

1                   **Expression analysis in a dispersal-fecundity**  
2                   **polyphenism identifies growth regulators and effectors**

3                   David R. Angelini <sup>1</sup>, Joshua L. Steele, Michael C. Yorsz, Devin M. O'Brien \*

4                   Department of Biology, Colby College, 5700 Mayflower Hill, Waterville, ME 04901

5                   \* Present address: Department of Natural Sciences & Mathematics, SUNY Cobleskill, 106  
6                   Suffolk Circle, Cobleskill, NY 12043

7                   <sup>1</sup> Corresponding author. E-mail: [david.r.angelini@gmail.com](mailto:david.r.angelini@gmail.com) Phone: 1-207-859-5734 Fax: 2-  
8                   207-859-57xx

9                   **ORCID**

10          0000-0002-2776-2158      David R. Angelini

11          0000-0000-0000-0000      Joshua L. Steele

12          0000-0002-7026-2327      Michael C. Yorsz

13          0000-0000-0000-0000      Devin M. O'Brien

14                   **Running title**

15                   Expression analysis in a polyphenism

16                   **Keywords**

17                   Polyphenism; trade-off; wing morphs; *Jadera haematoloma*; muscle development; ovaries;  
18                   testes; nutrient limitation; starvation response

19                   **Word count**

20                   6770

21                   **Synopsis**

22                   Polyphenism allows organisms to respond to varying environmental conditions by adopting  
23                   alternative collections of morphological traits. Polyphenic morphs evolve to optimize  
24                   reproductive success using different strategies. In many insects, polyphenism affecting the  
25                   development of flight trades dispersal ability for increased fecundity. The soapberry bug *Jadera*  
26                   *haematoloma* (Hemiptera: Rhopalidae) exhibits wing polyphenism in response to juvenile  
27                   nutritional resources and cohort density. Development of full-length wings and flight-capable  
28                   thoracic muscles occurs more frequently in cohorts raised under low food density conditions, and  
29                   these features are correlated to reduced female fecundity. Soapberry bugs represent an example  
30                   of polyphenic dispersal-fecundity trade-off. Short-wing development is not sex-limited, and male  
31                   morphs differ in fertility. We have previously shown, via a candidate gene approach, that  
32                   manipulation of insulin signaling can alter the threshold for nutritional response and that changes  
33                   in the activity of this pathway underlie, at least in part, differences in the polyphenic thresholds

34 in different host-adapted populations of *J. haematoloma*. We now expand the examination of this  
35 system using transcriptome sequencing across a multidimensional matrix of life stage, tissue,  
36 sex, food density and host population. We also examine the use of wing and thorax shape as  
37 factors modeling gene expression. In addition to insulin signaling, we find that components of  
38 the TOR, Hippo, Toll and estrogen-related receptor pathways are differentially expressed in the  
39 thorax of polyphenic morphs. The transcription factor *Sox14* was one of the few genes  
40 differentially expressed in the gonads of morphs, being up-regulated in ovaries. We identify two  
41 transcription factors as potential mediators of morph-specific male fertility differences. We also  
42 find that bugs respond to nutrient limitation with expression of genes linked to cuticle structure  
43 and spermatogenesis. These findings provide a broad perspective from which to view this  
44 nutrition-dependent polyphenism.

## 45 Introduction

46 The poet Ikkyu once wrote, “Many paths lead from the foot of the mountain, but at the peak we  
47 all gaze at the single bright moon.” While intended to describe the diversity of human  
48 experience, this quote is also a fitting metaphor for polyphenism, the ability of many organisms  
49 to follow different developmental paths to different reproductive strategies. In many diverse  
50 insect groups polyphenism creates morphs that trade-off dispersal and fecundity. Developmental  
51 plasticity shifts nutritional resources towards optimizing an individual for the biomechanical  
52 demands of dispersal by flight or for early or increased activation of gamete production or the  
53 provisioning offspring (Roff 1990). As in Ikkyu’s poem, success can be achieved via many  
54 paths. The evolutionary persistence of multiple strategies and of the developmental mechanisms  
55 that facilitate polyphenism suggest that neither strategy alone would be as successful in the long  
56 term, and it implies a history of environmental fluctuations (Levins 1968). One of the major  
57 goals of comparative biology has been to understand the mechanisms by which organisms  
58 integrate environmental signals to affect development. Polyphenisms provide useful systems to  
59 explore this question, because the degree of environmental influence is often large and  
60 stereotyped. However, these systems can also be used to explore more subtle responses to the  
61 environment.

62 The red-shouldered soapberry bug *Jadera haematoloma* (Herrich-Schäffer 1847) is a hemipteran  
63 common in temperate North America, and it is a promising model for the developmental genetic  
64 study of polyphenism and phenotypic plasticity. Two adult morphs exist in this species (Fig. 1).  
65 One morph is long-winged (Fig. 1A), and these nascent adults have robust flight muscles (Fig.  
66 1D). The second morph has fore-shortened, brachypterous wings (Fig. 1C). These individuals  
67 lack functional flight muscles (Fig. 1E), but females produce eggs at a faster rate than long-  
68 winged females (Fawcett et al. 2018). Within a cohort, the frequency of each morph is dependent  
69 on food availability during juvenile stages, with greater food access inducing more individuals to  
70 develop as the high-fecundity, short-winged morph (Dingle & Winchell 1997; Fawcett et al.  
71 2018). The bugs have diversified on multiple host plants (Carroll & Boyd 1992), and bugs  
72 associated with high-resource plants show heritable, decreased thresholds for development of  
73 short-winged morphs (Fawcett et al. 2018). While flight-fecundity polyphenisms are found in  
74 many insects groups (Harrison 1980), *J. haematoloma* has the uncommon feature that the  
75 flightless morph is not sex-limited. Females and males can develop as either morph and do so at  
76 roughly equal frequencies. In populations where short-winged individuals develop less frequently,  
77 laboratory crosses involving short-winged males produced fewer offspring than those with long-  
78 winged males (Fawcett et al. 2018), suggesting that sexual conflict may affect the evolution of  
79 polyphenism in *J. haematoloma*.

80 Functional genetic approaches have begun to reveal insights into the developmental mechanisms  
81 controlling the polyphenism in *J. haematoloma*. RNA interference, as well as insulin injection,  
82 suggest that insulin signaling plays a critical role in the regulation of the threshold for short-wing

83 development (Fawcett et al. 2018). Numerous studies of polyphenism in other insects have  
84 highlighted the importance of insulin signaling as a mediator of nutritional information  
85 influencing tissue growth (Emlen et al. 2006; Snell-Rood & Moczek 2012; Xu et al. 2015;  
86 Nijhout & McKenna 2018). Functional tests of specific genes and pathways are important  
87 confirmations of their mechanistic roles. However, these approaches rely on candidates  
88 suggested by work in different systems. Thus, it is not clear right now whether insulin signaling  
89 really is a hotspot for the evolution of polyphenism and a central component of its regulation in  
90 diverse species, or whether this is an instance where biased sampling has led to a false  
91 conclusion of consistency.

92 Unbiased genetic insights into polyphenism are possible using high-throughput gene expression  
93 studies. In order to characterize the development of polyphenic traits in the wings, flight muscles  
94 and gonads of the soapberry bug, *Jadera haematoloma*, we sequenced transcriptomes from a  
95 multidimensional matrix of life stage (juvenile and nascent adult), tissue (dorsal thorax and  
96 gonads), sex and food availability. Modeling allows the identification of transcripts with  
97 significant associations to these factors. Our null expectation was to identify genes with known  
98 roles in the differentiation of ovaries, testes and muscle, the major constituent of the dorsal  
99 thorax. The data were also used to examine three questions: (1) Is there evidence for expression  
100 bias in regulatory transcription factors or signaling pathways between morphs? Is insulin  
101 signaling the only such pathway involved? (2) Are there transcripts that are differentially  
102 regulated in ovaries and testes that might explain reduced short-winged male fertility? (3) Are  
103 there transcripts associated with the residual influence of the nutritional environment after  
104 accounting for differences associated with morph?

## 105 Materials and methods

106 Detailed supplementary methods are available at <https://aphanotus.github.io/morphDE/>.

### 107 Sampling design

108 Soapberry bugs were collected from multiple field locations in the United States and reared in  
109 the lab as described by Fawcett et al. (2018). Transcriptome studies proceeded through three  
110 phases. Initially, we sampled whole bodies from 12 individuals, including three biological  
111 replicates of each sex-and-morph combination from the Plantation Key, FL population (phase 1).  
112 Next we sampled dorsal thorax and gonads from three individuals of each sex-and-morph  
113 combination from the Plantation Key, FL and Aurora, CO populations (phase 2). The head,  
114 thorax and abdomen were separated in transverse sections with a scalpel. The thorax was then  
115 bisected in a frontal section, just dorsal of the base of the legs. The ventral thorax was discarded.  
116 The gonads were then removed from the adult abdomen using forceps. Because juvenile gonads  
117 are very small, we instead sampled the juvenile abdomen after removing the gut. Samples from  
118 these phases were prepared as full-length 150-bp, paired-end Illumina libraries and used to

119 assemble a reference transcriptome. The final project phase utilized 3'-tag sequencing with the  
120 exclusive goal of quantifying gene expression. The sampling design at this phase included two  
121 stages (fifth instars and nascent adults), two tissues (dorsal thorax and gonads), two sexes, long-  
122 and short-wing adult morphs, high and low food regimes, four populations (collected from Key  
123 Largo, FL, Plantation Key, FL, Aurora, CO, and Frederick, MD), and three biological replicates  
124 of each unique combination of the preceding factors. The influence of population was not  
125 examined in this study. Instead sampling across populations ensures that any results apply  
126 broadly for the species. The full matrix contained 288 samples. Seven samples were lost or did  
127 not pass library quality controls, resulting in 281 samples in the analysis.

