

1 The evolution of ovary-biased gene expression in Hawaiian
2 *Drosophila*

3 Samuel H. Church^{1,2,*} Catriona Munro³ Casey W. Dunn²
4 Cassandra G. Extavour^{1,4,5}

5 ¹ Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA 02138, USA

6 ² Current address: Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT 06520,
7 USA

8 ³ Collège de France, PSL Research University, CNRS, Inserm, Center for Interdisciplinary Research in
9 Biology, 75005 Paris, France

10 ⁴ Department of Molecular and Cellular Biology, Harvard University, Cambridge, MA 02138, USA

11 ⁵ Howard Hughes Medical Institute, Chevy Chase, MD 20815

12 * corresponding author: samuelhchurch@gmail.com

13 **1 Abstract**

14 With detailed data on gene expression accessible from an increasingly broad array of species, we can test the
15 extent to which our developmental genetic knowledge from model organisms predicts expression patterns and
16 variation across species. But to know when differences in gene expression across species are significant, we
17 first need to know how much evolutionary variation in gene expression we expect to observe. Here we provide
18 an answer by analyzing RNAseq data across twelve species of Hawaiian Drosophilidae flies, focusing on gene
19 expression differences between the ovary and other tissues. We show that over evolutionary time, there exists
20 a cohort of ovary specific genes that is stable and that largely corresponds to described expression patterns
21 from laboratory model *Drosophila* species. Our results also provide a demonstration of the prediction that,
22 as phylogenetic distance increases, variation between species overwhelms variation between tissue types.
23 Using ancestral state reconstruction of expression, we describe the distribution of evolutionary changes in
24 tissue-biased expression, and use this to identify gains and losses of ovary-biased expression across these
25 twelve species. We then use this distribution to calculate the evolutionary correlation in expression changes
26 between genes, and demonstrate that genes with known interactions in *D. melanogaster* are significantly more
27 correlated in their evolution than genes with no or unknown interactions. Finally, we use this correlation
28 matrix to infer new networks of genes that share evolutionary trajectories, and we present these results as a
29 dataset of new testable hypotheses about genetic roles and interactions in the function and evolution of the
30 *Drosophila* ovary.

31 2 Introduction

32 Data on when and where genes are expressed are now fundamental to the study of development and disease¹.
33 With continually advancing RNA sequencing technologies, these data have been collected using RNA sequencing
34 from a wide variety of cells, treatments and species^{2,3}. Statistical analysis of gene expression across these
35 differentials generates insights into how gene expression is connected to phenotypic differences in morphology
36 and behavior⁴. However, when comparing gene expression across species, most studies have been restricted
37 to pairwise comparisons, often between one model laboratory species and one other species of interest⁵. One
38 challenge with such pairwise comparisons is that they lack robust information about how much evolutionary
39 variation in expression we expect to observe, making it difficult to evaluate the significance of any inter-
40 specific difference in variation^{5,6}. Instead, we need phylogenetic comparisons of expression that take into
41 account the shared history between species^{7,8}, and that describe significant changes in expression in relation
42 to other phenotypic traits of interest.⁹ In this study we perform a phylogenetic comparison of gene expression
43 across the organs of twelve species of Hawaiian Drosophilidae flies with highly divergent ovary and egg
44 morphologies. From our results we identify individual genes that have undergone significant evolutionary
45 shifts in organ-specific expression, and describe global patterns in transcriptome variation across species that
46 can serve as a benchmark for future interspecific comparisons of gene expression.

47 Phylogenetic comparisons of developmental traits are particularly valuable for building context around com-
48 parisons between well-studied model organisms and their non-model relatives¹⁰. Much more has been learned
49 about the genetics and development of laboratory model species like *D. melanogaster* than may ever be pos-
50 sible for the vast majority of life¹¹. But the usefulness of model species to understand general principles
51 depends in part on the extent to which biology in these species reflects the biology of other taxa, rather
52 than species-specific phenomena¹². In the case of gene expression, there has been substantial debate about
53 the degree to which patterns observed in model organisms may be representative across species^{13–16}. Where
54 several studies showed that the expression profiles of organs within a species are more different the pro-
55 files of homologous organs across species^{17–20}, other work has questioned this finding^{13,14}. More recently,
56 Breschi and colleagues (2016)²¹ demonstrated that, consistent with an evolutionary model of trait evolu-
57 tion, species-level variation in gene expression increases with the time since divergence from the most recent
58 common ancestor. In addition, previous work by authors on this manuscript⁸ showed that, while expression
59 patterns across tissues tend to be consistent between species, lineage-specific shifts in expression enrichment
60 can be identified by applying phylogenetic comparative methods. With the exception of the work by Munro
61 and colleagues (2021)⁸, these studies have been, to our knowledge, performed almost exclusively in vertebrate
62 species^{17,18,20}, and for the most part placental mammals^{13,14,16}, meaning that far less is known about organ
63 and species-level expression differences when comparing across the tree of life.

64 The detailed atlases of expression data across organs²² and developmental timepoints²³ is one of the strengths
65 of model systems like *D. melanogaster*. These public resources make it possible to explore global patterns of
66 expression to gain insight into potential gene regulation, interaction, and function^{23–25}. As atlases such as
67 these have become increasingly detailed and available from more taxa, a new goal has been to compare these
68 expression profiles across species^{7,26,27}. One objective of these cross-species comparisons is to shed light on
69 potential regulatory associations between genes^{7,9}. This is especially advantageous for complex processes
70 such as ovarian function for which we have a fragmented understanding of gene regulation despite genetic
71 and transcriptome studies within single model organisms. Another objective of phylogenetic comparisons
72 of expression atlases is to estimate the evolutionary distance between species at which we might expect a
73 given gene to demonstrate a divergent pattern of expression⁶. If this distance is relatively small, then we
74 predict atlases to contain large amounts of species-specific patterns. Alternatively, if as described above,
75 variation across tissues outweighs variation across the species being compared, we predict atlases to contain
76 large cohorts of tissue-specific genes that have been evolutionarily conserved. In this study we test for the
77 existence of a core suite of ovary-specific genes across species of Hawaiian Drosophilidae and describe its size
78 and composition in relation to the described atlas of expression in *D. melanogaster*.

79 The *Drosophila* ovary has several features²⁸ that make it a compelling organ in which to test hypotheses
80 about expression evolution. Analyses of the FlyAtlas2 dataset²⁹ show that in *D. melanogaster*, more genes
81 demonstrate highest expression enrichment in the ovary than any other adult female organ (Fig. S1). Ad-

82 ditionally, all described signaling pathways are known to have a role in regulating ovarian development³⁰.
83 The ovary performs several critical functions, including maintaining the germ line and manufacturing spe-
84 cialized egg cells, yolk, and egg-shell materials³¹. Genetic screens^{30,32} and experimental manipulation in *D.*
85 *melanogaster* have revealed functions of many genes involved in these processes, including yolk-protein genes
86 required for oogenesis³³ and embryonic patterning genes with localized mRNA like *nanos*³⁴ and *swallow*³⁵.
87 Here we compare whole-ovary RNA profiles to assess the extent to which these genes and others demonstrate
88 consistent patterns of ovary-enrichment over evolutionary timescales in a clade with highly divergent ovary
89 and egg morphologies.

90 The Hawaiian Drosophilidae clade contains an estimated 1,000 extant species³⁶ that diverged from a common
91 ancestor with *D. melanogaster* between 25 and 40 million years ago³⁷. Extant species have been studied
92 in particular for the variation in ovary and egg morphology^{38,39}. Species of Hawaiian Drosophilidae show
93 the largest range within the family of egg size, shape, and the number of egg-producing units in the ovary,
94 known as ovarioles⁴⁰⁻⁴². Previous studies by our research group and others have shown that these traits are
95 likely associated with evolutionary changes in the egg-laying substrate (e.g. rotting bark, flowers, leaves)^{38,40}.
96 Furthermore, our previous work demonstrated that at least one developmental process, governing how the
97 number of ovarioles is specified in the adult *D. melanogaster* ovary, is conserved in Hawaiian *Drosophila*⁴⁰.
98 The diversity of Hawaiian species and their relationship to model species make them a strong candidate
99 model clade for evo-devo research^{36,43}. However, their relatively long generation times and species-specific
100 breeding requirements make laboratory culture more challenging than classic *Drosophila* models³⁶. In this
101 study we leverage technologies that can be deployed on wild-caught individuals to gather rich developmental
102 data to compare across species.

103 We compared the expression profiles of twelve species of wild-caught Hawaiian Drosophilidae species across
104 three body parts: the adult ovary, head, and the remaining carcass (Fig. 1). We use these tissues to make
105 two comparisons for calculating differential expression: one between the ovary and the carcass, and the
106 other between the head and the carcass. These comparisons allow us to assess ovary-specific and head-
107 specific gene expression over evolutionary time, both of which are relevant to fundamental questions in
108 Hawaiian Drosophilidae biology. Given that the ovary and head are not equivalent body parts in terms of
109 functional complexity (the ovary is primarily dedicated to producing oocytes, while the head contains the
110 eyes, brain, and mouthparts, all dedicated to different tasks), we present these analyses in parallel, and do
111 not draw conclusions based on direct comparisons between the head and ovary. In our description of results
112 we prioritize the ovary-carcass comparison.

113 For each analyses, we first characterized the differentially expressed genes in the ovary of each species
114 individually. By comparing these to each other, and to records of ovary-enriched and head-enriched genes
115 from *D. melanogaster*, we identified a core suite of tissue-specific genes shared across species. We applied
116 linear modeling to this dataset to test the overall contribution of species- and tissue-level differences to
117 expression variation across genes, and describe the circumstances under which one is likely to dominate over
118 the other. Finally, we used a phylogenetic analysis of expression changes over evolutionary time to identify
119 genes likely to have gained and lost tissue-enriched expression. This evolutionary screen of expression changes
120 allowed us to identify networks of genes that demonstrate correlated changes in expression evolution. We
121 provide these networks as a searchable dataset of novel, testable hypotheses for gene regulation with respect
122 to ovarian function. The results of this study demonstrate both the power of Hawaiian *Drosophila* as a
123 model clade for evo-devo, and the potential of using phylogenetic methods to identify evolutionary variation
124 in gene expression underlying phenotypic differences.

