slaidina, et al. 2020). Similar to the BULKSG assessment, among the genes that were upregulated per cell type, we identified the genes that had signs of rapid protein sequence divergence (M0 dN/dS >0.20 in genes with five-species orthologs) in the *melanogaster* subgroup (Stanley and Kulathinal 2016; Gramates, et al. 2017), for further analyses.

PGLS Analysis

Any species branch for a gene with an infinity value for dN/dS (denoted as "dN/dS>1" and cited in the tables; that is, dN>0.001, and dS near or at 0); table 1, table 2, table3) was assigned a value of 1.5 for PGLS analysis to indicate that the value was larger than one, while conservatively ascribing a low range value (between one and infinity). If more than one branch had "dN/dS>1", then the gene was excluded from the PGLS (one case, FBgn0002868 in table 3). Any species branch for a gene without a dN/dS value (dN and dS<0.001) was assigned "N/A" for PGLS.

***tau* Analysis**

The *tau* value is a relative measure, and can vary with the number of tissues/stages studied. For instance, we found that *tau* values across all studied genes in the genome determined using all 59 tissues/stages yielded elevated values as compared to using only the 30 developmental stages alone (median= 0.830 versus 0.727, fig. S2), and the two sets of values were strongly correlated ($R=0.873$, $P<10^{-6}$) suggesting that both values effectively reflect *tau*. Throughout this study, we used the complete available dataset of 59 tissues/stages for all *tau* analyses, and considered genes with values >0.90 as narrowly transcribed (just below the 75th percentile), while those with values below 0.73 (at the 25th percentile) were considered as broadly transcribed, with all other values considered relative to each other and intermediate in *tau*.

Supplementary Methods for the Three-Species Hawaiian Clade

As a supplementary analyses we studied a three species clade of Hawaiian *Drosophila*, shown in fig. S5 (phylogeny from (Kim, et al. 2021; Suvorov, et al. 2022)), that included *D. sproati* (Dspr, N=65.6 ovarioles), *D. murphyi* (Dmur, N=41.6 ovarioles) and *D. grimshawi* (47.8 ovarioles) (ovariole numbers from (Starmer, et al. 2003; Sarikaya, et al. 2019). For *D. sproati* and *D. murphyi* and the genomes were available as unannotated scaffolds from the SRA (Starmer, et al. 2003; Suvorov, et al. 2022) and the closely related species *D. grimshawi* was used as a reference (that has an annotated genome, N=12,825 CDS; longest CDS per gene; all three species genome data under project PRJNA675888 at the SRA database). Using the program Maker2 (Holt and Yandell 2011) and the RNA and proteins sequence list for *D. grimshawi*, we searched *D. sproati*

and *D. murphyi* scaffolds for evidence-based gene predictions. We also obtained ab initio gene predictions in Augustus (Hoff and Stanke 2013) which recommends *D. melanogaster* gene models as a reference for predictions in Dipteran species (Stanke and Waack 2003). In the assessments, we excluded any CDS that lacked a start or a stop codon, or had an internal stop codon, such that only complete CDS were used for study. Using these criteria, we identified 13,395 CDS for *D. murphyi* and 12,670 CDS for *D. sproati* . Thus, these comprise high quality full-length CDS. To assess the completeness of our final *D. sproati* and *D. murphyi* CDS lists we applied BUSCO v. 5.5.2 (Seppey, et al. 2019; Manni, et al. 2021), a program that evaluates the completeness of a given transcriptome by comparing it to a subset of "universal" single-copy genes. For this, we used as the control the Dipteran gene set, named diptera_odb10 (N genes=3,250). The results showed that for the 13,395 full-length CDS for *D. murphyi*, 93.4% had a complete (unfragmented) match in the BUSCO dataset, indicating we obtained a very complete CDS list for this species (note that this value would be elevated using all 14,108 CDS identified for *D. murphyi*, but we focused solely on the obtained full-length CDS). In turn, for *D. sproati* , we found 93.2% of the full-length CDS were represented in the Dipteran BUSCO dataset. Taken together, the results show that our approach yielded a high confidence and highly complete CDS list for *D. sproati* and *D. murphyi.*

We next used reciprocal BLASTX (BLAST v. 2.13.0+ (Altschul, et al. 1997)) to identify orthologs between all three Hawaiian species using all CDS per genome in each contrast, for each combination of paired species, namely *D. grimshawi-D. murphyi*, *D. grimshawi-D. sproati* and *D. murphyi-D. sproati*. For each pair, in order to be identified as orthologs, two CDS had to be the best match (lowest e-score, any matched with a tied e-score the best match was the one with the highest bit value, $e<10^{-6}$) in both the forward and reverse contrast and have an e-value $<10^{-6}$. Using the three sets of paired orthologs, we then compared the orthologs identified in the newly annotated *D. murphyi-D. sproati* pair to the *D. grimshawi*.-*D. murphyi* and *D. grimshawi*. -*D. murphyi* pairs, to ensure that the orthologs per gene identified between *D. murphyi* and *D. sproati* each independently matched the same single CDS in *D. grimshawi*. Thus, this comprises very stringent criteria for identification of three-way reciprocal orthologs among *D. grimshawi*.-*D. murphyi-D. sproati*.