## 128 **Shape analysis**

129 Most *J. haematoloma* can be qualitatively identified as "long-winged" or "short-winged" based  
130 on the shape of the wings and whether their distal tips extend beyond the posterior of the  
131 abdomen when folded at rest. However, closer inspection reveals that some individuals have  
132 intermediate shapes and lengths (Fig. 1). We reasoned that gene expression was likely to also  
133 follow a more continuous distribution. Therefore, rather than relying on a categorical assignment  
134 of specimens into morphs, gene expression was modeled based on the quantification of wing  
135 shapes. (Categorical modeling by morph was also performed, and the two approaches identified  
136 similar lists of top DEGs.) Thoracic muscles are required for flight, therefore we also tested for  
137 correlations between gene expression and thorax shape.

138 Adult wing shapes were quantified using 24 landmarks as described by Fawcett et al. (2018; Fig.  
139 1G; Table S1). This system of landmarks was adapted to quantify juvenile wing pad shapes  
140 (Table S2). Thorax shapes of juvenile and adult bugs were digitized using a separate set of  
141 landmarks outlining the pronotum and mesonotum (Table S3). Bugs were anesthetized using  
142 CO<sub>2</sub> and imaged on a trinocular stereo microscope equipped with a Moticam 5 digital camera.  
143 Dorsal and ventral images of each specimen were recorded with a millimeter-scale ruler to scale  
144 pixel measurements. ImageJ was used to place landmarks. Variation in shapes was analyzed  
145 using the geometric morphometric methods implemented in the R packages *geomorph* (Baken et  
146 al. 2021) with visualizations provided by functions in the package *borealis* (Angelini 2021).  
147 Generalized Procrustes analysis with partial Procrustes superimposition was performed using  
148 minimized bending energy. For subsequent analyses, the shapes of individual wings, wing pads  
149 and thoraces were quantified along the first principal component axis (Fig. 1H, S1,S2).  
150 Procrustes ANOVA with permutation was used to assess hypotheses for patterns of shape  
151 variation among the aligned specimens, using 10,000 iterations (Collyer & Adams 2018).  
152 Modeling started by testing a simple allometry model, shape ~ log(centroid size). Factors of  
153 interest were tested individually, and if their effects were significant, they were tested in  
154 combination in more complex models. ANOVA tables for all models appear in the  
155 Supplementary Materials.

## 156 **Nucleic acid extractions**

157 Individuals were collected from laboratory cultures, anesthetized using CO<sub>2</sub> exposure,  
158 photographed, and flash frozen in liquid nitrogen. Samples were stored at -80°C until further  
159 processing. The thorax was removed from head and abdomen with a clean scalpel blade, and the  
160 dorsal thorax was separated above the legs while tissue was still frozen. Gonads were then  
161 dissected from the abdomen of adults. For juveniles, the entire abdomen was included. Nucleic  
162 acid extractions used the Invitrogen PureLink RNA extraction kit.

### 163 **Sequencing and assembly of a reference transcriptome**

164 Samples used for transcriptome assembly (phases 1 and 2) were delivered on dry ice to Beckman  
165 Coulter Genomics (Danvers, Massachusetts) for poly-A selection, preparation of 125 bp paired-  
166 end TruSeq libraries, and sequencing using Illumina HiSeq. Libraries from phases 1 and 2 were  
167 sequenced separately. Reads from all samples were trimmed using [Trimmomatic](#) version 0.33  
168 ([Bolger et al. 2014](#)) with the parameters `ILLUMINACLIP:/export/local/src/Trimmomatic-`  
169 `0.33/adapters/TruSeq3-SE.fa:2:30:10 HEADCROP:12 LEADING:3 TRAILING:3`  
170 `SLIDINGWINDOW:3:15 MINLEN:50`. The resulting reads were assembled into transcripts by  
171 [Trinity](#) version 2.0.6 ([Grabherr et al. 2011](#)). Before annotation, potentially redundant transcripts  
172 with greater than 90% identity were collapsed using CD-HIT-EST version 4.6 ([Li & Godzik](#)  
173 [2006](#)) with default parameters. [TransDecoder](#) (<https://github.com/TransDecoder/TransDecoder/>)  
174 was used to identify coding sequences. At each stage, [BUSCO](#) version 5.2.2 ([Manni et al. 2021](#))  
175 was used to assess transcriptome completeness. Consolidation by `cd-hit-est` dramatically  
176 reduced the number of sequences flagged as duplicates (Table S4).

### 177 **Annotation of the transcriptome**

178 Transcript annotation was performed by EnTAP ([Hart et al. 2019](#)) using the NCBI invertebrate  
179 RefSeq dataset, SwissProt and TrEMBL databases. EnTAP assigns sequence names, GO terms,  
180 KEGG pathway associations, and eggNOG protein domains based on sequence similarity. The *J.  
haematoloma* sequences were matched to other Hemiptera, primarily *Halyomorpha halys*, *Cimex  
lectularius* and *Nilaparvata lugens* for the RefSeq comparisons and to *Oncopeltus*, *Lygus* and  
183 *Riptortus* for comparisons to the TrEMBL database. A small number of contaminant sequences  
184 were identified from bacteria and yeasts. These sequences were flagged as contaminants, but  
185 remained in the analysis. Subsequent examination of differentially expressed transcripts and  
186 enrichment analysis did not find significant contributions from contaminant or xenic transcripts.  
187 From each model, differentially expressed transcripts were ranked by FDR-adjusted p-values,  
188 and those among the top 100 with no annotations were used as queries in BLASTx searches to  
189 the NCBI nr database for insects. Transcripts were manually annotated if they had significant  
190 matches (E-values < 10<sup>-5</sup>) to annotated genes from other insects.

191 **Sequencing for gene expression quantification**

192 For gene expression analysis, RNA samples were shipped overnight on dry ice to the [DNA](#)  
193 [Technologies Core](#) at the University of California at Davis. Libraries were prepared using the  
194 QuantSeq 3' mRNA-Seq Library Prep Kit (Lexogen, Greenland, New Hampshire, USA) in three  
195 batches for sequencing in separate Illumina HiSeq lanes. Batch 3 was sequenced twice and read  
196 counts for each run were added together for each sample after filtering.

197 These batches were sequenced in three lanes yielding 1.147 billion reads ( $115.9 \times 10^9$  bp). Short  
198 read sequences were inspected for quality using FastQC  
199 (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>), before and after trimming with  
200 Trimmomatic, using more stringent parameters than in the earlier reference assembly phase:  
201 ILLUMINACLIP:/export/local/src/Trimmomatic-0.33/adapters/TruSeq3-  
202 SE.fa:2:30:10 HEADCROP:16 LEADING:3 TRAILING:3 SLIDINGWINDOW:3:15 MINLEN:79.  
203 This filter retained 1,056,710,191 reads (92.1%; Table S5).

204 **Mapping 3seq reads to the reference transcriptome**

205 [Kallisto](#) was used to quantify the abundance of transcripts from 3seq data using a pseudo-  
206 alignment method ([Bray et al. 2016](#)). First, a Kallisto index was created from the assembled  
207 transcripts, collapsed by cd-hit-est. In order to run Kallisto on single-end reads, the program  
208 requires the mean and standard deviation of read lengths, which was determined using Linux  
209 commands (see Supplementary Methods). A [bash script](#) was used to run kallisto quant for each  
210 sample and to rename the output files with the sample names. The resulting output files were  
211 then merged into a single comma-separated values (CSV) file with columns for each sample.

212 **Differential expression modeling**

213 Transcript expression differences were tested for several models using the Bioconductor package  
214 DESeq2 version 1.32.0 ([Love et al. 2014](#)). The function DESeq fits a generalized linear model in  
215 which counts for each transcript are modeled using a negative binomial distribution with a  
216 specific dispersion parameter. Means were estimated using approximate posterior estimation  
217 ([Zhu et al. 2018](#)). We extracted  $\log_2$  fold-change and p-values corrected by independent  
218 hypothesis weighing ([Ignatiadis et al. 2016](#)) and applied a critical threshold of 0.05. All models  
219 included a batch effect to account for variance introduced by multiple sequencing runs. Variance  
220 stabilizing transformation, using the function DESeq::vst, was applied before clustering and  
221 ordination of samples by read count. Hierarchical clustering (Fig. S3) by k-means and single-  
222 linkage was examined using the pheatmap R package (Kolde 2019). Principal components axes  
223 and differential expression results were visualized by plots customized using the R package  
224 ggplot2 (Wickham 2016).