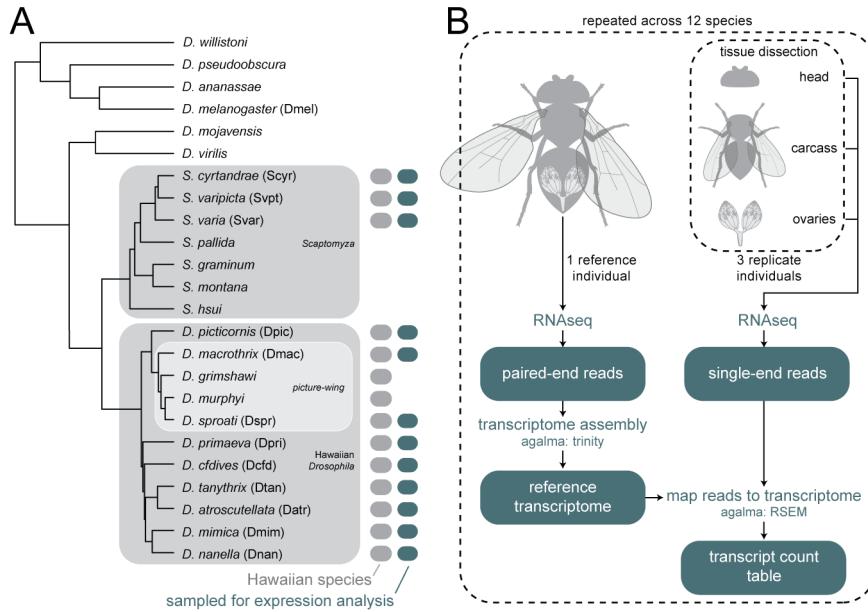


Figure 1: **Phylogeny of species and RNA sampling strategy.** A, Twelve species of Hawaiian Drosophilidae flies were collected in the wild and processed for RNA sequencing. The twelve reference transcriptomes assembled from these species were combined with twelve published genomes to generate the phylogeny shown here (originally published in Church and Extavour, 2021⁴⁴). Three clades within the group are highlighted: the genus *Scaptomyza*, nested within the paraphyletic genus *Drosophila*; the Hawaiian *Drosophila*, which, along with *Scaptomyza*, make up the Hawaiian Drosophilidae; and the well-known *picture-wing* clade. Adjacent to tip labels are four letter species codes used throughout the manuscript. B, The experimental design used to generate the data in this manuscript. When sufficient specimens were available per species, one whole individual was used as a reference and three whole individuals were dissected into three separate tissues: the head, ovaries, and all remaining material (carcass). Reference individuals were sequenced to generate paired-end RNA reads and dissected tissues were sequenced to generate single-end RNA reads. Tissue libraries were then mapped to the assembled reference to quantify transcript expression. Teal boxes indicate data files. Dashed-line boxes indicate a repeated step.

125 3 Results

126 3.1 Differential gene expression reveals a cohort of consistently ovary-specific 127 genes

128 We observed several patterns in tissue-specific gene expression that are consistent across all twelve species.
129 First, in all species the main axis of variation separated ovary RNA libraries from head and carcass (Fig.
130 S4). In all species this axis accounted for at least 50% of variation, and in several species greater than 70%
131 of variation. To test for possible variation due to different runs on the sequencer, we resequenced several
132 libraries and compared them using principle component analysis. We found variation between sequencing
133 runs to be negligible compared to variation across tissues and individuals (Fig. S7). Second, in all species
134 we observed that there was a larger amount of significantly downregulated transcripts than upregulated in
135 the ovary relative to the carcass (Fig. 2A-B, S5). Across species, we observed an average of 27.7% to be
136 significantly downregulated and 15.5% of transcripts to be significantly upregulated. When comparing the
137 head to the carcass, we observed an average of 10% of transcripts to be significantly upregulated and 10.5%
138 to be significantly downregulated (Fig. S6). These differences may reflect variation in the complexity and
139 diversity of functions of the tissues being compared.

140 We used the results of our differential gene expression analysis within species to test for the existence of a
141 suite of genes that show consistent ovary-specific expression across species. We found a cohort of 131 genes,
142 grouped according to BLAST sequence similarity to *D. melanogaster*, for which at least one transcript was
143 significantly upregulated in the ovaries of more than ten species (Fig. 2C). Transcripts matching these
144 genes made up on average 24.6% of the significantly ovary-upregulated transcripts across species, meaning
145 roughly one quarter of ovary-specific genes have conserved expression patterns over evolutionary time. When
146 excluding the species *S. varia*, this average decreased to 17.7%, as this species has the smallest set of ovary-
147 upregulated transcripts, 100% of which match core ovary genes.

148 We then tested the extent to which these core ovary genes correspond to observations in well-studied labor-
149 atory *Drosophila* models. To accomplish this, we compared expression across Hawaiian species to reported
150 tissue-specific expression levels from *D. melanogaster*²⁹. We found that Hawaiian core ovary-specific genes
151 show nearly universal enrichment in the ovary of *D. melanogaster* as well, as reported in the FlyAtlas2
152 dataset²⁹ (Fig. 2D). We likewise observed that genes reported in *D. melanogaster* to have highest enrich-
153 ment in the ovary largely correspond to genes that are significantly upregulated in the ovaries of Hawaiian
154 species (Fig. S8).

155 The 131 core ovary genes include several well-known members involved in oogenesis and germline stem cell
156 renewal such as *nanos*³⁴, *swallow*³⁵, and *oskar*⁴⁵ (Fig. 2E). We found only two genes that were identified as
157 Hawaiian core ovary genes that are not reported in the FlyAtlas2 dataset²⁹ to be enriched in the ovary of *D.*
158 *melanogaster*: the SET domain binding factor *sbf*, and *Rfx*, which are reported to be enriched in the heart,
159 brain, and other non-reproductive tissues²⁹.

160 We used the same approach to identify a core suite of 52 head-specific genes (Fig. S9). There was no
161 overlap between the sets of core head genes and core ovary genes. To test whether the correspondence
162 between expression observations in Hawaiian flies and *D. melanogaster* might be due to factors beyond
163 tissue identity, we compared head expression values to ovary enrichment data from *D. melanogaster*, as we
164 had done for ovary expression values above. We did not observe a correspondence in either direction between
165 expression in the head of Hawaiian species and enrichment in the ovary of *D. melanogaster* (Fig. S10A).
166 In contrast, we did find a correspondence between head-specific expression and genes enriched in the *D.*
167 *melanogaster* brain, eye, and head (Fig. S10B). Core head genes include *Rhodopsin* photoreceptor genes and
168 genes such as *hikaru genki* with involvement in synaptic centers⁴⁶.

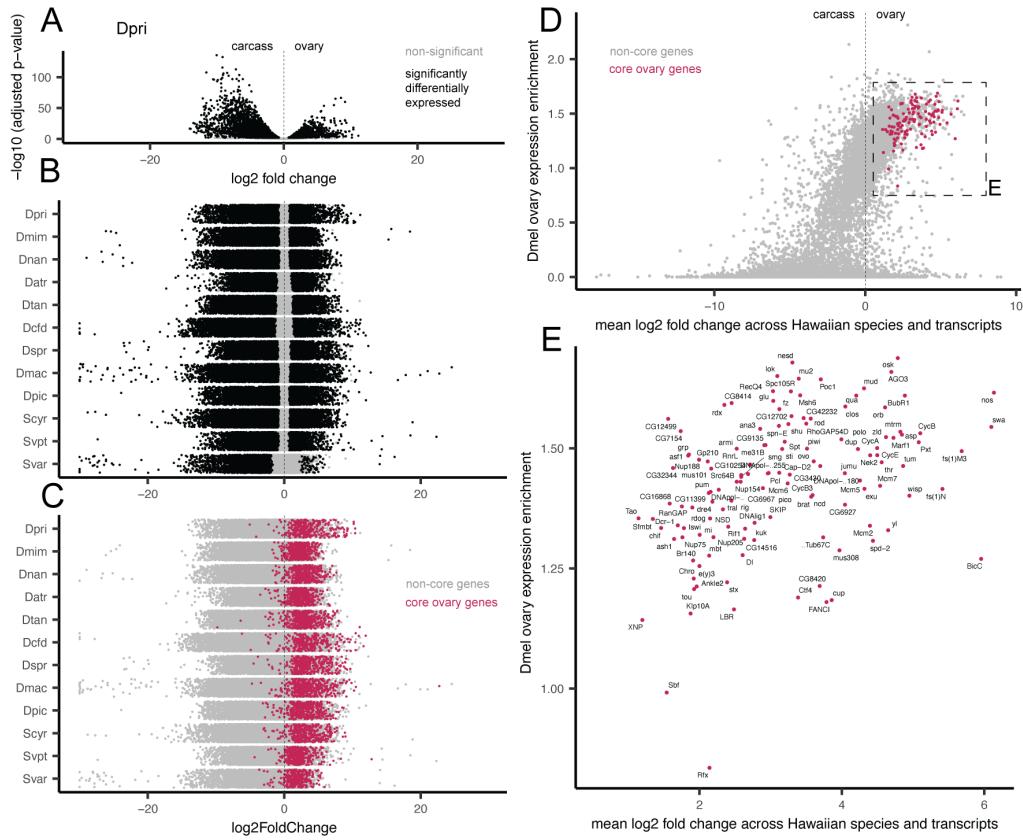


Figure 2: Identifying a cohort of ovary-specific genes across drosophilid species. A, Volcano plot for one example species, *D. primaeva* (Dpri), showing the results of a differential gene expression analysis comparing the ovary to the carcass. The x-axis shows the log₂ fold change of expression across transcripts, and the y-axis shows the adjusted p-value, log₁₀ transformed. Points that are significantly differentially expressed are shown in black. B, Jitter plots showing the results of the same analysis across the twelve species studied here. The x-axis shows the log₂ fold change of expression across transcripts, and points are arranged with random jitter within species on the y-axis. C, The same jitter plots as in B, now colored according to whether or not transcripts belong to a cohort of core ovary genes. These are defined as genes, grouped by BLAST similarity to *D. melanogaster* transcripts, for which at least one transcript is upregulated in the ovary of ten or more of the twelve species. D, A comparison of mean expression change across Hawaiian species to reported ovary-enrichment values from *D. melanogaster*, as reported in FlyAtlas2²⁹. Core ovary genes are marked in magenta. E, The boxed region shown in D, magnified and now showing only core ovary genes, annotated with the gene symbol from *D. melanogaster*.

169 **3.2 Modeling reveals the phylogenetic decay of expression similarity between**
 170 **tissues**

171 Many studies have investigated the question of whether we expect expression to be more similar across the
 172 same organ in different species, or across different organs within the same species^{13–20}. Recent studies have
 173 suggested that the answer to this question will depend on the phylogenetic distance separating the species
 174 being compared²¹. Here we used a modeling approach to investigate this question with respect to the ovaries
 175 of Hawaiian drosophilids.