To identify Hawaiian species orthologs to the genes linked to ovariole/egg number evolution obtained from analysis within the *melanogaster* subgroup (SIGNALC in table 1;

SIGNALC in fig. 4), and for genome-wide orthologs, we compared the CDS of the most well annotated *melanogaster* species *D. melanogaster* (13,986 coding genes with longest isoform per gene) to the annotated Hawaiian species *D. grimshawi* CDS lists using reciprocal BLASTX (using same criteria described for Hawaiian *Drosophila* above). We report that 11,011 high confidence orthologs were identified between *D. melanogaster* and *D. grimshawi* (78.7% of the Dmel CDS list). A total of 10,276 of those genes had high confidence three-species orthologs in Hawaiian clade and were used for study of dN/dS and ortholog detection rates.

Each set of three-species orthologs among *D. grimshawi*.-*D. murphyi*-*D. sproati* were aligned using MUSCLE (Edgar 2004) set to default parameters in MEGA (Kumar, et al. 2016). Each alignment was filtered in GBlocks v. 0.91b set at default parameters (Castresana 2000; Talavera and Castresana 2007) to remove gaps and any highly divergent segments. For each aligned CDS in *D. grimshawi*, *D. murphyi* and *D. sproati*(gaps removed), we determined dN/dS per species branch using the free-ratios model of PAML (M1) and conducted branch-site analyses (Yang 2007), as outlined for the *melanogaster* subgroup in the main text. Branches were unsaturated in substitutions, and the 90th percentile of dS values across all genes was <0.090 and of dN was <0.032 for each of the three Hawaiian species branches. The median gene-wide dN/dS value across all genes for each species branch were 0.152, 0.164, and 0.160 for *D. murphyi*, *D. sproati* and *D. grimshawi* respectively.

Supplementary Results for the *Drosophila melanogaster* subgroup

Descriptions of Genes Linked to Ovariole Evolution

Each of the ovariole-related genes listed in the tables herein has properties suggesting a role in interspecies ovariole number divergence. In the following sections, we highlight some of these genes, which we consider especially promising candidates for functional analyses in future work, and briefly describe published evidence for their role in ovariole number determination and/or interspecies ovariole number divergence. We note that none of the 42 ovariole-related genes identified herein as linked to interspecies divergence in ovariole numbers (tables 1-5, fig. 4) overlapped with any of the 24 genes identified using a recent genome-wide association assessment of variation in 205 lines of the *D. melanogaster* Raleigh population (Lobell, et al. 2017). This is unsurprising, given the latter study was focused on intraspecies variation and ovariole numbers in a single *D. melanogaster* population (unfixed variation), while the present study focused on fixed nonsynonymous (to synonymous, dN/dS) and adaptive changes at the interspecies level

SIGNALC genes (table 1)

unpaired 2 (unp2)

The gene *upd2* is a ligand for the JAK/STAT pathway, which interacts with the Hippo pathway during ovariole development (Sarikaya and Extavour 2015). Of the five species in the *melanogaster* subgroup, *upd2* exhibited the highest terminal branch dN/dS value in *D. sechellia* (value of 0.417; table 1) and had signals of branch-site positive selection ($P<0.05$) in this same lineage (and in *D. erecta*, table 1). Thus, the gene protein changes in *D. sechellia* were coupled with the fact that this species has the lowest ovariole number of the species examined (fig. 2). *unp2* also exhibited the highest *tau* value (0.962) of all 27 studied SIGNALC genes, indicating that this gene is narrowly transcribed, which could potentially provide flexibility to adaptively evolve new functions in the ovarioles (table 1; (Otto 2004; Larracuente, et al. 2008; Mank and Ellegren 2009; Whittle, et al. 2021)). Upon inspection, we found that *upd2* was maximally transcribed ($\hat{x}$) in 4-6 hour old *D. melanogaster* embryos, which is also the same stage at which *hpo*, a known effector

of ovariole numbers in larval ovaries (Sarikaya and Extavour 2015) is reported as maximally transcribed (Li, et al. 2014). The 4-6 hour old embryos comprise an embryonic stage during which the somatic gonad precursors (SGPs) are being established (Richardson and Lehmann 2010), and thus speculatively these two genes (*upd2* and *hpo*) could potentially influence ovariole numbers by affecting SGP number or behaviour (King, et al. 1968; Sarikaya, et al. 2012; Sarikaya and Extavour 2015). In sum, *upd2* protein sequence changes may contribute to the divergence in ovariole numbers in the *melanogaster* subgroup, particularly the marked decline in the *D. sechellia* branch.