225 **Results**

226 **Shape variation**

227 Before modeling gene expression, we first used landmark-based geometric morphometrics to  
228 quantify phenotypes. We examined the shapes of the juvenile and adult thorax (Fig. S2), juvenile  
229 wing pads and adult wings (Fig. S1). Juvenile thorax shapes varied primarily in the relative  
230 width of mesonotum. We tested models of this shape variation using thorax centroid size, sex,  
231 food regime, food density and population of origin. None of these factors showed significant  
232 correlations to juvenile thorax shape. Adult thorax shapes (Fig. 1F) varied primarily in their  
233 relative width, for both pronotum and mesonotum (Fig. S2B), and could be modeled by centroid  
234 size (Procrustes ANOVA,  $F_{1,86} = 2.50, p = 0.028$ ), sex ( $F = 8.56, p < 10^{-4}$ ) and morph ( $F = 3.15,$   
235  $p = 0.0064$ ). Juvenile wing pads differed primarily in the relative proximo-distal position of the  
236 fixed internal landmarks (Fig. S1A). This shape variation correlated with centroid size ( $F_{1,46} =$   
237  $2.19, p = 0.0126$ ) and sex ( $F = 2.42, p = 0.0059$ ). Adult wing pads varied primarily along an axis  
238 capturing the differences between short- and long-winged morphs (Fig. 1; S1B). Variation in  
239 adult wing shapes could be described by a model including centroid size ( $F_{1,86} = 290.2, p < 10^{-4}$ ),  
240 sex ( $F = 36.7, p < 10^{-4}$ ) and morph ( $F = 20.4, p < 10^{-4}$ ). Relative effect sizes were comparable  
241 for each of these factors ( $z = 5.24, 5.34$ , and  $5.56$ , respectively).

242 **Modeling transcript expression**

243 We first assembled a reference transcriptome from whole bodies or the dorsal thorax and gonads  
244 of 36 nascent adults from two populations of *J. haematoloma*. This assembly contained 96.6% of  
245 predicted conserved genes according to BUSCO (Table S4). We then leveraged 3'-tag  
246 sequencing to quantify gene expression by mapping short 3'-seq reads to the reference  
247 transcriptome. We sampled across a multi-dimensional matrix of factors including population,  
248 sex, morph, tissue and food regime, producing a dataset with 281 samples.

249 As sequencing capacity increases the number of samples in gene expression studies, there has  
250 been a growing appreciation for the need to account for batch effects (e.g. Zhou et al. 2019; Chen  
251 et al. 2020). In this study, batch effects were introduced by the need to sequence samples in  
252 multiple lanes (Fig. 2A-B). DESeq2 allows variance from confounding factors, such as multiple  
253 sequencing lanes, to be accounted for as fixed factors in generalized linear models. Because each  
254 transcript is modeled independently, there is no overall effect size or metric of goodness-of-fit to  
255 compare models. However, ordination of the normalized counts that result from batch modeling,  
256 via principal components analysis, better reflects biologically-meaningful factors (Fig. 2C).  
257 Hierarchical clustering grouped samples mostly by tissue, sex and stage, without obvious  
258 clustering by wing morph, food regime or population of origin (Fig. S3). More complex models  
259 can be applied to examine specific contrasts and correlations of biological interest. These models  
260 partition the variance in expression sequentially to factors in the model. For more complex

261 models, less power exists to detect differences in the last factor, but it allows for the influence of  
262 confounding factors to first be disentangled.

### 263 **Expression comparisons of sex**

264 Sex is often an important factor effecting gene expression. While it is not our focus of study,  
265 sexually dimorphic gene expression has not previously been described in *J. haematoloma*, and  
266 examining our dataset for differences between females and males in each tissue is a useful proof  
267 of concept for our approach. Transcripts were differentially expressed by sex at both sampled  
268 stages and in both sampled tissues (Fig. 3). Adults had a larger number of sex-biased transcripts  
269 (84% in gonad; 6.1% in thorax) compared to juveniles (30% in gonads, 0.4% in thorax). The  
270 gonads also had more sexually dimorphic expression than did the thorax. Within the gonads,  
271 males had more significantly biased transcripts than did females, at both sampled stages (Fig.  
272 3C-D).

273 Several genes with roles in somatic sex determination in Hemiptera were among those  
274 differentially expressed, including *doublesex* (Just et al. 2022). Two *dsx* transcripts were  
275 consistently female-biased, while one was male-biased. Alternative splicing of *dsx* is a well-  
276 described determinant of sexually dimorphic somatic development (Salz & Erickson 2010).  
277 Expression of *fruitless*, a gene that promotes male courtship behavior in *D. melanogaster* (Demir  
278 & Dickson 2005), is strongly up-regulated (367-fold difference) in *J. haematoloma* testes. The  
279 DEAD-box helicase encoded by *maleless* (*mle*) increases chromatin accessibility on the X  
280 chromosome in *D. melanogaster* males to effect dosage compensation (Gelbart & Kuroda 2009).  
281 Expression of *mle* is strongly (6-fold) male-biased in the *J. haematoloma* adult gonads (Fig  
282 S4B). Genes involved in differences between testes and ovary development are also  
283 differentially expressed (Fig. 3C-D). Dimorphic expression occurs earlier for males (Fig. 3C),  
284 and males at both stages have greater expression of tubulin, microtubule motor proteins and  
285 chromatin remodeling factors. These proteins are all consistent with spermatogenesis. We also  
286 noted strongly male-biased (26-fold in juveniles; 706-fold in adult) expression of a transcript  
287 annotated as *shaggy/GSK3* in the gonads, reminiscent of the testes-limited expression of  
288 *gasket/mojoless* in *D. melanogaster*, a GSK3 paralog originated through transposition. The male  
289 gonad also has biased expression of *ERK*, which is required in somatic cells of *D. melanogaster*  
290 testes to for regulation of cell divisions in germline (Gupta et al. 2018). The germline regulator  
291 *encore* was strongly male-biased at both stages, although it is expressed in testes and ovaries of  
292 *D. melanogaster* (van Buskirk et al. 2000). The adult female gonad also shows differential  
293 expression of some chromatin remodeling factors, such as *Mcm6*, *CBX3A* and *NAP1*. Genes  
294 related to transcription and mRNA provisioning during oogenesis are also upregulated, including  
295 *TIF-IA*, *SSB*, and *me31B*, several cyclins, as well as some signaling pathway components,  
296 including *Delta*, *shade* and *FoxO*.

297 With a large number of differentially expressed genes, larger insights can be gained from  
298 examining the enrichment of annotation terms (Table S6), such as gene ontology (GO) terms,

299 among the differentially expressed transcripts. We chose to test for enrichment independently  
300 among transcripts with positive and negative correlations in each model (Hong et al. 2014). We  
301 also tested for enrichment of KEGG pathway terms and protein domains applied from the  
302 eggNOG database (Huerta-Cepas et al. 2018).

303 Juvenile stage comparisons by sex in the thorax had no enrichment of GO or KEGG terms,  
304 consistent with the relatively small proportion of differentially expressed transcripts in those  
305 models. In the adult thorax, females showed enrichment for GO and KEGG terms related to  
306 ribosome function, protein synthesis and cell division, suggesting potentially greater or earlier  
307 cell growth than males. In contrast, the adult male thorax was enriched for GO terms related to  
308 DNA replication and transposable element activity. Somatic movement of mobile elements has  
309 been linked to aging and neurological disorders in humans, mice and fruit flies (Siudeja et al.  
310 2021). In the juvenile gonad, females show a bias for terms related to metabolism and  
311 intracellular organization. Males were enriched for transcripts annotated with GO terms related  
312 to the cytoskeleton and with KEGG terms related to metabolism, protein production and folding,  
313 and the Hedgehog signaling pathway, which is required to maintain the male germline stem cell  
314 niche in *D. melanogaster* (Michel et al. 2012). Similarly *traffic jam*, a gene promoting ovarian  
315 germline stem cell differentiation, had strongly (37-fold) female-biased expression in adult  
316 gonads. Comparisons of adult gonads had relatively few enriched terms, including female-bias  
317 for GO terms related to cell structure and development.

### 318 **Expression comparisons of wing morph**

319 One of the main goals of this study was to characterize gene expression in different wing  
320 morphs. Because morphs differ in thoracic wing and muscle tissue, as well as in fecundity, we  
321 expected to identify genes related to differentiation of these tissues. We first modeled expression  
322 by batch and sex, then examined differential expression by qualitatively categorizing adults as  
323 long- or short-winged (Fig. 4A). As expected, expression in the thorax of long-winged bugs was  
324 biased for genes encoding muscle proteins including flightin, tropomyosin, troponin, titin, actin,  
325 myosin and paramyosin. This group also included multiple enzymes of cellular respiration and  
326 oxidative phosphorylation, ribosomal proteins, enzymes of fatty acid catabolism, endosome  
327 regulators such as *Rab5*, as well as *sarcomere length short (sals)*, which regulates muscle actin  
328 polymerization in *D. melanogaster* (Bai et al. 2007). Several transcription factors were up-  
329 regulated in long-winged morphs, including several with homology to LIM/PDZ domain proteins  
330 and the zinc-finger transcription factor, such as *castor* (Fig. S4I), and the Iroquois-family protein  
331 known as *CG11617* in *D. melanogaster* or *Mohawk* in mice (Fig. S4E). The Hippo pathway is a  
332 well-described mechanism regulating cell and organ growth. *Hippo* transcripts were significantly  
333 biased towards expression in long-winged thorax. Other signaling pathway components such as  
334 the estrogen-related receptor (*ErR*) and the Toll pathway component *ECSIT* were also up-  
335 regulated in the long-winged thorax. In fruit fly embryos, *ErR* promotes carbohydrate  
336 metabolism essential for growth (Tennessee et al. 2011). *FGF* is also up-regulated in the long-  
337 wing thorax, and it is expressed in the flight muscles of *D. melanogaster* where it attracts the

338 developing tracheae (Jarecki et al. 1999). In contrast to the large percentage of transcripts that  
339 were differentially expressed by morph in the adult thorax, there was very little differential  
340 expression in the gonads (Fig. S4; 0.23% in adult gonads overall; 0.13% in ovaries; 0.55% in  
341 testes). A notable exception is a 2.6-fold bias in expression of *Sox14* in the ovaries of long-  
342 winged bugs (Fig. S4B; S6N), compared to short-winged females. *Sox14* is also expressed in  
343 late-stage somatic follicle cells in the *D. melanogaster* ovary (Rust et al. 2020).