176 First, we determined an appropriate unit of comparison across species, based on an assessment of homologous
 177 features between reference transcriptomes. The agalma pipeline provides a method for determining

178 homologous and orthologous sequences using an all-by-all BLAST approach to determine clusters of reciprocally similar sequences (homology groups). These can then be divided into orthology groups by estimating
179 gene trees and identifying maximally inclusive subtrees with no more than one sequence per taxon⁴⁷. We
180 compared the representation of species across homology and orthology groups, and observed that while the
181 representation of homology groups increases with the number of species compared, representation of orthology
182 groups decreases (Fig. S11). This is a known obstacle in comparative transcriptomics, attributed to
183 many transcripts being artificially fragmented during reference transcriptome assembly⁴⁸. To reduce the
184 impact of this on our downstream analyses, we averaged TPM values across all transcripts within a homology
185 group for each sequenced RNA library. Principle component analysis of this average expression dataset
186 showed that the first principle component divides ovary libraries from the rest, while the second component
187 separates samples along an axis that largely corresponds to phylogenetic distance between species (Fig.
188 S12). While this averaging approach reduces noise due to variable mapping affinities of fragments of the
189 same transcript, it comes at the cost of averaging over potential variation between genuine transcripts that
190 fall into the same homology group. Future analyses using improved assemblies for transcriptomes or genomes
191 will likely be able to avoid this trade off and compare transcript counts directly. To test the robustness of
192 results to this averaging, we also performed key analyses over a dataset of the identifiable strict orthologs.
193
194 With average expression counts for homologous transcripts across species, we tested the degree to which
195 variation across this dataset could be attributed to tissue-specific variation (here, ovary vs. carcass), species-
196 specific variation, or neither (residual variation). Using the linear modeling approach adapted from Breschi
197 and colleagues (2016)²¹, we found the proportion of variance across the dataset attributed to tissue differences
198 decreased with phylogenetic distance, while the proportion attributed to species difference increased (Fig.
199 3A-C). In addition, we found that, when comparing ovary and carcass tissues, the Hawaiian drosophilid clade
200 encompasses the crossover point where variation across species swamps variation across tissues (crossed lines,
201 Fig. 3A). When comparing across the two species from the *picture-wing* group included in this study, an
202 average of 45.6% of the variation can be attributed to tissue differences. For the same comparison, 960 genes
203 were identified as tissue-variable genes (TVGs), defined as residual variation accounting for <25% and a two-
204 fold increase in variation attributed to tissues than to species (Fig. 3B, S13). In contrast, when comparing
205 across all twelve Hawaiian drosophilid species studied here, 34.7% of the variation can be attributed to tissue,
206 with 240 TVGs (Fig. 3B, S13). Across different clades of comparisons, the number of species-variable genes
207 (SVGs) remains relatively stable (from 304 to 260, Fig. 3B).
208
209 We then leveraged the results of this linear modeling approach across all twelve species to perform an
210 additional screen for genes that are consistently upregulated in ovaries across species. We compared the
211 proportion of variation explained by tissue for each homology group to the average log₂ fold change from the
212 results of our differential gene expression analysis (Fig 3D). This comparison allowed us to identify genes
213 that fall above our threshold for TVGs that are also upregulated in the ovary (Fig 3E). This group of genes
214 includes many of the same members as the core ovary genes (e.g. *nanos* and *swallow*), as well as several new
215 candidates (e.g. *singed*).
216
217 To test the importance of tissue identity, we repeated the same analysis comparing variation across species
218 and tissues using the head in place of the ovary. Consistent with what we describe for the ovary and carcass,
219 as phylogenetic distance increases the proportion of variation across tissues decreases while variation across
220 species increases. In contrast to the above findings, however, for the head and carcass far less of the variation
221 in gene expression can be attributed to tissue differences (Fig. S14). For these tissues, the crossover point
222 between total proportion of variation occurs roughly at the distance separating the two *picture-wing* species.
223
224 To verify these results were not driven by the species *S. varia*, which had the most distinct expression
225 patterns of all species, we repeated these analyses excluding this species and recovered largely equivalent
226 results (Fig. S15). To test robustness to homology group averaging, we repeated this analysis over strict
227 orthologs, again recovering the same results (S16). Finally, we also compared our findings to those that
228 would be recovered using a more typical pairwise approach, by repeating the linear modeling analysis on
229 ovary and carcass data using every pairwise combination of the twelve species. We recovered the same trend
230 of decreasing contribution of tissue-level variation with increasing phylogenetic distance. While several pairs
231 of species show more variation between species than tissues, we note that not every pair, nor the average
232 across pairs, captures the crossover point where variation across species overwhelms variation across tissues

²³⁰ (Fig. S17). This reflects the inherent variability between individual pairwise comparisons of species, and
²³¹ highlights the importance of phylogenetic analytical approaches on entire clades.

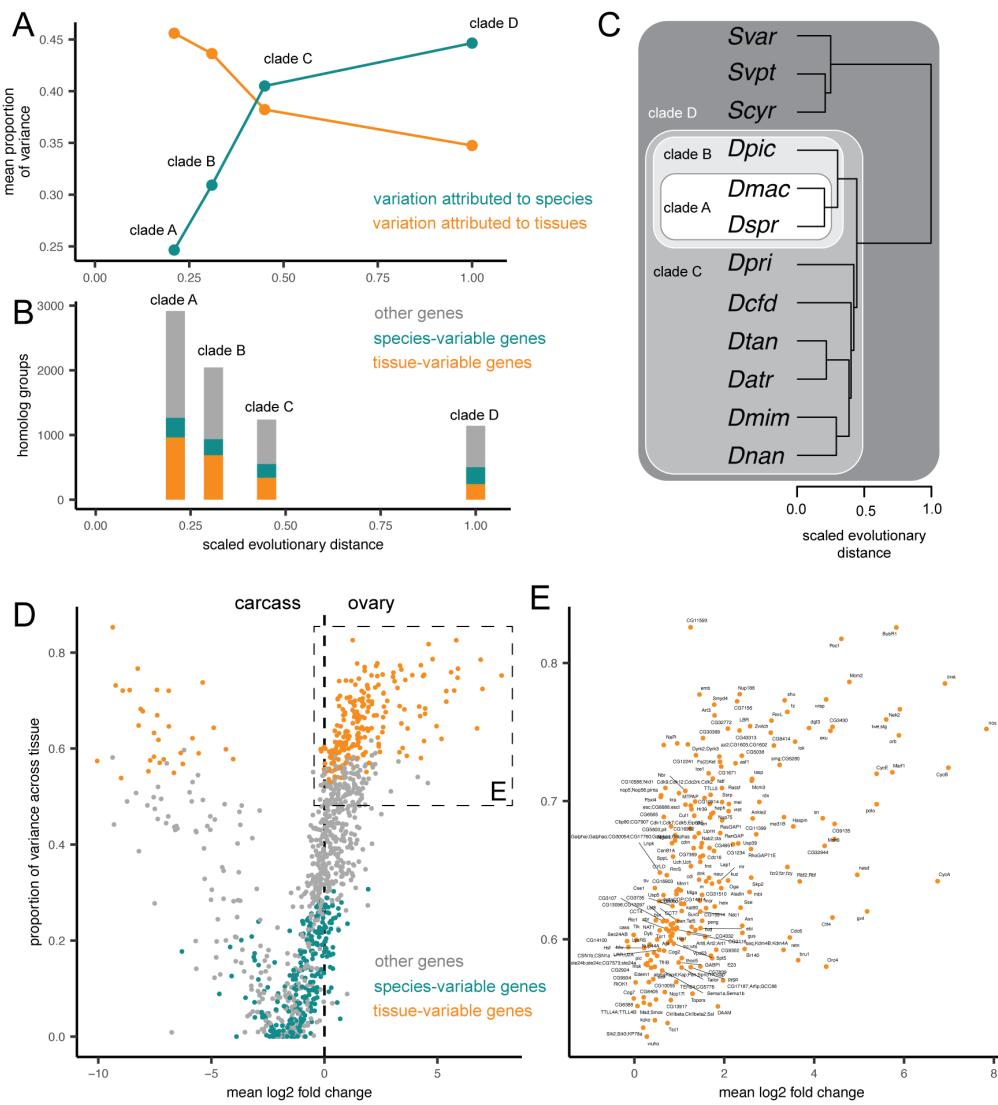


Figure 3: Linear modeling shows the proportion of variance explained by differences across tissues and species. A, The results of a linear modeling approach to calculate expression variation for each gene, attributed to variation across organs, species, or residual variation, as described in Breschi and colleagues (2016)²¹. The average proportion of variation attributed to tissues is higher than that attributed to species for the two *picture-wing* species in clade A, while the opposite is true for all twelve species in clade D. Sample sizes are for clade A: 2,918 homology groups, clade B: 2,044, clade C: 1,239, and clade D: 1,143. B, The number of genes, defined by homology group, classified as tissue variable genes (TVGs), species variable genes (SVGs), or neither in each clade comparison. C, The phylogeny of the twelve species studied here, showing the four clades compared in A-B. Scaled evolutionary distance is calculated as the relative distance from the most recent common ancestor of Hawaiian drosophilids to extant species. D, Comparing results of the differential gene expression approach (\log_2 fold change) on the x-axis to results of the modeling approach on the y-axis (variation across tissues). Genes are colored according to TVGs and SVGs. The inset box highlights TVGs that are upregulated in the ovary relative to the carcass. E, The same plot, now showing only upregulated TVGs, annotated with the gene symbol from the *D. melanogaster* sequences in the same homology group.

232 3.3 Identifying gains and losses of ovary bias across genes and the phylogeny

233 While many ovary-specific transcripts belong to the cohort of core ovary genes, on average 75.4% of transcripts
234 are upregulated in the ovaries of one or several species, but not consistently across ten or more of the species
235 studied here (Fig 2B-C). This is suggestive of many evolutionary gains and losses of ovary-specific expression
236 of genes. We characterized the evolution of these gains and losses using an ancestral state reconstruction
237 approach. First we quantified expression bias between tissues as the ratio of read counts⁷, then reconstructed
238 the value of this continuous trait for each gene (defined using homology groups) at each node of the estimated
239 species tree (Fig. S3). We then calculated the scaled change of expression bias along each branch, which
240 allowed us to describe how relative expression values between tissues had changed the course of evolutionary
241 time (Fig 4A). Visualizing the distribution of scaled changes by genes shows that most scaled changes are
242 small and centered around zero, representing little change in gene expression bias between tissues (Fig.
243 4B-C).