Zyxin (Zyx)

For the SIGNALC ovariole-related gene *Zyx*, which modulates activity of the Hippo pathway (table S3; Rauskolb et al. 2011; Gaspar et al. 2014; Ma et al. 2016), each of the five species of the *melanogaster* subgroup had a gene-wide dN/dS value (ranging from 0.267 up to >1, table 1) that was more than threefold higher than the genome-wide median per respective species branch (the genome-wide median dN/dS values for each of the five branches are shown in fig. S1). The patterns suggest potential functional divergence of *Zyx* throughout the clade. *Zyx* exhibited gene-wide dN/dS>1 in *D. sechellia*, suggesting a history of positive selection in that lineage, that co-occurred with its marked decline in ovariole numbers (fig. 2), suggesting the potential for a causative relationship. Upon close examination, we found that *Zyx* in the *D. sechellia* branch was a case when dS approached 0 (the other species branches had dS between 0.006 to 0.010), however the dN value was 0.004 (above the cut-off of 0.001), which was double the dN for *D. simulans* and *D. melanogaster* (even though dS was higher for the latter two species, their dN was lower), together indicating that the elevated dN/dS in *D. sechellia* was due to elevated dN (not low dS, and is conservatively denoted as dN/dS>1). Overall, the patterns suggest positive selection of *Zyx* in *D. sechellia*.

There are plausible cell biological mechanisms whereby adaptive changes in the *Zyx* protein may potentially influence ovariole number divergence. The gene *Zyx* encodes an actin cytoskeleton regulator, with roles in Hippo pathway regulation (table S3, (Rauskolb, et al. 2011)), and evidence suggests it can regulate cell behaviors and movements during development (Amsellem, et al. 2005; Matsui and Lai 2013). Given that actin cytoskeleton activity is a prime

candidate for coordinating the cell movements required for the formation of the TFs (Chen, et al. 2001), it may be speculated that Zyx may contribute to this process, and thereby changes in its amino acid sequence may affect TF formation (table 1), and in turn, alter the ovariole numbers (King, et al. 1968; Sarikaya, et al. 2012; Sarikaya and Extavour 2015). This provides a possible mechanism whereby functional changes in the Zyx protein may affect interspecies transitions in ovariole numbers. Further, *Zyx* positively regulates Yki (Rauskolb, et al. 2011), a transcriptional coactivator involved in many processes including cell proliferation (Huang, et al. 2005), and that accumulates in the nucleus in *hpo* RNAi larval ovary somatic cells (Sarikaya and Extavour 2015). We previously showed that elevating Yki levels in the somatic ovary results in an increased number of TFs and ovarioles (Sarikaya and Extavour 2015). Moreover, loss of *Zyx* reduces Yki activity (Rauskolb, et al. 2011). Thus, taken together, it may be hypothesized that the low ovariole numbers in *D. sechellia* as compared to all other *melanogaster* subgroup species (fig. 2) may have arisen from a reduction of Yki activity, that was caused by lowered function/activity of Zyx protein due to its amino acid sequence changes (table 1). Given that Z*yx* exhibited signs of positive selection in the *D. sechellia* branch (gene-wide dN/dS, table 1), this would in theory imply that the reduction in ovariole numbers (rather than an increase) in *D. sechellia* may comprise an adaptive reproductive change.

Phosphatase and tensin homolog (Pten)

The SIGNALC ovariole-related gene *Pten* (Kumar, et al. 2020) is a regulator of the mTOR and insulin signalling pathways (table S3). *Pten* exhibited an exceptionally high dN/dS value (0.594) in the *D. erecta* branch, more than five-fold higher than the *D. erecta* genome-wide median (0.104, fig. S1) and had branch-site positive selection (in *D. erecta*, $P<0.05$). This gene also exhibited a markedly elevated dN/dS value in the *D. melanogaster* branch (0.3122, table 1) and showed positive selection using the McDonald and Kreitman (1991) tests that were based on inter- versus intraspecies divergence solely for *D. melanogaster-D. simulans* ($P<0.012$, see Methods, and table 1 Notes (Murga-Moreno, et al. 2019)). *D. erecta* has an intermediate number of ovarioles per female (value of 27, fig. 2) among the five-species of the *melanogaster* subgroup (fig. 2) and forms an outgroup clade with *D. yakuba* (ovariole number 25.8). Similar to numerous other ovariole-related SIGNALC genes in table 1, *Pten* has core functions in regulating the actin

cytoskeleton (von Stein, et al. 2005) (table S3), and is involved in cell size, cell growth and tumor suppression (Goberdhan, et al. 1999). Mutations in *Pten* impair cytoskeletal activity in *Drosophila* ovaries, and induce various female reproductive phenotypes including impaired localization of the germ plasm components *oskar* mRNA and Vasa protein in laid eggs and an absence of pole cells (von Stein, et al. 2005). Further, *Pten* RNAi in the *D. melanogaster* larval ovaries increases ovariole number (the *hpo[RNAi]* Ovariole Number phenotype (Kumar, et al. 2020), see table 1). Among all of the 59 tissues/stages used for *tau* analyses (which was a relatively low value of 0.663 in *D melanogaster* for this gene), *Pten* had maximal expression, $\hat{x}$=1, for the adult virgin ovaries (table S3), which is also consistent with an essential female ovarian role. In sum, given its functions in female reproductive system, and the rapid evolution of this signalling gene in four of five of the *Drosophila* lineages (0.128-0.594), and explicit evidence of positive selection in *D. erecta* (table 1), we hypothesize that *Pten* is a putative factor involved in ovariole number transitions in the *melanogaster* subgroup, and that may shape ovariole numbers by adaptive evolution, particularly in the *D. erecta* branch (table 1).