344 Binary categorization of bugs into morphs is subjective, and while uncommon, intermediate  
345 individuals do occur. Therefore, we examined gene expression models using the first principal  
346 component (PC1) of wing shape as a continuous factor (Fig. 4B). This approach identified all of  
347 the same transcript groups described above, but also highlighted others. For example, the long  
348 wing shapes were positively correlated with thoracic expression of genes both promoting and  
349 inhibiting cell growth, including several cell cycle regulators, *tumorous imaginal discs* and  
350 apoptosis-inducing factor. The TOR activator encoded by *Rheb* was up-regulated in the thorax of  
351 bugs with more long-winged shapes (Fig. S8E). *schnurri*, a transcription factor that mediates  
352 TGF $\beta$  signaling during development of *D. melanogaster* muscle and trachea (Staudt et al. 2005),  
353 was also up-regulated with more long-winged shapes (Fig. S6A). Short-winged bugs had higher  
354 thoracic expression of *FoxO* (Fig. S7I) and *charybde* (Fig. S8I). *FoxO* encodes a transcription  
355 factor that inhibits cell growth, and is negatively regulated by insulin signaling (Kramer et al.  
356 2003). *charybde* encodes an inhibitor of TOR, acting to limit cell growth and promote apoptosis  
357 (Reiling & Hafen 2004).

358 Enrichment of GO and KEGG terms was similar for differentially expressed transcripts in the  
359 thorax identified through modeling by morph and by wing shape (Table S6). These analyses  
360 highlighted GO terms associated with the mitochondrion, intracellular anatomical structure,  
361 cellular metabolism and the ribosome. The KEGG metabolic pathway map (map01100) was the  
362 top enrichment term for both models. These enrichment results are consistent with the increased  
363 energetic demands of flight muscle in the long-winged bug thorax.

364 Enrichment analyses for the gonad datasets did not agree between models using morph and wing  
365 shape. Because of the relatively small number of differentially expressed transcripts in the  
366 gonads, fewer terms were identified as enriched and these had much higher p-values than in the  
367 thorax analyses (Table S6). Therefore, these results should be considered skeptically. Analysis of  
368 the dataset containing gonad expression from males and females, modeling batch, sex and  
369 morph, highlighted GO terms related to ectoderm and mesoderm development and to cAMP  
370 signaling in the gonads of long-winged bugs. In contrast, modeling by wing shape associated  
371 long-winged bug gonads with GO terms including endopeptidase activity, the lysosome, cell  
372 death, autophagy and histolysis. Modeled by morph, the short-winged bug gonad was enriched  
373 for terms related to dormancy and diapause, phosphatidylinositol metabolism, and synapse  
374 function. Wing shape models did not identify any terms enriched in transcripts biased towards  
375 expression in the gonads of bugs with short-wing shapes. Expression in the ovaries of long-  
376 winged morphs was enriched for GO terms related to the spindle and microtubules,

377 metamorphosis, and cAMP metabolism, and for KEGG terms related to calcium signaling and  
378 synapse function. Short-winged ovary expression was enriched for GO terms related to peptidase  
379 and hydrolase inhibition. The testes of short-winged bugs were enriched for expression related to  
380 the cell cycle, axon guidance, immune function, cell cycle regulation and neurotrophin signaling.

381 Since polyphenism in *J. haematoloma* also affects thorax shape, we examined models of  
382 expression based on thorax shape PC1. No transcripts in the gonads were significantly correlated  
383 with thorax shape. Models of gene expression in the thorax based on thorax shape correlated  
384 with relatively few transcripts (Fig. 4C) with no obvious theme. Several cuticle proteins were  
385 biased towards wider thorax shapes, as was expression of *takeout*, a gene involved in male  
386 courtship and circadian behaviors in fruit flies, and the Zn-finger transcription factor *bowl*. The  
387 homeobox transcription factor *H2.0/CG11607* was strongly biased toward narrower thorax  
388 shapes.

### 389 **Expression comparisons of food availability**

390 The fecundity-dispersal trade-off in *J. haematoloma* is mediated by food availability during  
391 juvenile development (Dingle & Winchell 1997; Fawcett et al. 2018). There have also been  
392 reports that variation may exist within morphs for muscle development, muscle histolysis and the  
393 rates of gamete production (Carroll et al. 2003). Therefore, we examined the influence of food  
394 availability on gene expression. We modeled food availability in two ways. First we categorized  
395 cohorts as experiencing either a high or low food regime, if they had more or less than 1 host  
396 plant seed per individual. Food regime was modeled as a categorical factor. Second we modeled  
397 food density ( $\log_{10}$  seeds / bug) as a continuous covariate. These models first accounted for  
398 batch, sex and morph. Relatively little variance remains in the model after these factors, and we  
399 identified a small number of differentially expressed transcripts in more comparisons.

400 Differentially expressed transcripts included the lipid-binding protein vitellogenin, which was  
401 up-regulated in the thorax of well-fed bugs, under both models (Fig. S5B,F). Ankyrin, a cuticle  
402 protein and an odorant-binding protein were up-regulated in the thorax of low-food bugs (Fig.  
403 S5B,E,F). In the gonads, high food was associated with increased expression of a solute carrier  
404 (Fig. S5C,D), while low food was associated with expression of trypsin, multiple inositol  
405 polyphosphate phosphatase (*MNPP*), and *Mak10/CG15517*, a component of the N-terminal  
406 acetyltransferase C complex (Fig. S5H). The transcript most strongly differentially expressed in  
407 the gonad was *bursicon*, with 23-fold bias in the juvenile gonad under a low food regime (Fig.  
408 S5C). Bursicon is an endocrine protein that promotes cuticle hardening in insects (Fraenkel &  
409 Hsiao 1965).

410 The largest number of transcripts differentially expressed by food density were found in the  
411 juvenile testes (Fig. 5). Under high food, the testes showed biased expression of transcripts from  
412 several transposable elements, and genes with known roles in spermatogenesis, including *klhl10*  
413 (Arama et al. 2007), unconventional myosin (Hicks et al. 1999), formin (Kapoor et al. 2021) and

414 the nuclear pore importer *karyopherin- $\alpha$ 1* (*Kap- $\alpha$ 1*), which is required for male fertility in *D.*  
415 *melanogaster* (Ratan et al. 2008) and for the nuclear import of Yorkie, the transducer of the  
416 growth-regulating Hippo pathway (Wang et al. 2016). SNAP29 is also up-regulated in the testes  
417 of well-fed bugs. Other members of the tSNARE complex are known to be required for  
418 membrane fusions during spermatogenesis in fruit flies (Fabian & Brill 2012). This group  
419 includes an E2 ubiquitin-conjugating enzyme (*UbE2W*) as well. Similar genes, such as *D.*  
420 *melanogaster effete* are required for germline maintenance (Slaidina & Lehmann 2014). Low  
421 food density was correlated with biased expression of fatty acid synthase, a Zn-finger  
422 transcription factor (*ZMYM1*) and *parcas*, which promotes guanine exchange factor activity  
423 during development of muscle and the female germline in *D. melanogaster* (Beckett et al. 2006;  
424 Hamada-Kawaguchi et al. 2015). This group also included a transcript with homology to the *D.*  
425 *melanogaster ejaculatory bulb protein III*, a component of the mating plug.

426 Enrichment analysis identified several annotations related to nutrient uptake and metabolism in  
427 differentially expressed transcripts from the food regime and food density models (Table S6). In  
428 the juvenile and adult thorax, terms related to lipid-binding and transport were overrepresented in  
429 transcripts positively correlated with food density, while low food density and low food regime  
430 were associated with topoisomerase activity and terms related to signaling pathways including  
431 TGF $\beta$ , EGF and Hippo. In the juvenile gonad, transcripts expressed under high food conditions  
432 were enriched for GO terms related to redox and lipid catabolism, chromatin regulation and cell  
433 division and for KEGG pathways involved in MAP kinase signaling, protein secretion,  
434 endocytosis and immune function. In the adult gonads, high food was associated with terms  
435 related to thyroid hormone function. In insects, this likely reflects the involvement of JH  
436 metabolism. Transcripts biased towards expression in low food conditions in the gonads were  
437 enriched for terms linked to the hormonal control of molting and cuticle hardening, to protein  
438 digestion, and to inositol phosphate signaling. For models restricted to gene expression in the  
439 ovaries, transcripts biased toward high food density were enriched for terms related to  
440 spliceosome formation and mitochondrial structure in juveniles, and to lipid transport in adults.  
441 High food density was also associated with terms for mitochondrial function in the adult testes,  
442 as were terms related to germ cell determination and cell migration.

## 443 Discussion

444 In this study we have assembled transcriptomes from multiple life stages, tissues and morphs of  
445 the soapberry bug, *Jadera haematoloma*, in order to characterize the dispersal-fecundity  
446 polyphenism in this species at the developmental genetic level. We first assembled a high-quality  
447 reference transcriptome to provide a basis for gene expression quantification using 3'-tag  
448 sequencing. Generalized linear modeling implemented by DESeq2 was then employed to  
449 account for batch effects and identify transcripts differentially expressed with respect to each of  
450 the biological factors of interest.