244 Using this dataset of scaled changes across genes and branches, we identified branches for which the direction
245 of tissue bias had changed (e.g. from higher expression in the ovary than in the carcass to lower, or vice
246 versa). Visualizing this dataset according to branches reveals that the majority of these changes in bias are
247 located on the root and terminal branches, rather than internal branches (Fig 4D-E). This is likely because
248 internal branches for this rapid radiation tend to be very short; even when scaling evolutionary changes to
249 branch length, it is less probable for our analysis to identify a shift to and from ovary-biased expression on
250 a short branch than a long branch. Repeating the same analysis on a dataset of strict orthologs reflects the
251 same pattern, indicating that this result is not an artefact of expression averaging across homology groups
252 (Fig. S18).

253 Visualizing the distribution of genes by ancestral and descendant values allows us to identify shifts in bias
254 which represent the largest swings in expression values (Fig. 4F, points a-d). Highlighting the top two such
255 shifts in both directions, we identify four example genes which acquired or lost ovary-specific expression in
256 the phylogeny of Hawaiian Drosophilidae. In the case of *FMRFaR* and *GABA*, a few Hawaiian species have
257 gained ovary-biased expression of these genes, while most species and the ancestral state indicate non-ovary
258 bias (Fig. 4Fa-b). In the case of *vilya* and the unnamed gene *CG9109*, each shows a pattern where one
259 species has lost ovary bias from a biased ancestral state (Fig. 4Fc-d).

260 Repeating the same analysis using the head in place of the ovary revealed a set of evolutionary gains and
261 losses in head-specific expression (Fig. S19). Identifying the top four changes in head expression shows gains
262 and losses of head expression in the genes *hiro*, *stil*, *Jhe*, and, consistent with the ovary, *vilya*. In the case of
263 the latter, these results may be driven by substantial changes in expression of *vilya* in carcass tissues across
264 species, resulting in major differences in both ovary and head-biased expression.

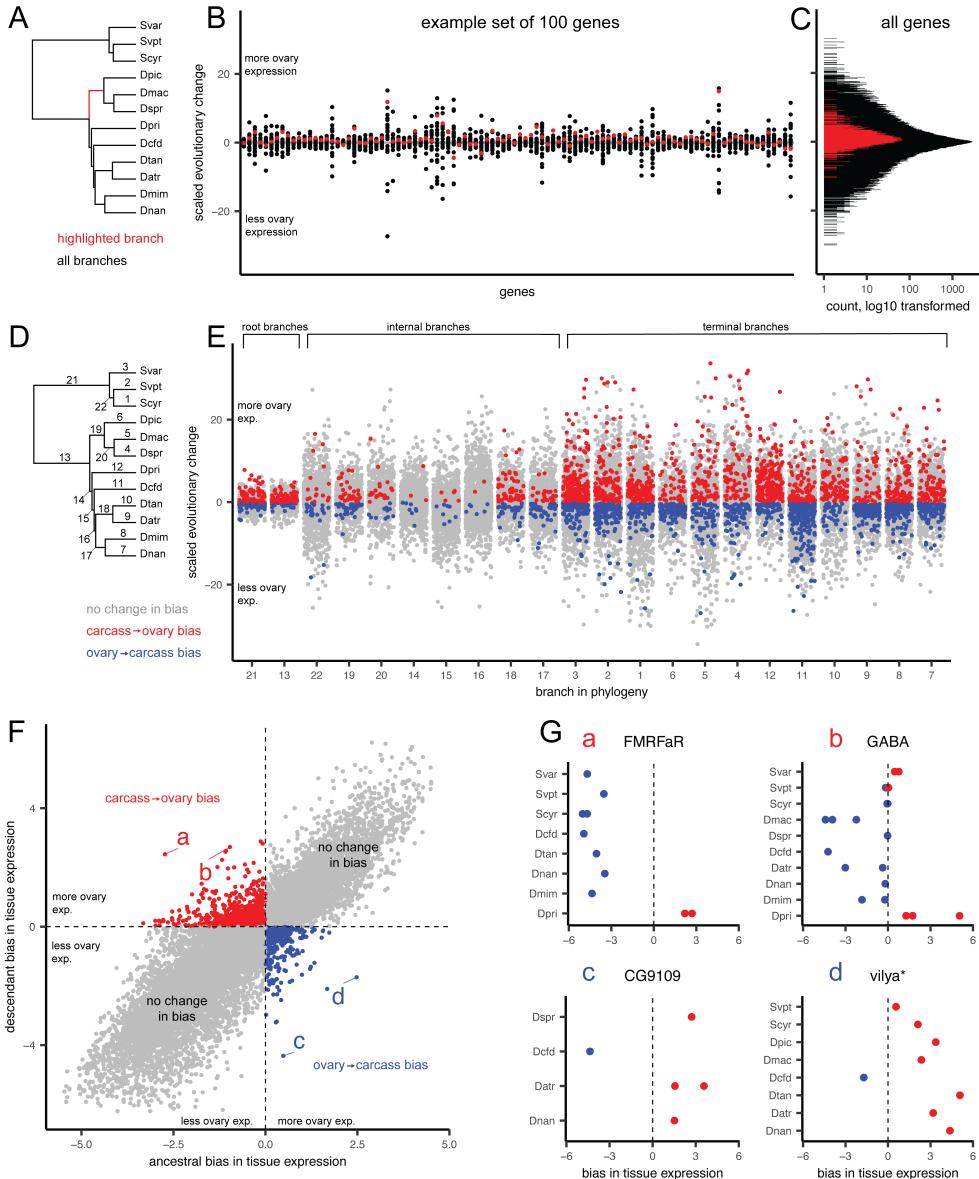


Figure 4: Identifying genes that have gained and lost ovary-biased expression across the phylogeny. A, The phylogeny of the twelve species studied here, highlighting one example branch of the 22 for which we inferred the scaled evolutionary change in expression bias. B, The distribution of changes, grouped by gene, for 100 randomly selected genes, defined by homology group. Each point represents one of the 22 branches from A, with the red point corresponding to the highlighted branch from that panel. C, The distribution, \log_{10} transformed, of scaled genes across all branches and all genes. Changes on the highlighted branch in red. D, The phylogeny with all 22 branches numbered. E, The distribution of changes, grouped by branch, with random jitter on the x-axis within each group. Points colored according to the qualitative change in bias, either from more expression in ovary than carcass to less (blue), the reverse (red), or no change in overall bias (gray). F, The distribution of ancestral and descendant values, showing the two quadrants that represent qualitative changes in bias. Points that represent large swings in expression within those quadrants are labeled a-d. G, The four genes with large swings from F, showing the expression bias for each transcript colored according to more expression in the ovary (red) or carcass (blue). Panels annotated with the gene symbol from the *D. melanogaster* sequences in the same homology group, with the exception of *vilya**, which was annotated using a direct BLAST search since no *D. melanogaster* sequence was present in that group.

265 3.4 Genes with a strong correlation of expression evolution

266 We tested the estimated evolutionary changes in expression bias for evidence of correlated expression evolution
267 between genes. For every gene represented across all species, we performed a pairwise comparison of
268 changes in expression bias, using as data points the scaled change in ovary bias on the 22 branches in the
269 phylogenetic tree. This resulted in 1,306,449 pairwise measures of evolutionary correlation between genes.
270 Because the number of gene pairs being compared is much larger than the number of values used to estimate
271 correlation, this method has the potential to produce many spurious correlations⁷. To test the degree
272 to which the correlations observed here reflect known biological interactions between genes, we compared
273 these measures to reported protein and genetic interactions between genes, using the database of published
274 genetic experiments in *D. melanogaster*, available at <http://flybase.org>. We found that the mean correlation
275 coefficient for genes that are known to physically interact as proteins was higher than for genes with
276 no or unknown interaction (maximum p-value=<0.001 over 100 replicates, Fig. 5A). This indicates that
277 even with a relatively small number of observations, there is sufficient information in the matrix to detect
278 biological signal between gene pairs. These results were calculated based on the correlation in expression
279 bias between the ovary and carcass. However, following the same procedure using correlations in changes in
280 head-biased expression showed no significant difference between the two groups (max. p-value=0.256, Fig.
281 S20), suggesting the strength of this signal may be dependent on the tissues being compared.

282 We also found that genes known to interact genetically have a significantly higher mean correlation than
283 genes with no or unknown genetic interactions (unknown vs. enhancement max. p-value <0.001, unknown
284 vs. suppression max. p-value=<0.001, Fig. 5B). Comparing genes with known genetic enhancement and
285 suppression interactions to each other showed no significant difference (p-value=0.497). However, for genetic
286 interactions, the range of correlation coefficients was higher in the group of no or unknown interactions (Fig.
287 5B). This indicates that, while the average correlation of expression evolution might be higher for interaction
288 partners, stronger positive and negative correlations exist between pairs of genes which do not interact, or
289 for which interactions have not yet been tested.

290 As evidence of this, we tested whether the network inferred based on strong correlation of expression evolution
291 was consistent with known interaction partners from *D. melanogaster*. We selected as an example the gene
292 yolk-protein gene family, which are known to be expressed in the reproductive system, among other tissues⁴⁹
293 (Fig. 5C). We found eight distinct homologous gene groups, comprising 14 unique *D. melanogaster* parent
294 genes, that had a strong evolutionary correlation with yolk-protein genes (absolute coefficient greater than
295 0.825, Fig. 5D). None of these correlated genes correspond to those listed on FlyBase⁵⁰ as having known
296 interactions with yolk-protein genes in *D. melanogaster* (Fig. 5E). We consider these strong evolutionary
297 correlations to be a set of new predictions about evolutionary and genetic relationships between genes which
298 can be tested in wild and laboratory model species of *Drosophila*. The dataset of pairwise correlation
299 coefficients can be visualized and interrogated at the accompanying data visualization for this manuscript
300 (https://github.com/shchurch/hawaiian_fly_dataviz_2021).