vein (vn)

The ovariole-related EGF pathway ligand gene *vn* (table S3) exhibited a dN/dS value that was between 1.6 to 7.2 fold higher than the genome-wide median per respective species branch in three of the *Drosophila* species branches studied (*D. simulans*, *D. sechellia* and *D. yakuba* had values of 0.4712, 0.2069, and 0.2802 respectively; table 1; genome-wide medians in fig. S1), and presented signals of branch-site positive selection in the *D. yakuba* branch ($P<0.05$, table 1). *vn* is involved in growth and tissue patterning, including roles in the *D. melanogaster* adult ovary in somatic escort cells and in the differentiation of female germ stem cells (Slaidina, et al. 2021; Tu, et al. 2021). Further, this gene is expressed in the *D. melanogaster* TFs and SH cells (Slaidina, et al. 2020), which regulate ovariole formation and number (King, et al. 1968; Sarikaya, et al. 2012; Sarikaya and Extavour 2015; Slaidina, et al. 2020), consistent with the finding that *vn* RNAi reduces ovariole number in *D. melanogaster* (the *hpo[RNAi]* Ovariole Number phenotype, table 1 (Kumar, et al. 2020)). The highly elevated dN/dS value for *vn* in both *D. simulans* and *D. sechellia* terminal branches in particular suggest that *vn* protein sequence evolution may contribute to the rapid and two-fold level change in ovariole numbers between these closely related species (fig. 2).

In sum, *vn* has signs of evolvability in its protein sequence, with frequent protein sequence changes and adaptive evolution (in *D. yakuba*) and yet also exhibited relatively slow evolution in the *D. melanogaster* branch (dN/dS=0.0841), perhaps due to high pleiotropy (*tau*=0.741, table S3; note *tau* was measured in *D. melanogaster* as the reference), suggesting the gene has been subjected to strong purifying selection in that lineage. Thus, *vn* appears to have a dynamic evolutionary pattern, suggestive of potential roles in ovariole number divergence in the *D. simulans* and *D. sechellia* branches, and in *D. yakuba*.

BULKSG genes (table 2)

Insulin-like peptide 5 (Ipl5)

The ovariole-related gene *Ilp5* in table 2 was particularly notable in that it exhibited strong purifying selection in *D. melanogaster* (dN/dS=0.001) and had high dN/dS values in all other branches (values of 0.2932-0.5843 in *D. simulans*, *D. sechellia*, *D. yakuba*, *D. erecta*, each markedly higher than the genome-wide branch median values, which ranged from 0.066 to 0.125, fig. S1). It may be the case that gene redundancy has contributed to the divergence patterns for this gene. The *D. melanogaster Ilp* family has eight genes that have evolved a range of functions and tissue type specializations, including ovarian functions for *Ilp5* (Gronke, et al. 2010). In gene families, one gene member (or another) may pass through various stages of partial redundancy of function while still retaining the original gene function (Ohno 1970; Kirschner and Gerhart 1998), creating a window for relaxed constraint, and evolvability, of protein-coding regions and potentially giving rise to adaptive new functions or subfunctions, including gene-dosage functions (Kirschner and Gerhart 1998; Presgraves 2005; Conant and Wolfe 2008; Kuzmin, et al. 2022). Otherwise, without a function, the genes may accumulate mutations to form a pseudogene or become lost from the genome (Ohno 1970; Presgraves 2005; Conant and Wolfe 2008; Kuzmin, et al. 2022). Given the gene *Ilp5* was retained in the genomes of all five species here (table 2), it may be theorized that its rapid sequence divergence in four of five terminal branches may be associated with the evolution of altered functions that may affect ovariole formation. In fact, *llp5* has been linked to female re-mating frequency in *D. sechellia* (Wigby, et al. 2011), and thus sexual selection pressures could potentially contribute to its very high dN/dS in that branch (0.5843, table 2), and to the evolved decrease in ovariole numbers in this lineage (fig. 2). With respect to the

comparatively slow evolution of this gene in *D. melanogaster*, a study of *Ilp5* knockouts found evidence for compensatory expression of *Ilp3* (Gronke et al. 2010). *Ilp5* loss of function conditions cause a decrease in egg laying but do not completely eliminate it (Kumar, et al. 2020), suggesting its redundancy with other *Ilp* family members (and gene-dose compensation) may limit the extent of the negative effects on egg production. Nonetheless, the overall effect on egg yield and thus likely on fitness (Gronke, et al. 2010) may explain the strong purifying selection and low dN/dS value that we observed for the *D. melanogaster* branch, particularly if a weaker effect of *Ilp5* mutations occurred in the other species. In sum, *Ilp5* presents a dynamic history in the *melanogaster* subgroup and may substantively contribute to ovariole number variation in this taxon.