451 The null expectation of this study was to identify numerous genes that are differentially  
452 expressed with respect to sex. Comparisons of expression in the gonads of adult females and  
453 males produced the most dramatic differences with 84% of transcripts differentially expressed  
454 (Fig. 3D). Even in the adult thorax, 6.1% of transcripts showed significantly sex-biased  
455 expression (Fig. 3B). Inspection of individual transcript annotations and enrichment analysis for  
456 GO and KEGG terms identifies many genes known to be involved in sex-specific processes such  
457 as sex determination, dosage compensation and gametogenesis. These data should be useful for  
458 future studies examining sexual differentiation in *J. haematoloma* and for comparative studies of  
459 sexually dimorphic development in insects. This confirmation of null expectations also gives us  
460 confidence to explore more complex models of gene expression that include additional factors.

## 461 **Morph development involves multiple pathways**

462 One of the major goals of this study is to identify regulatory transcription factors and signaling  
463 pathways that may regulate the specification and development of polyphenic morphs. Insulin  
464 signaling has already been implicated in this process (Fawcett et al. 2018), but other mechanisms  
465 may also contribute. Null expectations predicted higher expression of muscle-related genes in the  
466 thorax of long-winged bugs, and these were found to be a large number of the most strongly  
467 biased transcripts. Genes involved in protein synthesis, cellular respiration, fatty acid catabolism  
468 and endosome regulation were also identified. The growth-promoting kinase *Hippo* was also up-  
469 regulated in the long-wing thorax. *FGF*, which promotes tracheal branching in muscle, was also  
470 up-regulated, as was the transcription factor *schnurri* (Fig. S6A), which regulates TGF $\beta$   
471 signaling during muscle and tracheal development.

472 If we are interested in identifying early acting signals that induce development of each morph,  
473 we should look to early-acting regulators. Juveniles do not yet exhibit polyphenic traits.  
474 However, it may be possible to examine gene expression in juveniles raised under high and low  
475 food regimes, since this factor influences morph determination. Interestingly, none of the  
476 transcription factors or signaling pathway components above were found to differ in juvenile  
477 expression by food regime or food density. With the current dataset, it is not possible to say  
478 which individuals within a cohort obtain more or less nutrition. Juvenile *J. haematoloma* survive  
479 significantly better in groups (Ribeiro 1989). Juveniles reared in isolation often die, risking the  
480 introduction of confounding factors. This fact limits our power to detect genes responding  
481 directly to nutrition.

482 Therefore, while we can conclude that polyphenic development involves signaling via the  
483 insulin, TOR, Hippo, ErR and Toll pathways, it remains unclear which of these may act as  
484 primary inductive signals. In the brown planthopper, *Nilaparvata lugens*, RNA interference  
485 knockdown of the insulin receptor *InR1* causes development of almost entirely short-wing  
486 morphs, while RNAi of *InR2* and *FoxO* produce long-wing morphs (Xu et al. 2015). In contrast,  
487 RNAi targeting *InR1* and *InR2* in *J. haematoloma* did not alter wing morph frequencies, while  
488 *FoxO* RNAi increased the frequency of short-winged morphs (Fawcett et al. 2018). These

489 differences suggest that while insulin signaling may be a common component of polyphenisms  
490 affecting organ growth, the specifics of pathway involvement may differ among species. The  
491 proximate factors transferring nutritional information to the developmental system remain to be  
492 identified.

### 493 **Potential mediators of sexual conflict**

494 While short-winged *J. haematoloma* females benefit from increased gamete production, short-  
495 winged males of some populations have reduced fertility (Fawcett et al. 2018). This presents a  
496 sexual conflict, wherein the polyphenism is adaptive in females but maladaptive in males. To  
497 understand the mechanism of reduced male fertility, we looked for transcripts that were similarly  
498 up- or down-regulated by morph in the gonads of each sex. Such an expression pattern might be  
499 caused by pleiotropic regulation of transcripts. There were no transcripts that were differentially  
500 expressed in ovaries and testes, using a permissive cut-off of 0.1 for FDR-adjusted p-values.  
501 However, two transcription factors identified for biased expression in the long-winged thorax  
502 *Mohawk/CG11617* and *castor* were also slightly up-regulated in the testes of long-winged males,  
503 compared to short-winged males (Fig. S6H,L). These differences suggest possible effects on  
504 male fertility and pleiotropy linking polyphenic traits across the body.

### 505 **Nutrient utilization and responses to nutrient limitation**

506 Polyphenisms are dramatic examples of phenotypic plasticity because individuals show large,  
507 often nonlinear differences due to environmental influences. However, more subtle responses to  
508 the environment are much more common. All animals must adjust to changes in their nutritional  
509 environment. In *J. haematoloma*, while food availability during juvenile development has the  
510 dramatic effect of determining polyphenic morphs, this factor may also influence bugs in more  
511 subtle ways by affecting variation within morphs. For example, significantly more wing shape  
512 variation exists among short-wing bugs than long-wing bugs (Fawcett et al. 2018), with some  
513 individuals showing intermediate wing shapes (Fig. 1B,H).

514 By modeling the effects on gene expression of large influences, such as sex and morph, we were  
515 able to attribute residual variation by adding terms to consider food availability, either as a  
516 categorical determination (high vs. low food regimes) or as a continuous covariate (food  
517 density). These models highlight transcripts likely involved in nutrient utilization in the high  
518 food samples. Vitellogenin and other lipid-binding proteins are among this group. The testes of  
519 well-fed males show increased expression of several genes related to spermatogenesis. Up-  
520 regulation of the nuclear pore importer *karyopherin-α1*, which is known to facilitate Hippo  
521 signaling, also suggests that well-fed males invest their extra resources to increase sperm  
522 production. Interestingly, one of the most strongly biased transcripts in well-fed males was a  
523 reverse transcriptase (Fig. 5, r.t.) related to the R1DM retrotransposon of *D. melanogaster*,  
524 suggesting that this selfish-genetic element has evolved to exploit increased nutritional resources  
525 as well.

526 The response of bugs to low food conditions suggests several different ways in which the insects  
527 may deal with nutrient limitation. One such response involves changes to the abdominal cuticle,  
528 up-regulating cuticle protein in the thorax (Fig. S5F) and the cuticle-hardening hormone bursicon  
529 in the juvenile abdomen (Fig. S5C). The testes of poorly-fed juvenile males up-regulated  
530 expression of several regulatory proteins including the transcription factor *ZMYM1* and the  
531 guanine exchange factor *parcas*. Inactivation of *parcas* in *D. melanogaster* led to an increase in  
532 the size of oocyte ring canals (Hamada-Kawaguchi et al. 2015). Ring canals also exist during  
533 spermatogenesis, and it is possible that increased expression of *parcas* might reduce ring canal  
534 diameter, limiting the cytoplasmic provisioning of gametes produced under low nutrient  
535 conditions. Increased expression of the mating plug component *PebIII* also suggests that with  
536 lower nutritional resources males respond to maximize the potential reproductive success of  
537 sperm.

### 538 **Conclusions**

539 Polyphenic development allows organisms to respond to their environment with coordinated  
540 variations in multiple traits. We show how high-throughput RNA sequencing provides useful  
541 insights into this process in *J. haematoloma*. Several signaling pathways influence morph-  
542 specific thoracic development, including the insulin, TOR, Hippo, ErR and Toll pathways. We  
543 identify two transcription factors up-regulated in the long-wing thorax and testes of long-winged  
544 males, which suggest potential mechanisms for maladaptive male fertility differences by morph.  
545 Finally, we find that bugs respond to nutrient limitation with expression linked to changes in  
546 cuticle structure and spermatogenesis. In the future, functional genetic tests, using RNA  
547 interference or other methods, should test these hypotheses in soapberry bugs and other models  
548 of dispersal-fecundity polyphenism.

549 **Acknowledgements**

550 Randall Downter provided support for high-performance computing without which this work  
551 would not be possible. We would also like to thank Cory Williams and Lise Aubry for  
552 organizing the Symposium *Phenological plasticity: from molecular mechanisms to ecological*  
553 *and evolutionary implications* and for inviting us to be a part of it.

554 **Data Availability Statement**

555 The short read data underlying this article are available in the NCBI Short Read Archive under  
556 accessions TBD. Analysis scripts are available at <https://github.com/aphanotus/morphDE>.

557 **Funding**

558 Research reported in this publication was supported by the Colby College Division of Natural  
559 Sciences, by an Institutional Development Award (IDeA) from the National Institute of General  
560 Medical Sciences of the National Institutes of Health under grant number P20GM0103423, and  
561 by grant IOS-1350207 from the National Science Foundation to DRA.