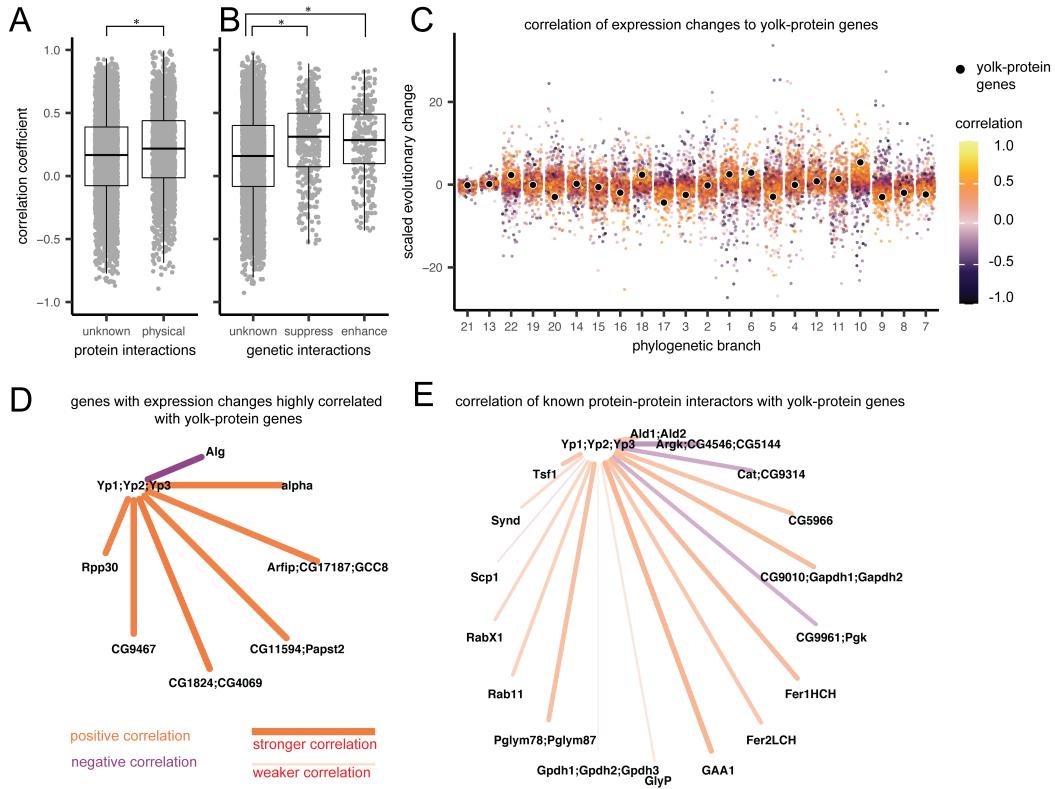


Figure 5: Estimating pairwise correlation coefficients across genes reveals new networks of correlated expression evolution. A-B Comparison of the distribution of Pearson's correlation coefficients based on ovary-biased expression evolution between genes. Box plots indicate mean, upper and lower quartiles, and 1.5x interquartile ranges. Asterisks indicate a significant t-test comparison. A, Genes with no or unknown protein-protein interactions compared to those with reported interactions in FlyBase⁵⁰ (maximum p-value=<0.001 over 100 replicates). B, Correlation comparison between genes with no or unknown genetic interactions and those reported to have enhancement or suppression interactions in FlyBase (unknown vs. enhancement max. p-value=<0.001; unknown vs. suppression max. p-value =<0.001; enhancement vs. suppression p-value=0.497). C, Each point represents a scaled change in expression bias, colored by Pearson's correlation coefficients relative to one example gene-family, the yolk-protein genes (black points), arranged by phylogenetic branch (numbers shown in Fig. 4D). Yellow=strong positive correlation, purple=strong negative correlation. D, The network of strong correlation partners (absolute correlation > 0.825) with the yolk-protein genes, colored by the direction of correlation. Stronger correlations are shown by brighter colors, and thicker, shorter lines. Nodes are annotated with the gene symbols from the *D. melanogaster* sequences from that homology group. E, The correlation between known protein-protein interaction partners⁵⁰ with the yolk-protein genes.

301 4 Discussion

302 The results of this study show the importance of placing any comparison of gene expression across species
 303 in an evolutionary context. When making comparisons that involve model organisms for the study of
 304 development and disease, this means identifying the crossover point at which variation between species
 305 begins to swamp variation across the tissues or treatments in question. In such comparisons, the possibility
 306 that any individual gene may show a divergent pattern of expression from the model organism increases
 307 substantially. This study provides evidence that confirms we should expect variation in gene expression to

308 increase with the phylogenetic distance separating the species being compared. In addition, our results using
309 ovary and head expression data show that our expectation should also depend on the identity of the tissues
310 being compared. Our dataset demonstrates that for some tissues, like the fly head, this crossover point may
311 be met even when comparing between two relatively closely related species.

312 Despite substantial variation across species, we identified core suites of ovary- and head-expressed genes
313 that have maintained conservation of expression patterns over millions of years of evolution. The core
314 ovary genes include some of the most well-studied genes in relation to *D. melanogaster* oogenesis, such as
315 *nanos* and *oskar*, as well as many genes that have yet to be studied in depth (e.g. unnamed genes such
316 as *CG3430*). We provide the full list of core ovary and head genes as a reference against which future
317 genetic studies may be informed and compared (Tables S6-S7). Furthermore, the existence of these suites of
318 genes suggests that equivalent groups are likely to exist within the many gene expression atlases currently
319 being published^{51,52}. New technologies such as single-cell RNA sequencing that use global signatures of gene
320 expression to identify cells are ripe for interspecific comparisons that may reveal evolutionarily conserved
321 gene modules⁵³. Developing robust comparative methods for comparing these atlases across species has
322 the potential to reveal ancestral expression patterns in cells and organs, as well as pinpoint important
323 evolutionary shifts in expression regulation.

324 Our results indicate that genes known to interact, both physically as the proteins they encode and through
325 genetic enhancement and suppression, likely experience more correlated changes in expression than would
326 be expected for genes chosen at random. However, we also find the difference in mean correlation between
327 these groups to be relatively small, and dependent on the context of the tissue in question. One possible
328 explanation for this finding is that interactions between genes with strong correlations of expression evolution
329 have yet to be described. We provide an interactive tool to explore highly correlated genes that can inform
330 future genetic studies in *D. melanogaster* and other related species (https://github.com/shchurch/hawaiian_fly_dataviz_2021). Another possibility we consider likely is that interactions between genes represent only
331 one factor among many that dictate the probability of correlated changes in expression. We hypothesize
332 that other features, such as shared regulatory or chromatin architecture, will also influence evolutionary
333 correlation of expression.

334 As more studies undertake phylogenetic comparisons of functional genomic data, new factors that influence
335 the evolutionary associations between genes are likely to be revealed⁷. The strength of these phylogenetic
336 comparisons will depend in part on comparing across a sufficient number of taxa such that there are multiple
337 branches on which to calculate and compare evolutionary changes. However, even as functional genomic data
338 become more accessible for more species, the number of features being compared (e.g. thousands of genes) will
339 likely continue to outnumber the number of evolutionary observations (e.g. changes along branches)⁷. One
340 encouraging result from this study is that, using our matrix of gene expression changes along 22 branches, we
341 find sufficient information to detect the biological signal associated with physical and genetic interactions.
342 While this is true, we assume that some fraction of the correlations that we report here represent false
343 positives, and that the strength of correlation of these genes would decrease with the addition of more taxa
344 to the comparison. For this reason we present the correlation matrix as a set of hypotheses to be tested in
345 future studies using additional lines of evidence.

346 One outstanding challenge in expression evolution is the quality of the references available against which
347 RNA reads can be mapped⁴⁸. In this study we account for the statistical noise in our data by averaging
348 expression values over groups of homologous genes, as identified by sequence similarity to high quality refer-
349 ence genomes. This approach has the advantage of accounting for problems associated with fragmentation of
350 genes in transcriptome assembly. However, it comes at the cost of averaging over possible biological variation
351 in expression between genes from the same gene family. The strong concordance of our results with published
352 records from *D. melanogaster* suggests that the approach we have used here is robust for our dataset. How-
353 ever, as the quality and accessibility of genomes from diverse species continue to increase, future studies will
354 likely be able to compare directly between orthologous genes without needing to account for fragmentation.
355 For those future studies, a phylogenetic comparative approach like the one used here and elsewhere⁸ can
356 serve as an analytical framework to move expression comparisons beyond pairwise comparisons.

357 One goal of evolutionary developmental biology is to identify changes in developmental mechanisms that
358 underlie phenotypic differences¹². Many studies approach this by identifying phenotypic variation between

360 species and then searching for differences in gene content or expression using one or several emerging model
361 organisms in the lab¹². To narrow down the field of search, this approach often requires a priori
362 knowledge of candidate genes, gained from developmental research in related models or other methods of
363 filtering the genome. Furthermore, because these approaches usually lack global measurements of gene
364 expression variation across species, identifying an expression difference does not always constitute a smoking
365 gun⁶. For example, observing a difference in candidate gene expression between taxa would not be unexpected
366 if we frequently observe differences of that magnitude between genes chosen at random. An alternative
367 approach, as demonstrated here, is to characterize all the evolutionary changes in expression across the
368 transcriptome, and then identify the changes that are significantly associated with traits of interest⁹. As
369 expression data become available from an ever wider array of species, this “evolutionary screen” approach
370 becomes increasingly possible. One advantage of this approach is that it may reveal associations that would
371 otherwise escape detection when comparisons are centered on model organisms; for example, when genes,
372 traits, or processes happen to not be present in our laboratory model species¹⁰. By leveraging phylogenetic
373 comparative methods on high-dimensional functional genomic data, the objective of connecting genomic
374 variation to developmental mechanisms and phenotypic differences will be accelerated.

375 5 Methods

376 5.1 Field collection

377 Specimens used for transcriptome sampling were caught on the Hawaiian islands between May of 2016 and
378 May of 2017. Specimens were caught using a combination of net sweeping and fermented banana-mushroom
379 baits in various field sites on the Hawaiian islands of Kaua’i and Hawai’i (see Table S1 for locality data).
380 Field collections were performed under permits issued by the following: Hawai’i Department of Land and
381 Natural Resources, Hawai’i Island Forest Reserves, Kaua’i Island Forest Reserves, Koke’e State Park, and
382 Hawai’i Volcanoes National Park. Adult flies were maintained in the field on vials with sugar media and kept
383 at cool temperatures. They were transported alive back to Cambridge, MA where they were maintained on
384 standard *Drosophila* media at 18°C. Samples were processed for RNA extraction between 5 and 31 days after
385 collecting them live in the field (average 10.8 days, see Table S1). One species, *Scaptomyza varia*, was caught
386 in the field before the adult stage by sampling rotting *Clermontia* sp. flowers (the oviposition substrate).
387 For this species, male and female adult flies emerged in the lab, and were kept together until sampled for
388 RNA extraction.