aquarius (*aqrs*)

The BULKSG ovariole-related gene *aquarius* (*aqrs*) (table 2) had the highest gene-wide dN/dS in the *D. yakuba* branch (0.3029), followed by *D. erecta* (0.1972), suggesting particularly rapid changes in the outgroup species (fig. 2). Aqrs protein has been associated with the functionality of the male seminal fluid proteins (SFPs) and the sex peptide (SP) in *D. melanogaster* (Findlay, et al. 2014), products that are transferred to the female reproductive tract during copulation and act to increase egg production and decrease female receptivity to courtship (Findlay, et al. 2014). The observed high levels of *aqrs* transcripts in larval somatic ovary cells (table 2) suggests that the gene also plays some role in these female reproductive cells. In this regard, *aqrs* may be a strong target for male-female sexual conflict, which is a factor known to cause rapid evolution of female (and male) proteins and of reproductive traits such as egg production, and may ultimately promote speciation (Rice 1996; Arnqvist, et al. 2000; Swanson and Vacquier 2002; Clark, et al. 2009). Thus, the rapid sequence divergence observed here for *aqrs* may putatively contribute to interspecies ovariole number variation.

SINGLEC Genes (table S7; fig. 4)

Ecdysone-inducible gene E1 (*ImpE1*), *sm*, and *CLIP-190*

In terms of the subset of rapidly evolving genes that we identified that were exclusively upregulated in TFs (TFa and/or TFp, and no other cell types, table S7 Notes), examples included *Ecdysone-inducible gene E1* (*ImpE1*), *smooth (sm*), and *Cytoplasmic linker protein 190* (*CLIP-190*), each of which exhibited positive selection in the *D. sechellia* branch (table S7). *ImpE1* has been linked to cell rearrangements during morphogenesis (Natzle, et al. 1988), *sm* is a nuclear ribonucleoprotein involved in diverse roles including muscle functions (Draper, et al. 2009), neuronal and chemosensory processes (Layalle, et al. 2005), and *CLIP-190* proteins coordinate binding between actins and microtubules (Lantz and Miller 1998), and may affect intracellular transport (Sanghavi, et al. 2012). Taken together, the rapidly evolving genes upregulated in TF cell types are involved in multiple cellular processes likely required for TF morphogenesis, and thus adaptive changes observed in these genes (fig. 4) may contribute to the proximate mechanisms underlying the rapid evolutionary divergence in ovariole numbers in *Drosophila*.

Genes downregulated in the BULKSG soma cells versus germ cells (table S7)

In terms of genes upregulated in the germ cells (and thus downregulated in the soma), the top ten genes with the highest degree of upregulation are shown in table S5. The genes exhibited elevated *tau* values ranging from 0.880 to 0.975, indicating a tendency of narrow transcription breadth (and several very narrow, with *tau* values above 0.950) for these genes, potentially reflecting high specialization to germ cell functions. Further, 8 of the 10 genes were largely uncharacterized genes (annotations as "CG number" only in table S5) based on available annotation from DAVID (Huang da, et al. 2009) and FlyBase (Gramates, et al. 2022). The two relatively well characterized genes *no long nerve cord* (*nolo*) and *orientation disruptor* (*ord*) play roles in the nerve cord, extracellular matrix, and meiosis (Huang da, et al. 2009). As the evidence to date suggests that somatic cells, rather than germ cells, are the principal regulators of ovariole number, we did not focus further on this particular genes of very highly and positively upregulated germ cell genes, and instead focused in the main text on the genes upregulated in somatic cells and with variable expression across larval ovary stages in tables 3 and 4.

BULKSG for three stages of larval ovary development

We also identified genes from the BULKSG dataset with respect to the data for three stages of larval ovary soma development (early, mid or late). The 30 genes with the highest degree of upregulation (log_2 change) in one stage (versus the other two) and that also exhibited rapid evolution (M0 dN /dS>0.20) are provided in table S6 (given the three types of comparisons, we included the top 30 genes, rather than ten as in tables 2, table S5). These genes were utilized in conjunction with sc-RNA (SINGLEC) in table 3.

Supplementary Results For The Three-Species Hawaiian *Drosophila* Clade

Table S10 contains the dN/dS results for the Hawaiian *Drosophila* orthologs of the 27 SIGNALC genes identified in table 1 from the *melanogaster* subgroup (Kumar, et al. 2020), and are described in the main text. Here, we note that six of the 27 studied genes of interest from the *melanogaster* subgroup (from table 1) had high confidence orthologs (reciprocal BLASTX) identified between *D. melanogaster-D. grimshawi* but not among all three species of *D. murphyi-D. sproati-D. grimshawi*. This may reflect rapid evolution of these genes in Hawaiian *Drosophila*, gene loss from the genome, and/or genes missing from the annotation (or lacking a complete CDS that were removed), although the latter has been largely mitigated given the high BUSCO scores from each genome (see the Supplementary Methods section above). Nonetheless, the data in table S10 infers rapid evolution of many of the SIGNALC genes (table S10), including positive selection, in the Hawaiian taxa (for those that had high confidence orthologs).