562 **References**

- 563 Angelini DR. 2021. borealis: Tools for reproducible geometric morphometric analysis. R package.
- 564 Arama E, Bader M, Rieckhof GE, Steller H. 2007. A Ubiquitin Ligase Complex Regulates Caspase  
565 Activation During Sperm Differentiation in Drosophila. PLOS Biol 5:1–18.
- 566 Bai J, Hartwig JH, Perrimon N. 2007. SALs, a WH2-Domain-Containing Protein, Promotes Sarcomeric  
567 Actin Filament Elongation from Pointed Ends during Drosophila Muscle Growth. Dev Cell 13:828–  
568 42.
- 569 Baken EK, Collyer ML, Kaliontzopoulou A, Adams DC. 2021. geomorph v4.0 and gmShiny: Enhanced  
570 analytics and a new graphical interface for a comprehensive morphometric experience. Methods Ecol  
571 Evol 12:2355–63.
- 572 Beckett K, Baylies MK. 2006. Parcas, a regulator of non-receptor tyrosine kinase signaling, acts during  
573 anterior-posterior patterning and somatic muscle development in *Drosophila melanogaster*. Dev Biol  
574 299:176–92.
- 575 Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data.  
576 Bioinformatics 30:2114–20.
- 577 Bray NL, Pimentel H, Melsted P, Pachter L. 2016. Near-optimal probabilistic RNA-seq quantification.  
578 Nat Biotechnol 34:525–27.
- 579 Carlson M. 2021. GO.db: A set of annotation maps describing the entire Gene Ontology. R package.
- 580 Carroll SP, Boyd C. 1992. Host Race Radiation in the Soapberry Bug: Natural History with the History.  
581 Evolution 46:1053–69.
- 582 Carroll SP, Marler M, Winchell R, Dingle H. 2003. Evolution of Cryptic Flight Morph and Life History  
583 Differences During Host Race Radiation in the Soapberry Bug, *Jadera haematoloma* Herrich-  
584 Schaeffer (Hemiptera: Rhopalidae). Ann Entomol Soc Am 96:135–43.
- 585 Chen W, Zhang S, Williams J, Ju B, Shaner B, Easton J, Wu G, Chen X. 2020. A comparison of methods  
586 accounting for batch effects in differential expression analysis of UMI count based single cell RNA  
587 sequencing. Comput Struct Biotechnol J 18:861–73.
- 588 Collyer ML, Adams DC. 2018. RRPP: An r package for fitting linear models to high-dimensional data  
589 using residual randomization. Methods Ecol Evol 9:1772–79.
- 590 Dingle H, Winchell R. 1997. Juvenile hormone as a mediator of plasticity in insect life histories. Arch  
591 Insect Biochem Physiol 35:359–73.
- 592 Emlen DJ, Szafran Q, Corley LS, Dworkin I. 2006. Insulin signaling and limb-patterning: candidate  
593 pathways for the origin and evolutionary diversification of beetle “horns.” Heredity 97:179–91.
- 594 Fabian L, Brill JA. 2012. Drosophila spermiogenesis. Spermatogenesis 2:197–212.
- 595 Fawcett MM, Parks MC, Tibbetts AE, Swart JS, Richards EM, Vanegas JC, Cenzer M, Crowley L,  
596 Simmons WR, Hou WS, Angelini DR. 2018. Manipulation of insulin signaling phenocopies evolution  
597 of a host-associated polyphenism. Nat Commun 9:1699.

- 598 Fraenkel G, Hsiao C. 1965. Bursicon, a hormone which mediates tanning of the cuticle in the adult fly  
599 and other insects. *J Insect Physiol* 11:513–56.
- 600 Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury  
601 R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C,  
602 Lindblad-Toh K, Friedman N, Regev A. 2011. Full-length transcriptome assembly from RNA-Seq  
603 data without a reference genome. *Nat Biotechnol* 29:644–52.
- 604 Gupta S, Varshney B, Chatterjee S, Ray K. 2018. Somatic ERK activation during transit amplification is  
605 essential for maintaining the synchrony of germline divisions in *Drosophila* testis. *Open Biol*  
606 8:180033.
- 607 Hamada-Kawaguchi N, Nishida Y, Yamamoto D. 2015. Btk29A-Mediated Tyrosine Phosphorylation of  
608 Armadillo/β-Catenin Promotes Ring Canal Growth in Drosophila Oogenesis. *PLoS One* 10:1–13.
- 609 Harrison RG. 1980. Dispersal Polymorphisms in Insects. *Annu Rev Ecol Syst* 11:95–118.
- 610 Hart AJ, Ginzburg S, Xu M (Sam), Fisher CR, Rahmatpour N, Mitton JB, Paul R, Wegrzyn JL. 2020.  
611 EnTAP: Bringing faster and smarter functional annotation to non-model eukaryotic transcriptomes.  
612 *Mol Ecol Resour* 20:591–604.
- 613 Hicks JL, Deng W-M, Rogat AD, Miller KG, Bownes M. 1999. Class VI Unconventional Myosin is  
614 Required for Spermatogenesis in *Drosophila*. *Mol Biol Cell* 10:4341–53.
- 615 Hong G, Zhang W, Li H, Shen X, Guo Z. 2014. Separate enrichment analysis of pathways for up- and  
616 downregulated genes. *J R Soc Interface* 11:20130950.
- 617 Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H, Mende DR, Letunic  
618 I, Rattei T, Jensen LJ, von Mering C, Bork P. 2018. eggNOG 5.0: a hierarchical, functionally and  
619 phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic  
620 Acids Res* 47:D309–14.
- 621 Ignatiadis N, Klaus B, Zaugg JB, Huber W. 2016. Data-driven hypothesis weighting increases detection  
622 power in genome-scale multiple testing. *Nat Methods* 13:577–80.
- 623 Jarecki J, Johnson E, Krasnow MA. 1999. Oxygen Regulation of Airway Branching in *Drosophila* Is  
624 Mediated by Branchless FGF. *Cell* 99:211–20.
- 625 Just J, Laslo M, Lee YJ, Yarnell M, Zhang Z, Angelini DR. 2022. Distinct developmental mechanisms  
626 influence sexual dimorphisms in the milkweed bug *Oncopeltus fasciatus*. *bioRxiv*.  
627 <https://www.biorxiv.org/content/10.1101/2021.05.12.443917v3>
- 628 Kapoor T, Dubey P, Shirolikar S, Ray K. 2021. An actomyosin clamp assembled by the Amphiphysin-  
629 Rho1-Dia/DAAM-Rok pathway reinforces somatic cell membrane folded around spermatid heads.  
630 *Cell Rep* 34:108918.
- 631 Kolde R. 2019. pheatmap: Pretty Heatmaps. R package.
- 632 Kramer J, Davidge J, Lockyer J, Staveley B. 2003. Expression of *Drosophila* FOXO regulates growth and  
633 can phenocopy starvation. *BMC Dev Biol* 3:5.
- 634 Laslo M, Just J, Angelini DR. 2022. Theme and variation in the evolution of insect sex determination. *J  
635 Exp Zool B Mol Dev Evol* 1–20.

- 636 Levins R. 1968. Evolution in changing environments: some theoretical explorations, Monographs in  
637 Population Biology Princeton, New Jersey: Princeton University Press.
- 638 Li W, Godzik A. 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or  
639 nucleotide sequences. Bioinformatics 22:1658–59.
- 640 Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq  
641 data with DESeq2. Genome Biol 15:1–34.
- 642 Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. 2021. BUSCO Update: Novel and  
643 Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of  
644 Eukaryotic, Prokaryotic, and Viral Genomes. Mol Biol Evol 38:4647–54.
- 645 Matunis EL, Stine RR, Cuevas M de. 2012. Recent advances in Drosophila male germline stem cell  
646 biology. Spermatogenesis 2:137–44.
- 647 Michel M, Kupinski AP, Raabe I, Bökel C. 2012. Hh signalling is essential for somatic stem cell  
648 maintenance in the Drosophila testis niche. Development 139:2663–69.
- 649 Nijhout HF, McKenna KZ. 2018. The distinct roles of insulin signaling in polyphenic development. Curr  
650 Opin Insect Sci 25:58–64.
- 651 Ratan R, Mason DA, Sinnot B, Goldfarb DS, Fleming RJ. 2008. Drosophila Importin  $\alpha$ 1 Performs  
652 Paralog-Specific Functions Essential For Gametogenesis. Genetics 178:839–50.
- 653 Ribeiro ST. 1989. Group Effects and Aposematism in Jadera haematoloma (Hemiptera: Rhopalidae). Ann  
654 Entomol Soc Am 82:466–75.
- 655 Roff DA. 1990. The Evolution of Flightlessness in Insects. Ecol Monogr 60:389–421.
- 656 Rust K, Byrnes LE, Yu KS, Park JS, Sneddon JB, Tward AD, Nystul TG. 2020. A single-cell atlas and  
657 lineage analysis of the adult Drosophila ovary. Nat Commun 11:5628.
- 658 Siudeja K, van den Beek M, Riddiford N, Boumard B, Wurmser A, Stefanutti M, Lameiras S, Bardin AJ.  
659 2021. Unraveling the features of somatic transposition in the *Drosophila* intestine. EMBO J  
660 40:e106388.
- 661 Slaidina M, Lehmann R. 2014. Translational control in germline stem cell development. J Cell Biol  
662 207:13–21.
- 663 Smith ER, Pannuti A, Gu W, Steurnagel A, Cook RG, Allis CD, Lucchesi JC. 2000. The Drosophila MSL  
664 Complex Acetylates Histone H4 at Lysine 16, a Chromatin Modification Linked to Dosage  
665 Compensation . Mol Cell Biol 20:312–18.
- 666 Snell-Rood EC, Moczek AP. 2012. Insulin signaling as a mechanism underlying developmental plasticity:  
667 the role of FOXO in a nutritional polyphenism. PLoS One 7:e34857.
- 668 Staudt N, Molitor A, Somogyi K, Mata J, Curado S, Eulenberg K, Meise M, Siegmund T, Häder T,  
669 Hilfiker A, Brönner G, Ephrussi A, Rørth P, Cohen SM, Fellert S, Chung H-R, Piepenburg O, Schäfer  
670 U, Jäckle H, Vorbrüggen G. 2005. Gain-of-Function Screen for Genes That Affect Drosophila Muscle  
671 Pattern Formation. PLOS Genet 1:1–8.
- 672 Temple Lang D. 2022. RCurl: General Network (HTTP/FTP/...) Client Interface for R. R package.