389 5.2 Species identification

390 Species were identified using dichotomous keys⁵⁴⁻⁵⁸, when possible. Many keys for Hawaiian Drosophilidae are written focusing on male specific characters (e.g. sexually dimorphic features or male genitalia)⁵⁶.
391 Therefore, for species where females could not be unambiguously identified by morphology, we verified their
392 identity using DNA barcoding. When males were caught from the same location, we identified males to
393 species using dichotomous keys and matched their barcode sequences to females included in our study. We
394 also matched barcodes from collected females to sequences previously uploaded to NCBI⁵⁹⁻⁶¹.

395 The following dichotomous keys were used to identify species: for *picture-wing* males and females, Magnacca and Price (2012)⁵⁴; for *antopocerus* males, Hardy (1977)⁵⁵; for *Scaptomyza*, Hackman (1959)⁵⁶; for species in the *mimica* subgroup of MM, O’Grady and colleagues (2003)⁵⁷; for other miscellaneous species, Hardy (1965)⁵⁸.

396 For DNA barcoding, DNA was extracted from one or two legs from male specimens using the Qiagen DNeasy blood and tissue extraction kit, or from the DNA of females isolated during RNA extraction (see below). We
397 amplified and sequenced the cytochrome oxidase I (COI), II (COII) and 16S rRNA genes using the primers
398 and protocols described in Sarikaya and colleagues (2019)⁴⁰.

404 For barcode matching, we aligned sequences using MAFFT, version v7.475⁶², and assembled gene trees
405 using RAxML, version 8.2.9⁶³. Definitive matches were considered when sequences for females formed a
406 monophyletic clade with reference males or reference sequences from NCBI; see Table S2.
407 Female *D. primaeva*, *D. macrothrix*, *D. sproati*, and *D. picticornis* could be identified unambiguously using
408 dichotomous keys. Female *D. atroscutellata*, *D. nanella*, *D. mimica*, *D. tanythrix*, *S. cyrtandrae*, *S. varipicta*,
409 and *S. varia* were identified by matching barcodes to reference sequences from NCBI, reference males, or
410 both. For the female *haleakalae* fly used in this study, no male flies were caught in the same location as these
411 individuals, and no other sequences for *haleakalae* males on NCBI were an exact match with this species.
412 Given its similar appearance to *Drosophila dives*, we are referring to it here as *Drosophila* cf *dives*, and we
413 await further molecular and taxonomic studies of this group that will resolve its identity.

414 5.3 Sampling strategy

415 The target number of mature, healthy female flies per species was four, with three intended for dissection
416 and species-specific expression libraries and one intended as a whole-body reference library (Fig. 1). When
417 four such individuals were not available, a reference library was assembled by combining the tissue-specific
418 libraries from one of the other individuals. This was the case for the following species: *D. sproati*, which
419 was dissected and had RNA extracted separately from the head, ovaries, and carcass, with RNA combined
420 prior to library preparation; and *S. varia*, *S. cyrtandrae* and *D. cf dives*, for which RNA was extracted and
421 libraries prepared for separate tissues, and raw reads were combined after sequencing.
422 For the other eight species, sufficient individual females were available such that reads for transcriptome
423 assembly were sequenced from a separate individual. In these cases one entire female fly was dissected and
424 photographed to assess whether vitellogenetic eggs were present in the ovary, and all tissues were combined in
425 the same tube and used for RNA extraction. Library preparation failed for one individual *D. atroscutellata*
426 fly, as well as two tissue-specific libraries: one head sample from *D. mimica*, and one head sample from *D.*
427 *sproati*.

428 5.4 Dissection and RNA sequencing

429 Female flies were anesthetized in 100% ethanol and were dissected in a 1x phosphate-buffered saline solution.
430 The ovary was separated from the abdomen, and the head was separated from the carcass. Photographs
431 of each tissue were taken, and tissues were moved to pre-frozen eppendorf tubes, kept in dry ice, and
432 immediately transported to a -80°C freezer. Dissections were performed as quickly as possible to prevent
433 RNA degradation. Samples were stored at -80°C for between 90 and 336 days before RNA extraction (average
434 281.9 days, see Table S1).

435 RNA was extracted from frozen samples using the standard TRIzol protocol (http://tools.thermofisher.com/content/sfs/manuals/trizol_reagent.pdf). One mL of TRIzol was added to each frozen sample, which
436 were then homogenized using a sterile motorized mortar. The recommended protocol was followed without
437 modifications, using 10 µg of glycogen, and resuspending in 20µL RNase-free water-EDTA-SDS solution.
438 DNA for subsequent barcoding was also extracted using the phenol-chloroform phase saved from the RNA
439 extraction.

441 RNA concentration was checked using a Qubit fluorometer, and integrity was assessed with a Agilent TapeS-
442 tation 4200. RNA libraries were prepared following the PrepX polyA mRNA Isolation kit and the PrepX
443 RNA-Seq for Illumina Library kit, using the 48 sample protocol on an Apollo 324 liquid handling robot in
444 the Harvard University Bauer Core Facilities. Final library concentration and integrity were again assessed
445 using the Qubit and TapeStation protocols.

446 Samples intended for transcriptome assembly were sequenced on an Illumina HiSeq 2500, using the standard
447 version 4 protocol, at 125 base pairs of paired-end reads. Samples intended for tissue-specific expression
448 analyses were sequenced on an Illumina NextSeq 500, using a high output flow cell, at 75 base pairs of
449 single-end reads. A table of total read counts for each library can be found in Tables S3-S4. To account

450 for any possible batch effects across separate rounds of sequencing, each sequencing run was performed with
451 one or several overlapping samples. Principle component analysis of these libraries showed variation between
452 sequencing runs to be negligible relative to variation between tissue and individual (see Results and Fig.
453 S7).

454 **5.5 Transcriptome assembling and expression mapping**

455 Transcriptome assembly and expression mapping was performed using the agalma pipeline, version 2.0.0⁴⁷.
456 For the twelve reference transcriptomes, reads from separate rounds of sequencing were concatenated and
457 inserted into the agalma catalog. Further details of transcriptome assembly and homology assessment are
458 included in our previous manuscript⁴⁴.

459 Each tissue-specific expression library was mapped to the corresponding reference transcriptome using the
460 ‘expression’ pipeline in agalma, which uses the software RSEM to estimate gene and isoform count levels
461 from RNAseq data⁶⁴. The agalma pipeline also includes steps to catalog the species, tissue type, and run
462 information, which were exported as a single JavaScript object notation (JSON) file. This file is available in
463 the GitHub repository in the directory `analysis/data`.

464 **5.6 Phylogenetic analysis**

465 The phylogenetic methods for inferring homology, orthology, and estimating gene and species trees are the
466 same as those described in our previous manuscript⁴⁴. Genetrees were additionally annotated with the
467 software Phyldog⁶⁵.

468 **5.7 Annotating transcripts by sequence similarity**

469 We leveraged the close relationship of these species to species of *Drosophila* with well-annotated genomes to
470 annotate the transcripts considered here. For each transcript in the reference transcriptome, we performed
471 four comparisons of sequence similarity using local BLAST: [1] comparing nucleotide transcript sequences to
472 nucleotide sequences from *D. melanogaster* (blastn), [2] comparing translated nucleotide sequences to protein
473 sequences of *D. melanogaster* (blastx), [3] comparing nucleotide sequences to a database of nucleotide se-
474 quences from *D. melanogaster*, *D. virilis*, and *D. grimshawi* (blastx), and [4] comparing translated nucleotide
475 sequences to a database of protein sequences from the same three species (blastn). For downstream analyses,
476 we prioritized annotations from the second comparison, but we provide all sequence similarity reports in the
477 GitHub repository under the directory `analysis/BLAST`.

478 To annotate homology groups as defined by the homology inference step of agalma, we extracted the name
479 and sequence ID from all *D. melanogaster* sequences in the group.

480 **5.8 Normalization and differential gene expression**

481 Transcript count tables were imported into R using the agalmar package, version 0.0.0.9000. Differential gene
482 expression analysis was performed using the package DESeq2, version 1.34.0. For these analyses we used only
483 one sequencing run per library, thereby excluding duplicate sequencing runs. Analyses of differential gene
484 expression were calculated using the default approaches in DESeq2 for estimating size factors, dispersions,
485 and calculating log₂ fold-change and p-values (Fig. S2A). Both individual and tissue were considered in the
486 design formula. Transcripts were considered differentially expressed at a significance threshold of 0.01.

487 We identified a cohort of core ovary-specific genes by first identifying a parent gene for each transcript using
488 a sequence similarity search against *D. melanogaster* (Fig. S2A). We then identified parent genes that had
489 at least one transcript significantly differentially upregulated in the ovary of more than ten of the twelve
490 species. Because multiple transcripts may match to a single parent-gene, core ovary-specific parent genes

491 may include transcripts that are also not differentially upregulated in the ovary, as long as at least one
492 transcript is for more than ten out of twelve species. This may be the case when transcripts are artificially
493 fragmented during reference transcriptome assembly, or when sequence-similar transcripts have biologically
494 distinct expression levels.

495 5.9 Comparison of expression to *D. melanogaster*

496 We compared our differential gene expression results to a reference database of tissue expression from *D.*
497 *melanogaster*, known as the FlyAtlas2²⁹. We downloaded this reference in July of 2021, from http://motif.gla.ac.uk/downloads/FlyAtlas2_21.04.18.sql. This dataset provides data on transcript abundance and tissue
498 enrichment, including for female ovaries. Tissue enrichment is calculated using the same methods as in the
499 FlyAtlas2 web browser, defined as the fragments per kilobase of transcript per million mapped reads (FPKM)
500 for a given tissue divided by that value for the reference tissue (here, female whole body), with a pseudocount
501 of two counts added to empty values to avoid division by zero. We considered a FlyAtlas gene to be enriched
502 in the ovary, comparable to our data, if the ovary was the maximum enrichment value across all tissues
503 excluding the head, brain, and eye tissues, as these were separated in our RNASeq procedure (Fig. S2A).
504 We considered a FlyAtlas gene to be head enriched if either the head, brain, or eye were the maximum
505 enrichment value, excluding the ovary.

507 5.10 Transforming data into comparable measurements of expression across 508 species

509 Transcript counts are reported in transcripts per million (TPM), but this measurement is known to not
510 be directly comparable across species due to differences in reference transcriptome size^{7,8}. Therefore, we
511 normalized TPM by species using the procedure described by Munro and colleagues (2021)⁸, where TPM
512 values are multiplied by the number of genes in the reference, and this value is divided by 10⁴ (Fig. S2B).
513 TPM10k values were natural-log transformed.