The ovariole related genes identified from the SINGLEC dataset in fig. 4 (and fig. S4, table S7), are less apt to share functions in Hawaiian *Drosophila* (than SIGNALC) given the fast evolution of gene expression of reproductive genes, including in *Drosophila* (Ranz, et al. 2003; Whittle and Extavour 2019). Thus, we excluded analyses of dN/dS and branch-site positive selection for this group pending future data on sc-RNA sequence or bulk-RNA seq of those cell types in *D. grimshawi* or other closely related Hawaiian taxa. Nonetheless, as cited in the main text, the genes that were highly upregulated in the *D. melanogaster* TF and SH cells (fig. 4, fig. S4, N values therein), were found to have the lowest frequency of high confidence ortholog gene sets identified in Hawaiian *Drosophila* among the nine cell types. The genes without orthologs included those from the *D. grimshawi-D. melanogaster* genome contrast (44.6% of all studied

genes in fig. 4without Hawaiian orthologs were from this class) and from the genome contrasts among the three Hawaiian species (55.4% of all studied genes without Hawaiian orthologs were from this class, fig. S6). This pattern suggests that the genes upregulated in the TF and SH cells (in *D. melanogaster*) as a group have evolved rapidly across the studied *Drosophila* species making orthologs unrecognizable, and/or have more commonly exhibited gene gains and losses, than genes from other cell types (Tautz and Domazet-Loso 2011; Tautz, et al. 2013), as described in the main text.

Supplementary References

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Research 25:3389-3402.

Amsellem V, Kryszke MH, Hervy M, Subra F, Athman R, Leh H, Brachet-Ducos C, Auclair C. 2005. The actin cytoskeleton-associated protein zyxin acts as a tumor suppressor in Ewing tumor cells. Experimental Cell Research 304:443-456.

Arnqvist G, Edvardsson M, Friberg U, Nilsson T. 2000. Sexual conflict promotes speciation in insects. Proceedings of the National Academy of Sciences of the United States of America 97:10460-10464.

Assis R, Zhou Q, Bachtrog D. 2012. Sex-biased transcriptome evolution in Drosophila. Genome Biology and Evolution 4:1189-1200.

Castresana J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol 17:540-552.

Chen J, Godt D, Gunsalus K, Kiss I, Goldberg M, Laski FA. 2001. Cofilin/ADF is required for cell motility during *Drosophila* ovary development and oogenesis. Nature Cell Biology 3:204-209.

Clark NL, Gasper J, Sekino M, Springer SA, Aquadro CF, Swanson WJ. 2009. Coevolution of interacting fertilization proteins. PLoS Genetics 5:e1000570.

Conant GC, Wolfe KH. 2008. Turning a hobby into a job: how duplicated genes find new functions. Nature Reviews: Genetics 9:938-950.

Cui Q, Purisima EO, Wang E. 2009. Protein evolution on a human signaling network. BMC Systems Biology 3:21.

Darfour-Oduro KA, Megens HJ, Roca AL, Groenen MA, Schook LB. 2016. Evolutionary patterns of Toll-like receptor signaling pathway genes in the Suidae. BMC Evolutionary Biology 16:33.

Draper I, Tabaka ME, Jackson FR, Salomon RN, Kopin AS. 2009. The evolutionarily conserved RNA binding protein SMOOTH is essential for maintaining normal muscle function. Fly (Austin) 3:235-246.

Edgar RC. (r20416 co-authors). 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics 5:113.

Findlay GD, Sitnik JL, Wang W, Aquadro CF, Clark NL, Wolfner MF. 2014. Evolutionary rate covariation identifies new members of a protein network required for *Drosophila* melanogaster female post-mating responses. PLoS Genetics 10:e1004108.

Goberdhan DC, Paricio N, Goodman EC, Mlodzik M, Wilson C. 1999. *Drosophila* tumor suppressor PTEN controls cell size and number by antagonizing the Chico/PI3-kinase signaling pathway. Genes Dev 13:3244-3258.

Gramates LS, Agapite J, Attrill H, Calvi BR, Crosby MA, Dos Santos G, Goodman JL, Goutte-Gattat D, Jenkins VK, Kaufman T, et al. 2022. FlyBase: a guided tour of highlighted features. Genetics 220.

Gramates LS, Marygold SJ, Santos GD, Urbano JM, Antonazzo G, Matthews BB, Rey AJ, Tabone CJ, Crosby MA, Emmert DB, et al. 2017. FlyBase at 25: looking to the future. Nucleic Acids Research 45:D663-D671.

Gronke S, Clarke DF, Broughton S, Andrews TD, Partridge L. 2010. Molecular evolution and functional characterization of *Drosophila* insulin-like peptides. PLoS Genetics 6:e1000857.