- 673 Tennessen JM, Baker KD, Lam G, Evans J, Thummel CS. 2011. The Drosophila Estrogen-Related  
674 Receptor Directs a Metabolic Switch that Supports Developmental Growth. *Cell Metab* 13:139–48.
- 675 Treidel LA, Quintanilla Ramirez GS, Chung DJ, Menze MA, Vázquez-Medina JP, Williams CM, Lisa  
676 Treidel CA. 2022. Selection on dispersal drives evolution of metabolic capacities for energy  
677 production in female wing-polymorphic sand field crickets, *Gryllus firmus*. *J Evol Biol* 00:1–11.
- 678 Van Buskirk C, Hawkins NC, Schupbach T. 2000. Encore is a member of a novel family of proteins and  
679 affects multiple processes in *Drosophila* oogenesis. *Development* 127:4753–62.
- 680 Wang S, Lu Y, Yin M-X, Wang C, Wu W, Li J, Wu W, Ge L, Hu L, Zhao Y, Zhang L. 2016. Importin  $\alpha$ 1  
681 Mediates Yorkie Nuclear Import via an N-terminal Non-canonical Nuclear Localization Signal\*. *J  
682 Biol Chem* 291:7926–37.
- 683 Wickham H. 2016. Ggplot2: Elegant Graphics for Data Analysis. R package.
- 684 Xu H-J, Xue J, Lu B, Zhang X-C, Zhuo J-C, He S-F, Ma X-F, Jiang Y-Q, Fan H-W, Xu J-Y, Ye Y-X,  
685 Pan P-L, Li Q, Bao Y-Y, Nijhout HF, Zhang C-X. 2015. Two insulin receptors determine alternative  
686 wing morphs in planthoppers. *Nature* 519:464–67.
- 687 Zhou W, Koudijs KKM, Böhringer S. 2019. Influence of batch effect correction methods on drug induced  
688 differential gene expression profiles. *BMC Bioinformatics* 20:437.
- 689 Zhu A, Ibrahim JG, Love MI. 2018. Heavy-tailed prior distributions for sequence count data: removing  
690 the noise and preserving large differences. *Bioinformatics* 35:2084–92.
- 691 Zhuo J-C, Zhang H-H, Hu Q-L, Zhang J-L, Lu J-B, Li H-J, Xie Y-C, Wang W-W, Zhang Y, Wang H-Q,  
692 Huang H-J, Lu G, Chen J-P, Li J-M, Tu Z-J, Zhang C-X. 2021. A feminizing switch in a  
693 hemimetabolous insect. *Sci Adv* 7:9237.

694 **Figure Legends**

695 **Fig. 1**

696 Morphs of *Jadera haematoloma* differ visibly in wing shape. A typical long-winged bug (A) has  
697 wings extending beyond the posterior abdomen. Short-winged bugs have a much greater wing  
698 shape variation with rare intermediate individuals (B) and more common short-winged  
699 individuals with much smaller, foreshortened wings (C). Panels A-C show all females. Flight  
700 muscles (D) are only present in flight-capable, long-winged adults. The thorax of short-winged  
701 adults contains tissue that is not organized into sarcomeres (E). Shape variation in the adult  
702 thorax (F) and wings (G) can be compared following Procrustes alignment. Anterior is up in  
703 panels F-G. Outlines connect the consensus landmark positions. Morphospace plotting the first  
704 two principal component axes for the wing shape (H) separates bugs by morph. Specimens are  
705 color-coded by sex and morph. The percent of total shape variance described by each principal  
706 component is given on the axis. The wing shapes of the bugs photographed in panels A-C are  
707 indicated within the shape space.

708 **Fig. 2 Null model PCA**

709 Principal components of transcript counts for each sample, before (A, B) and after (C) modeling  
710 and normalization for batch effects. Color-coded by biological factors (A, C) and sequencing  
711 batch (B). The percentage of variance described by each PC axis is given in parentheses. Axes  
712 display a fixed aspect ratio.

713 **Fig. 3 Expression comparisons by sex**

714 Volcano plots, showing  $\log_2$  fold-change (LFC) and  $-\log_{10}$  FDR-adjusted p-values for contrasts  
715 by sex in the juvenile thorax (A), adult thorax (B), juvenile gonads (C) and adult gonads (D).  
716 Transcripts are denoted by solid dots if they have an absolute LFC value  $> 1$  (two-fold  
717 difference) and an FDR-adjusted p-value  $< 0.05$  (A) or  $10^{-3}$  (B-D). Transcripts associated with  
718 transposons are labeled “r.t.”.

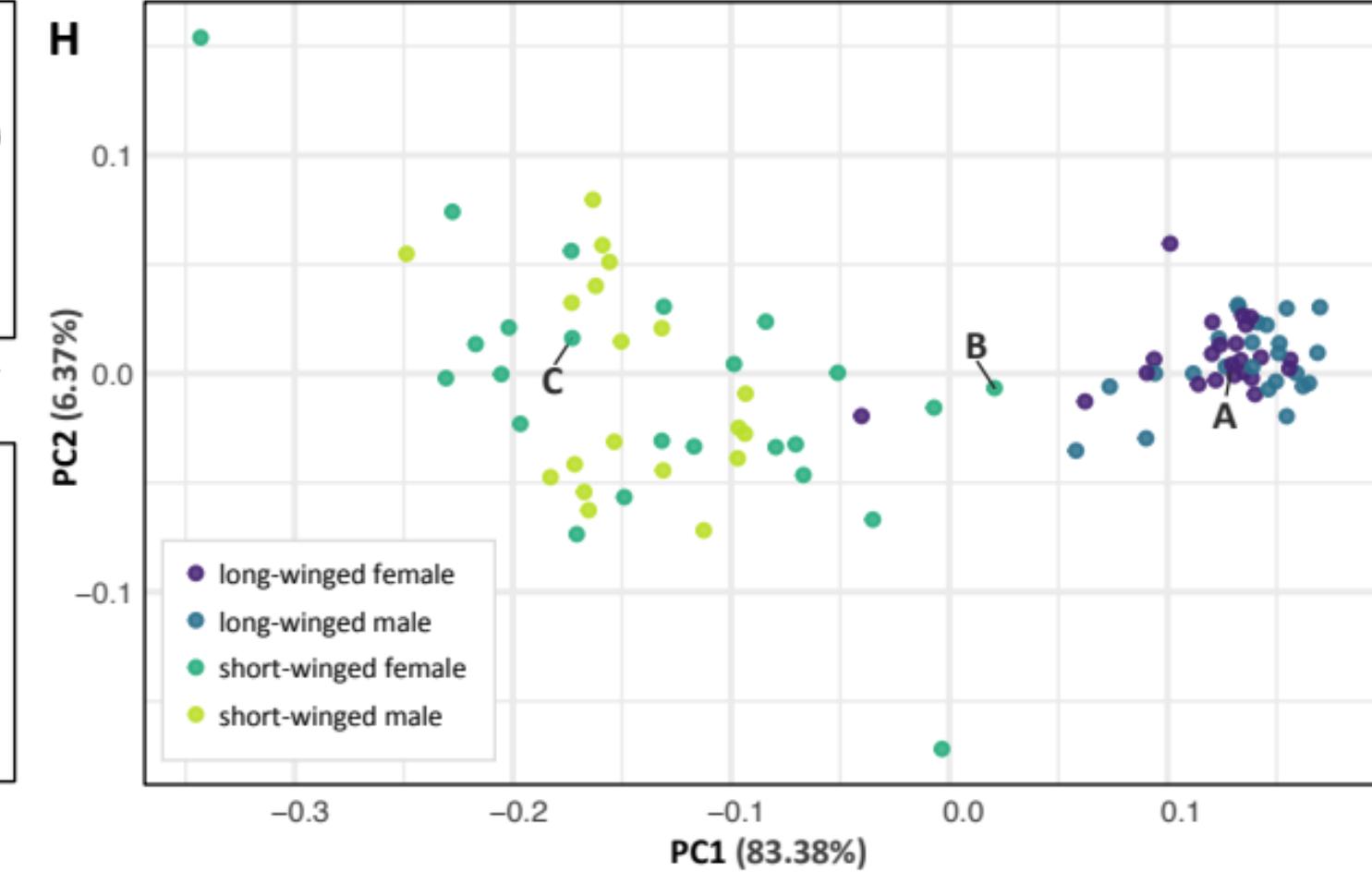
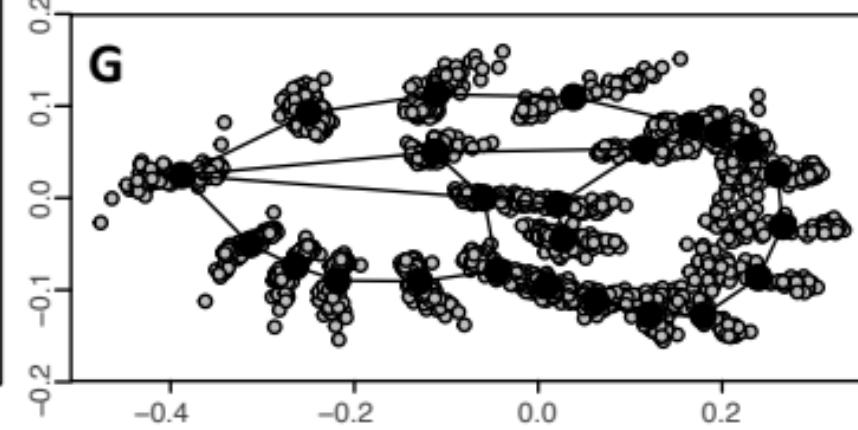
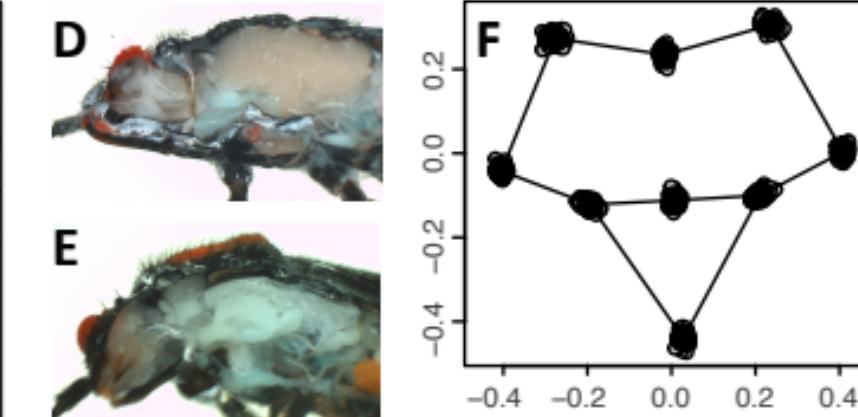
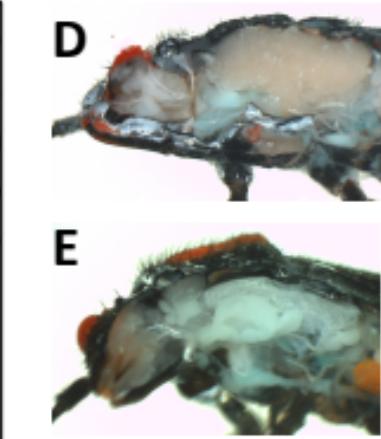
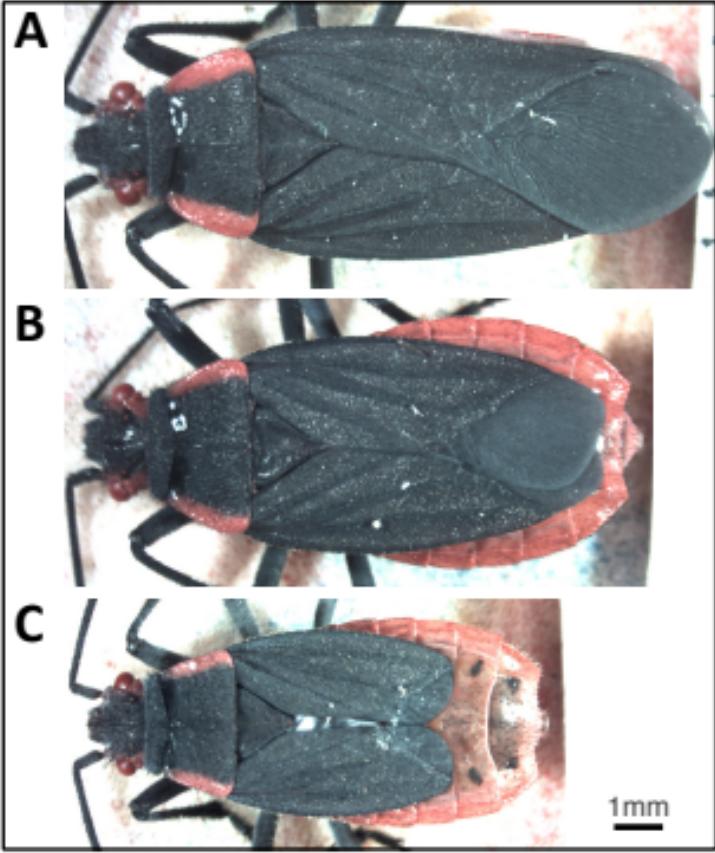
719 **Fig. 4 Expression comparisons by morph**