514 An additional challenge when working with reference transcriptomes is the presence of fragmented transcripts
515 created during the assembly process⁴⁸. This fragmentation can result in noise in estimating the amount of
516 transcript as reads are differentially mapped to these fragments. To reduce the impact of this noise on our
517 analysis, we undertook a novel approach where transcripts were grouped according to inferred homology
518 as estimated by the agalma pipeline using an all-by-all BLAST approach (Fig. S2B). For each sequenced
519 library, we then found the average count value across all transcripts from the same homology group (see
520 Table S5 for statistics on homology group composition). For each species-tissue pair, we then averaged this
521 value across all biological replicates, here replicate individuals.

522 5.11 Linear modeling

523 We performed linear modeling to calculate the relative contribution of tissue- and species-level differences
524 to variation in gene expression (Fig. S2B), following the approach of Breschi and colleagues (2016)²¹. These
525 analyses were performed separately on datasets of ovary vs. carcass and head vs. carcass expression. Using
526 the ANOVA script provided at <https://github.com/abreschi/Rscripts/blob/master/anova.R>, we built
527 a linear model for each gene that accounts for the contribution of the organ, species, and any residual error.
528 We then calculated the relative proportion of each factor divided by the total sum of squares for all factors.
529 We identified groups of highly variable genes, using the same metrics defined by Breschi and colleagues
530 (2016)²¹, as any gene for which either tissues or species explains at least 75% of the variance. Species
531 variable genes (SVGs) were defined as highly variable genes whose relative variation was two-fold greater
532 across species than tissues (vice-versa for tissue variable genes, TVGs).

533 We performed these linear model analyses over four nested clades: a clade containing two *picture-wing* species
534 (*D. sproati* and *D. macrothrix*); a clade containing the four *picture-wing*-*Nudidrosophila*-*Ateledrosophila*

535 species in this study; a clade containing the nine Hawaiian *Drosophila* species in this study; and a clade of
536 all 12 Hawaiian *Drosophila* and *Scaptomyza* species in this study. We repeated these analyses excluding the
537 species *S. varia*, which showed the lowest similarity in expression to the other eleven species. To compare
538 our analysis to the more typical approach undertaken, we also performed these analyses on all pairwise
539 combinations of these twelve species.

540 5.12 Reconstructing evolutionary history of differential expression

541 We calculated tissue bias as the ratio of counts in TPM10k for each tissue (ovary and head) to the reference
542 tissue⁷, here the carcass (Fig. S3A). We subsequently performed the same transformation steps described
543 above, averaging over ratios from the same homology group and across biological replicates, to calculate
544 average expression bias per homology group per library. To avoid division by zero, we added a pseudocount
545 of 0.01 to each TPM10k value. Ratio values were natural-log transformed so that positive values indicate
546 enrichment in the tissue of interest relative to the reference tissue, negative values indicate the opposite, and
547 values of zero indicate equivalent expression.

548 We reconstructed the evolutionary history of tissue bias for each homology group using the species tree
549 published in Church and Extavour, 2021⁴⁴, based on the same reference transcriptome data (Fig. S2C).
550 First, we calibrated the tree estimated using IQtree (Fig 1A of that publication) to be ultrametric using the
551 R function `chronos` in the package `ape`, version 5.6.2 (using a correlated model and a lambda value of 1).
552 We then subset this tree to only include tips for which expression data was available, and annotated this
553 tree to be able to identify specific branches and nodes in ancestral state reconstruction analyses.

554 Ancestral expression bias values were estimated with the R package Rphylopars, version 0.3.8, using the fast
555 ancestral state reconstruction algorithm based on Ho and Ané, 2014⁶⁶ (Fig. S3A). Tips for which expression
556 data were not available were dropped from each reconstruction, and ancestral state reconstruction was only
557 performed when more than three tips had data. Following ancestral state reconstruction, we calculated the
558 scaled change as the difference between the value at the ancestral and descendant nodes, divided by the
559 length of the branch. Scaled changes were compared between homology groups by identifying equivalent
560 branches as those that share the same parent and child node, following the procedure described in Munro
561 and colleagues (2021)⁸. We identified qualitative changes in expression bias as changes that resulted in a
562 ratio changing from negative to positive values or vice versa.

563 5.13 Estimating correlated evolution of expression across genes

564 For each homology group that had representation across all twelve species, we calculated pairwise Pearson's
565 correlation coefficients by comparing scaled changes in expression bias across equivalent branches (Fig.
566 S3B). For the twelve-species phylogeny, this meant each correlation coefficient was calculated using 22
567 individual data points (branches). This resulted in a correlation matrix of 1,306,449 pairwise comparisons of
568 evolutionary correlation.

569 We compared this correlation network to data on protein interactions and genetic interactions downloaded
570 from <http://flybase.org> in July, 2021. These data include pairwise observations of genetic enhancement
571 and suppression interactions between parent genes in *D. melanogaster*. These interactions were matched to
572 pairwise correlation coefficients by identifying the corresponding homology group for each *D. melanogaster*
573 parent gene ID (more than one parent gene may fall into the same homology group).

574 We tested whether correlation coefficients for known physical and genetic interaction partners were higher
575 than in genes with unknown interactions using two-sample t-tests. The sample size for physical interaction
576 partners was 1,953, for genetic enhancement was 280, and for genetic suppression was 497. In each test we
577 compared the coefficients for either enhancement or suppression interactions to a random sample of 5000
578 coefficients for which interactions are unknown. We repeated these t-tests 100 times using different random
579 samples, and report the maximum p-value observed. We also compared the distribution of enhancement and
580 suppression interaction coefficients to each other using a single t-test.

581 Strong correlations for the visualization of co-evolutionary networks were selected using a threshold correlation
582 coefficient of 0.825.

583 6 Data Availability

584 All data are available at GitHub, under the repository `shchurch/hawaiian_drosophilidae_expression_2021`,
585 commit `67d8e6f`. The correlation matrix can be interactively visualized and queried at the accompanying
586 data visualization for this paper (https://github.com/shchurch/hawaiian_fly_dataviz_2021). Raw RNA
587 sequencing data are available at the Sequence Read Archive of the National Center for Biotechnology Information
588 (NCBI), under BioProject PRJNA731506. Assembled transcriptomes and DNA barcode sequences are
589 available at GitHub, under the repository http://github.com/shchurch/hawaiian_drosophilidae_phylogeny_2021,
590 commit `b12cbb10`.

591 7 Code Availability

592 All code and results for this manuscript are available at GitHub, under the repository `shchurch/hawaiian_drosophilidae_ex`
593 commit `67d8e6f`. The code to perform all agalma commands was performed in clean anaconda environment,
594 installed following the instructions at <https://bitbucket.org/caseywdunn/agalma>. All R commands were
595 performed with a fresh install of R, and the session information including all package versions is available
596 in the GitHub repository under the file `r_session_info.txt`. The code to generate all plots as well as the
597 text of this manuscript is available in several R scripts and Rmarkdown files at the same location.

598 8 Acknowledgments

599 This work was partially supported by National Science Foundation Graduate Research Fellowship Program
600 DGE1745303 to SHC, National Institutes of Health award R01 HD073499-01 (NICHD) to CGE, and funds
601 from Harvard University to support SHC and CGE. We thank Didem Sarikaya, Karl Magnacca, and Steve
602 Montgomery for their field assistance and expertise in Hawaiian fly identification and husbandry. We thank
603 Kenneth Kaneshiro for the use of his lab space in preparing wild caught specimens. We thank Bruno de
604 Medeiros, Seth Donoughe, and members of the Dunn and Extavour labs for discussion of ideas.

605 9 Competing Interest

606 The authors declare no competing interests.

607 References

- 608 1. Wang, Z., Gerstein, M. & Snyder, M. RNA-seq: A revolutionary tool for transcriptomics. *Nature Reviews Genetics* **10**, 57–63 (2009).
- 610 2. Ozsolak, F. & Milos, P. M. RNA sequencing: Advances, challenges and opportunities. *Nature Reviews Genetics* **12**, 87–98 (2011).
- 612 3. Van Dijk, E. L., Auger, H., Jaszczyszyn, Y. & Thermes, C. Ten years of next-generation sequencing technology. *Trends in Genetics* **30**, 418–426 (2014).
- 614 4. McDermaid, A., Monier, B., Zhao, J., Liu, B. & Ma, Q. Interpretation of differential gene expression results of RNA-seq data: Review and integration. *Briefings in Bioinformatics* **20**, 2044–2054 (2019).
- 616 5. Dunn, C. W., Zapata, F., Munro, C., Siebert, S. & Hejnol, A. Pairwise comparisons across species are problematic when analyzing functional genomic data. *Proceedings of the National Academy of Sciences* **115**, E409–E417 (2018).
- 619 6. Church, S. H. & Extavour, C. G. Null hypotheses for developmental evolution. *Development* **147**, dev178004 (2020).
- 621 7. Dunn, C. W., Luo, X. & Wu, Z. Phylogenetic analysis of gene expression. *Integrative and Comparative Biology* **53**, 847–856 (2013).
- 623 8. Munro, C., Zapata, F., Howison, M., Siebert, S. & Dunn, C. W. Evolution of gene expression across species and specialized zooids in Siphonophora. *bioRxiv* (2021).
- 625 9. Smith, S. D., Pennell, M. W., Dunn, C. W. & Edwards, S. V. Phylogenetics is the new genetics (for most of biodiversity). *Trends in Ecology & Evolution* **35**, 415–425 (2020).
- 627 10. Dunn, C. W., Leys, S. P. & Haddock, S. H. The hidden biology of sponges and ctenophores. *Trends in Ecology & Evolution* **30**, 282–291 (2015).
- 629 11. Arias, A. M. *Drosophila melanogaster* and the development of biology in the 20th century. *Drosophila* **420**, 1–25 (2008).
- 631 12. Jenner, R. A. & Wills, M. A. The choice of model organisms in evo–devo. *Nature Reviews Genetics* **8**, 311–314 (2007).
- 633 13. Yanai, I., Graur, D. & Ophir, R. Incongruent expression profiles between human and mouse orthologous genes suggest widespread neutral evolution of transcription control. *Omics: a Journal of Integrative Biology* **8**, 15–24 (2004).
- 636 14. Lin, S. *et al.* Comparison of the transcriptional landscapes between human and mouse tissues. *Proceedings of the National Academy of Sciences* **111**, 17224–17229 (2014).
- 638 15. Gilad, Y. & Mizrahi-Man, O. A reanalysis of mouse ENCODE comparative gene expression data. *F1000Research* **4**, 1–30 (2015).
- 640 16. Liao, B.-Y. & Zhang, J. Evolutionary conservation of expression profiles between human and mouse orthologous genes. *Molecular Biology and Evolution* **23**, 530–540 (2006).
- 642 17. Cardoso-Moreira, M. *et al.* Gene expression across mammalian organ development. *Nature* **571**, 505–509 (2019).
- 644 18. Fukushima, K. & Pollock, D. D. Amalgamated cross-species transcriptomes reveal organ-specific propensity in gene expression evolution. *Nature Communications* **11**, 1–14 (2020).
- 646 19. Merkin, J., Russell, C., Chen, P. & Burge, C. B. Evolutionary dynamics of gene and isoform regulation in mammalian tissues. *Science* **338**, 1593–1599 (2012).
- 648 20. Brawand, D. *et al.* The evolution of gene expression levels in mammalian organs. *Nature* **478**, 343–348 (2011).