Hoff KJ, Stanke M. 2013. WebAUGUSTUS--a web service for training AUGUSTUS and predicting genes in eukaryotes. Nucleic Acids Research 41:W123-128.

Holt C, Yandell M. 2011. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. BMC Bioinformatics 12:491.

Huang da W, Sherman BT, Lempicki RA. 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nature protocols 4:44-57.

Huang J, Wu S, Barrera J, Matthews K, Pan D. (p07161 co-authors). 2005. The Hippo signaling pathway coordinately regulates cell proliferation and apoptosis by inactivating Yorkie, the *Drosophila* Homolog of YAP. Cell 122:421-434.

Kim BY, Wang JR, Miller DE, Barmina O, Delaney E, Thompson A, Comeault AA, Peede D, D'Agostino ER, Pelaez J, et al. 2021. Highly contiguous assemblies of 101 drosophilid genomes. Elife 10.

King RC, Aggarwal SK, Aggarwal U. 1968. The development of the female *Drosophila* reproductive system. Journal of Morphology 124:143-166.

Kirschner M, Gerhart J. 1998. Evolvability. Proceedings of the National Academy of Sciences of the United States of America 95:8420-8427.

Kumar S, Stecher G, Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Molecular Biology and Evolution 33:1870-1874.

Kumar T, Blondel L, Extavour CG. 2020. Topology-driven protein-protein interaction network analysis detects genetic sub-networks regulating reproductive capacity. Elife 9.

Kuzmin E, Taylor JS, Boone C. 2022. Retention of duplicated genes in evolution. Trends in Genetics 38:59-72.

Lantz VA, Miller KG. 1998. A class VI unconventional myosin is associated with a homologue of a microtubule-binding protein, cytoplasmic linker protein-170, in neurons and at the posterior pole of *Drosophila* embryos. Journal of Cell Biology 140:897-910.

Larracuente AM, Sackton TB, Greenberg AJ, Wong A, Singh ND, Sturgill D, Zhang Y, Oliver B, Clark AG. 2008. Evolution of protein-coding genes in *Drosophila*. Trends in Genetics 24:114-123.

Layalle S, Coessens E, Ghysen A, Dambly-Chaudiere C. 2005. Smooth, a hnRNP encoding gene, controls axonal navigation in *Drosophila*. Genes to Cells 10:119-125.

Li JJ, Huang H, Bickel PJ, Brenner SE. 2014. Comparison of *D. melanogaster* and *C. elegans* developmental stages, tissues, and cells by modENCODE RNA-seq data. Genome Research 24:1086-1101.

Lobell AS, Kaspari RR, Serrano Negron YL, Harbison ST. 2017. The Genetic Architecture of Ovariole Number in *Drosophila melanogaster*: Genes with Major, Quantitative, and Pleiotropic Effects. G3 (Bethesda) 7:2391-2403.

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology 15:550.

Mank JE, Ellegren H. 2009. Are sex-biased genes more dispensable? Biology letters 5:409-412.

Manni M, Berkeley MR, Seppey M, Simao FA, Zdobnov EM. 2021. BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. Mol Biol Evol 38:4647-4654.

Masalia RR, Bewick AJ, Burke JM. 2017. Connectivity in gene coexpression networks negatively correlates with rates of molecular evolution in flowering plants. PloS One 12:e0182289.

Matsui Y, Lai ZC. 2013. Mutual regulation between Hippo signaling and actin cytoskeleton. Protein Cell 4:904-910.

McDonald JH, Kreitman M. 1991. Adaptive protein evolution at the Adh locus in Drosophila. Nature 351:652-654.

Meisel RP. 2011. Towards a more nuanced understanding of the relationship between sex-biased gene expression and rates of protein-coding sequence evolution. Molecular Biology and Evolution 28:1893-1900.

Murga-Moreno J, Coronado-Zamora M, Hervas S, Casillas S, Barbadilla A. 2019. iMKT: the integrative McDonald and Kreitman test. Nucleic Acids Research 47:W283-W288.

Natzle JE, Fristrom DK, Fristrom JW. 1988. Genes expressed during imaginal disc morphogenesis: IMP-E1, a gene associated with epithelial cell rearrangement. Developmental Biology 129:428-438.

Ohno S. 1970. Evolution by gene duplication. New York: Springer Verlag.

Otto SP. 2004. Two steps forward, one step back: the pleiotropic effects of favoured alleles. Proceedings: Biological Sciences 271:705-714.

Presgraves DC. 2005. Evolutionary genomics: new genes for new jobs. Current Biology 15:R52-53.

Ranz JM, Castillo-Davis CI, Meiklejohn CD, Hartl DL. 2003. Sex-dependent gene expression and evolution of the *Drosophila* transcriptome. Science 300:1742-1745.

Rauskolb C, Pan G, Reddy BV, Oh H, Irvine KD. 2011. Zyxin links fat signaling to the hippo pathway. PLoS Biology 9:e1000624.