650 21. Breschi, A. *et al.* Gene-specific patterns of expression variation across organs and species. *Genome Biology* **17**, 1–13 (2016).

652 22. Robinson, S. W., Herzyk, P., Dow, J. A. & Leader, D. P. FlyAtlas: Database of gene expression in the
653 tissues of *Drosophila melanogaster*. *Nucleic Acids Research* **41**, D744–D750 (2013).

654 23. Graveley, B. R. *et al.* The developmental transcriptome of *Drosophila melanogaster*. *Nature* **471**,
655 473–479 (2011).

656 24. St. Pierre, S. E., Ponting, L., Stefancsik, R., McQuilton, P. & Consortium, F. FlyBase 102—advanced
657 approaches to interrogating flybase. *Nucleic Acids Research* **42**, D780–D788 (2014).

658 25. Chintapalli, V. R., Wang, J. & Dow, J. A. Using flyatlas to identify better *Drosophila melanogaster*
659 models of human disease. *Nature Genetics* **39**, 715–720 (2007).

660 26. Papatheodorou, I. *et al.* Expression atlas: Gene and protein expression across multiple studies and
661 organisms. *Nucleic Acids Research* **46**, D246–D251 (2018).

662 27. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic
663 data across different conditions, technologies, and species. *Nature Biotechnology* **36**, 411–420 (2018).

664 28. Parisi, M. *et al.* A survey of ovary-, testis-, and soma-biased gene expression in *Drosophila melanogaster*
665 adults. *Genome Biology* **5**, 1–18 (2004).

666 29. Leader, D. P., Krause, S. A., Pandit, A., Davies, S. A. & Dow, J. A. T. FlyAtlas 2: A new version of the
667 *Drosophila melanogaster* expression atlas with RNA-seq, miRNA-seq and sex-specific data. *Nucleic Acids
668 Research* **46**, D809–D815 (2018).

669 30. Kumar, T., Blondel, L. & Extavour, C. G. Topology-driven protein-protein interaction network analysis
670 detects genetic sub-networks regulating reproductive capacity. *Elife* **9**, e54082 (2020).

671 31. Kirilly, D. & Xie, T. The *Drosophila* ovary: An active stem cell community. *Cell Research* **17**, 15–25
672 (2007).

673 32. Jagut, M. *et al.* A mosaic genetic screen for genes involved in the early steps of *Drosophila* oogenesis.
674 *G3: Genes, Genomes, Genetics* **3**, 409–425 (2013).

675 33. Barnett, T., Pachl, C., Gergen, J. P. & Wensink, P. C. The isolation and characterization of *Drosophila*
676 yolk protein genes. *Cell* **21**, 729–738 (1980).

677 34. Kugler, J.-M. & Lasko, P. Localization, anchoring and translational control of oskar, gurken, bicoid and
678 nanos mRNA during *Drosophila* oogenesis. *Fly* **3**, 15–28 (2009).

679 35. Stephenson, E. C., Chao, Y.-C. & Fackenthal, J. D. Molecular analysis of the *swallow* gene of *Drosophila*
680 *melanogaster*. *Genes & development* **2**, 1655–1665 (1988).

681 36. O’Grady, P. & DeSalle, R. Hawaiian *Drosophila* as an evolutionary model clade: Days of future past.
682 *Bioessays* **40**, 1700246 (2018).

683 37. Obbard, D. J. *et al.* Estimating divergence dates and substitution rates in the *Drosophila* phylogeny.
684 *Molecular Biology and Evolution* **29**, 3459–3473 (2012).

685 38. Kambysellis, M. P. *et al.* Pattern of ecological shifts in the diversification of Hawaiian *Drosophila* inferred
686 from a molecular phylogeny. *Current Biology* **5**, 1129–1139 (1995).

687 39. Montague, J. R., Mangan, R. L. & Starmer, W. T. Reproductive allocation in the Hawaiian Drosophilidae:
688 Egg size and number. *The American Naturalist* **118**, 865–871 (1981).

689 40. Sarikaya, D. P. *et al.* Reproductive capacity evolves in response to ecology through common changes in
690 cell number in Hawaiian *Drosophila*. *Current Biology* **29**, 1877–1884 (2019).

691 41. Church, S. H., Donoughe, S., Medeiros, B. A. de & Extavour, C. G. A dataset of egg size and shape
692 from more than 6,700 insect species. *Scientific Data* **6**, 1–11 (2019).

693 42. Church, S. H., Medeiros, B. A. de, Donoughe, S., Márquez Reyes, N. L. & Extavour, C. G. Repeated
694 loss of variation in insect ovary morphology highlights the role of development in life-history evolution.
695 *Proceedings of the Royal Society B* **288**, 20210150 (2021).

696 43. Edwards, K. A., Doescher, L. T., Kaneshiro, K. Y. & Yamamoto, D. A database of wing diversity in the
697 Hawaiian *Drosophila*. *PLoS One* **2**, e487 (2007).

698 44. Church, S. H. & Extavour, C. G. Phylotranscriptomics reveals discordance in the phylogeny of Hawaiian
699 *Drosophila* and *Scaptomyza* (Diptera: Drosophilidae). *bioRxiv* (2021).

700 45. McLaughlin, J. M. & Bratu, D. P. *Drosophila melanogaster* oogenesis: An Overview. *Methods in
701 molecular biology (Clifton, NJ)* **1328**, 1–20 (2015).

702 46. Hoshino, M., Suzuki, E., Nabeshima, Y.-i. & Hama, C. Hikaru genki protein is secreted into synaptic
703 clefts from an early stage of synapse formation in *Drosophila*. *Development* **122**, 589–597 (1996).

704 47. Dunn, C. W., Howison, M. & Zapata, F. Agalma: An automated phylogenomics workflow. *BMC
705 Bioinformatics* **14**, 1–9 (2013).

706 48. Freedman, A. H., Clamp, M. & Sackton, T. B. Error, noise and bias in de novo transcriptome assemblies.
707 *Molecular Ecology Resources* **21**, 18–29 (2021).

708 49. Garabedian, M. J., Shepherd, B. M. & Wensink, P. C. A tissue-specific transcription enhancer from the
709 *Drosophila* yolk protein 1 gene. *Cell* **45**, 859–867 (1986).

710 50. Larkin, A. *et al.* FlyBase: Updates to the *Drosophila melanogaster* knowledge base. *Nucleic Acids
711 Research* **49**, D899–D907 (2021).

712 51. Consortium, T. M. & others. Single-cell transcriptomics of 20 mouse organs creates a *Tabula Muris*.
713 *Nature* **562**, 367–372 (2018).

714 52. Quake, S. R., Consortium, T. S. & others. The *Tabula Sapiens*: A single cell transcriptomic atlas of
715 multiple organs from individual human donors. *BioRxiv* (2021).

716 53. Tanay, A. & Sebé-Pedrós, A. Evolutionary cell type mapping with single-cell genomics. *Trends in
717 Genetics* **37**, (2021).

718 54. Magnacca, K. N. & Price, D. K. New species of Hawaiian picture wing *Drosophila* (Diptera: Drosophilidae), with a key to species. *Zootaxa* **3188**, 1–30 (2012).

720 55. Hardy, D. Review of the Hawaiian *Drosophila* (antopocerus) Hardy [insects]. *Proceedings Entomological
721 Society of Washington* **79**, 82–95 (1977).

722 56. Hackman, W. On the genus *Scaptomyza* Hardy (Dipt., Drosophilidae) with descriptions of new species
723 from various parts of the world. *Acta Zoologica Fennica* **97**, 1–73 (1959).

724 57. O’Grady, P., Kam, M., Val, F. do & Perreira, W. Revision of the *Drosophila* mimica subgroup, with
725 descriptions of ten new species. *Annals of the Entomological Society of America* **96**, 12–38 (2003).

726 58. Hardy, D. *Diptera: Cyclorrhapha II, Series Schizophora Section Acalypterae I. Family Drosophilidae*.
727 vol. 12 (University of Hawai’i Press, 1965).

728 59. O’Grady, P. M. *et al.* Phylogenetic and ecological relationships of the Hawaiian *Drosophila* inferred by
729 mitochondrial DNA analysis. *Molecular Phylogenetics and Evolution* **58**, 244–256 (2011).

730 60. Lapoint, R. T., Magnacca, K. N. & O’Grady, P. M. Phylogenetics of the Antopocerus-Modified Tarsus
731 clade of Hawaiian *Drosophila*: Diversification across the Hawaiian islands. *PLoS One* **9**, e113227 (2014).

732 61. Katoh, T., Izumitani, H. F., Yamashita, S. & Watada, M. Multiple origins of Hawaiian drosophilids:
733 Phylogeography of *Scaptomyza* hardy (Diptera: Drosophilidae). *Entomological Science* **20**, 33–44 (2017).

734 62. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: Improvements in
735 performance and usability. *Molecular Biology and Evolution* **30**, 772–780 (2013).

736 63. Stamatakis, A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies.
737 *Bioinformatics* **30**, 1312–1313 (2014).

738 64. Li, B. & Dewey, C. N. RSEM: Accurate transcript quantification from RNA-seq data with or without a
739 reference genome. *BMC Bioinformatics* **12**, 1–16 (2011).

740 65. Boussau, B. *et al.* Genome-scale coestimation of species and gene trees. *Genome Research* **23**, 323–330
741 (2013).

742 66. Tung Ho, L. si & Ané, C. A linear-time algorithm for gaussian and non-gaussian trait evolution models.
743 *Systematic Biology* **63**, 397–408 (2014).