Rice WR. 1996. Sexually antagonistic male adaptation triggered by experimental arrest of female evolution. Nature 381:232-234.

Richardson BE, Lehmann R. 2010. Mechanisms guiding primordial germ cell migration: strategies from different organisms. Nature Reviews: Molecular Cell Biology 11:37-49.

Sanghavi P, Lu S, Gonsalvez GB. 2012. A functional link between localized Oskar, dynamic microtubules, and endocytosis. Developmental Biology 367:66-77.

Sarikaya DP, Belay AA, Ahuja A, Dorta A, Green DA, 2nd, Extavour CG. 2012. The roles of cell size and cell number in determining ovariole number in *Drosophila*. Developmental Biology 363:279-289.

Sarikaya DP, Church SH, Lagomarsino LP, Magnacca KN, Montgomery SL, Price DK, Kaneshiro KY, Extavour CG. 2019. Reproductive Capacity Evolves in Response to Ecology through Common Changes in Cell Number in Hawaiian Drosophila. Current Biology 29:1877-1884.e1876.

Sarikaya DP, Extavour CG. 2015. The Hippo pathway regulates homeostatic growth of stem cell niche precursors in the *Drosophila* ovary. PLoS Genetics 11:e1004962.

Satija R, Farrell JA, Gennert D, Schier AF, Regev A. 2015. Spatial reconstruction of single-cell gene expression data. Nature Biotechnology 33:495-502.

Seppey M, Manni M, Zdobnov EM. 2019. BUSCO: Assessing Genome Assembly and Annotation Completeness. Methods in Molecular Biology 1962:227-245.

Slaidina M, Banisch TU, Gupta S, Lehmann R. 2020. A single-cell atlas of the developing Drosophila ovary identifies follicle stem cell progenitors. Genes Dev 34:239-249.

Slaidina M, Gupta S, Banisch TU, Lehmann R. 2021. A single-cell atlas reveals unanticipated cell type complexity in *Drosophila* ovaries. Genome Research 31:1938-1951.

Stanke M, Waack S. 2003. Gene prediction with a hidden Markov model and a new intron submodel. Bioinformatics 19 Suppl 2:ii215-225.

Stanley CE, Jr., Kulathinal RJ. 2016. flyDIVaS: A Comparative Genomics Resource for Drosophila Divergence and Selection. G3 (Bethesda) 6:2355-2363.

Starmer W, Polak M, Pitnick S, McEvey S, Barke rJ, Wolf L. 2003. Phylogenetic, Geographical, and Temporal Analysis of Female Reproductive Trade-Offs in Drosophilidae. Evolutionary Biology 33:139-171.

Suvorov A, Kim BY, Wang J, Armstrong EE, Peede D, D'Agostino ERR, Price DK, Waddell P, Lang M, Courtier-Orgogozo V, et al. 2022. Widespread introgression across a phylogeny of 155 *Drosophila* genomes. Current Biology 32:111-123 e115.

Swanson WJ, Vacquier VD. 2002. The rapid evolution of reproductive proteins. Nature Reviews: Genetics 3:137-144.

Talavera G, Castresana J. (r20429 co-authors). 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Systematic Biology 56:564-577.

Tarikere S, Ylla G, Extavour CG. 2022. Distinct gene expression dynamics in germ line and somatic tissue during ovariole morphogenesis in *Drosophila melanogaster*. G3 (Bethesda) 12: DOI: 10.1093/g1093journal/jkab1305

Tautz D, Domazet-Loso T. 2011. The evolutionary origin of orphan genes. Nature Reviews Genetics 12:692-702.

Tautz D, Neme R, Domazet-Loso T. 2013. Evolutionary Origin of Orphan Genes. eLS.

Tu R, Duan B, Song X, Chen S, Scott A, Hall K, Blanck J, DeGraffenreid D, Li H, Perera A, et al. 2021. Multiple Niche Compartments Orchestrate Stepwise Germline Stem Cell Progeny Differentiation. Current Biology 31:827-839 e823.

von Stein W, Ramrath A, Grimm A, Muller-Borg M, Wodarz A. 2005. Direct association of Bazooka/PAR-3 with the lipid phosphatase PTEN reveals a link between the PAR/aPKC complex and phosphoinositide signaling. Development 132:1675-1686.

Whittle CA, Extavour CG. 2019. Selection shapes turnover and magnitude of sex-biased expression in *Drosophila* gonads. BMC Evolutionary Biology 19:60 (doi: 10.1186/s12862-12019-11377-12864).

Whittle CA, Kulkarni A, Extavour CG. 2021. Evolutionary dynamics of sex-biased genes expressed in cricket brains and gonads. Journal of Evolutionary Biology 34:1188-1211.

Wigby S, Slack C, Gronke S, Martinez P, Calboli FC, Chapman T, Partridge L. 2011. Insulin signalling regulates remating in female *Drosophila*. Proceedings: Biological Sciences 278:424-431.

Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Molecular Biology and Evolution 24:1586-1591.