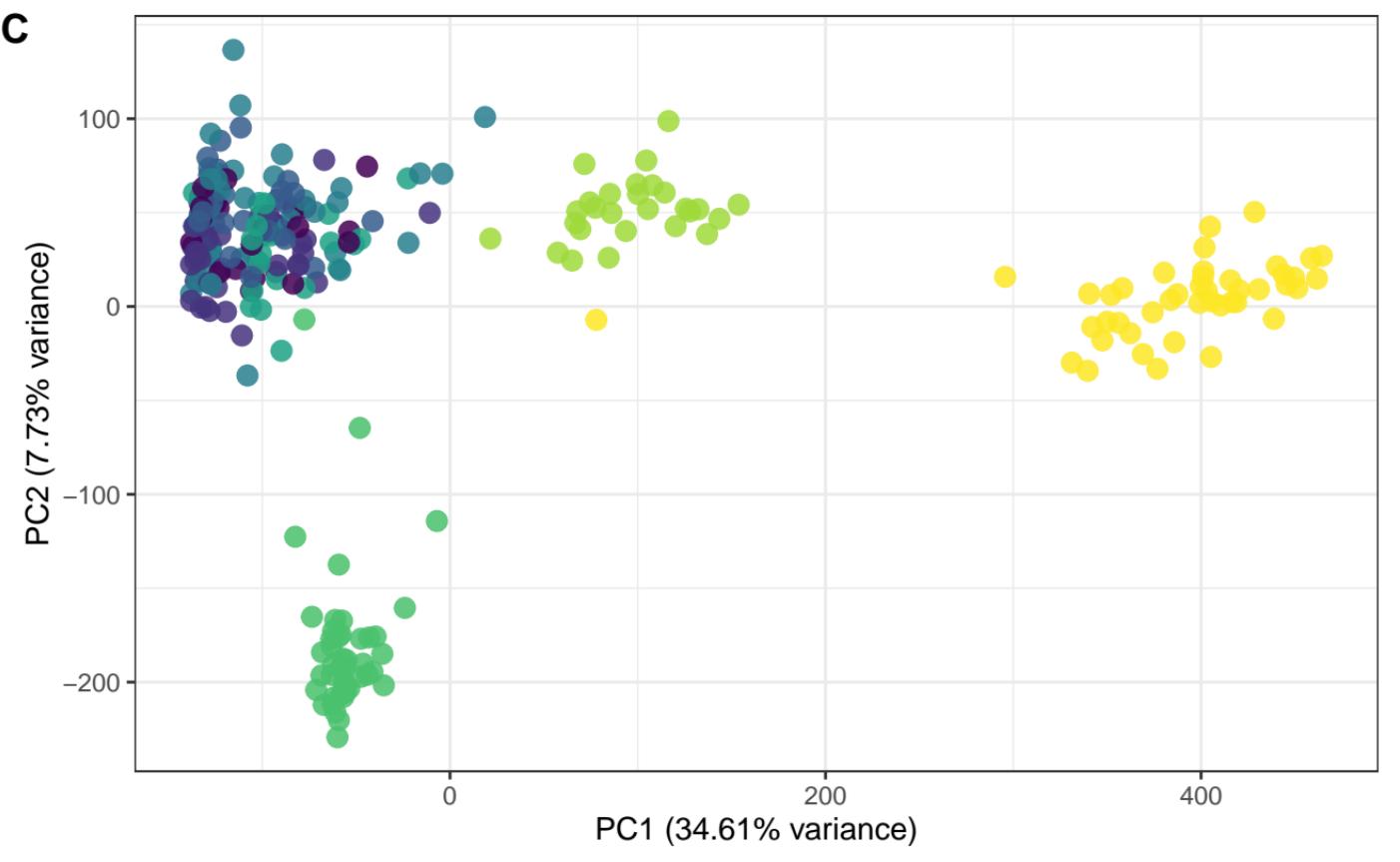
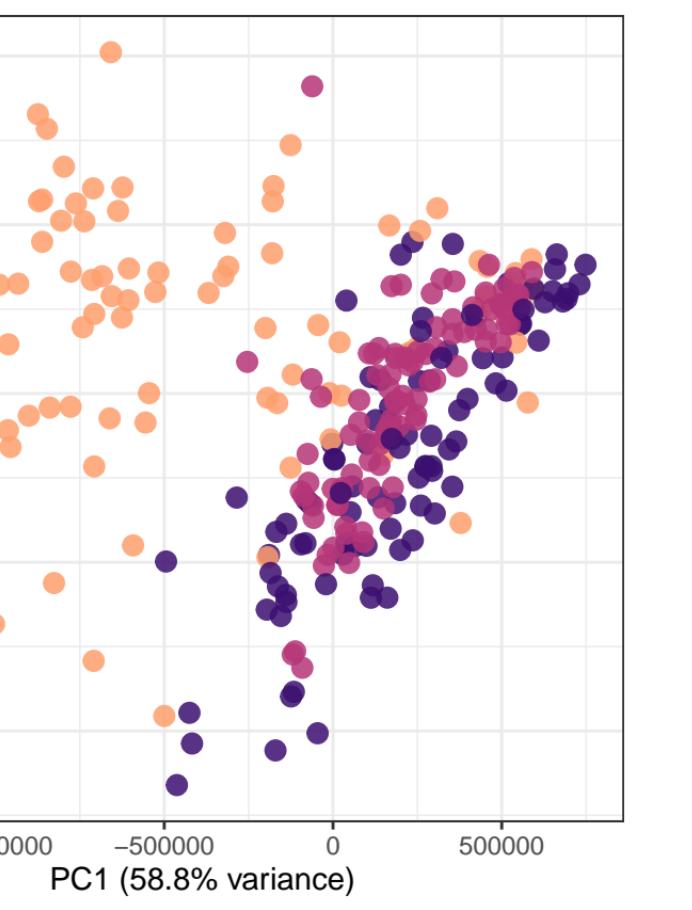
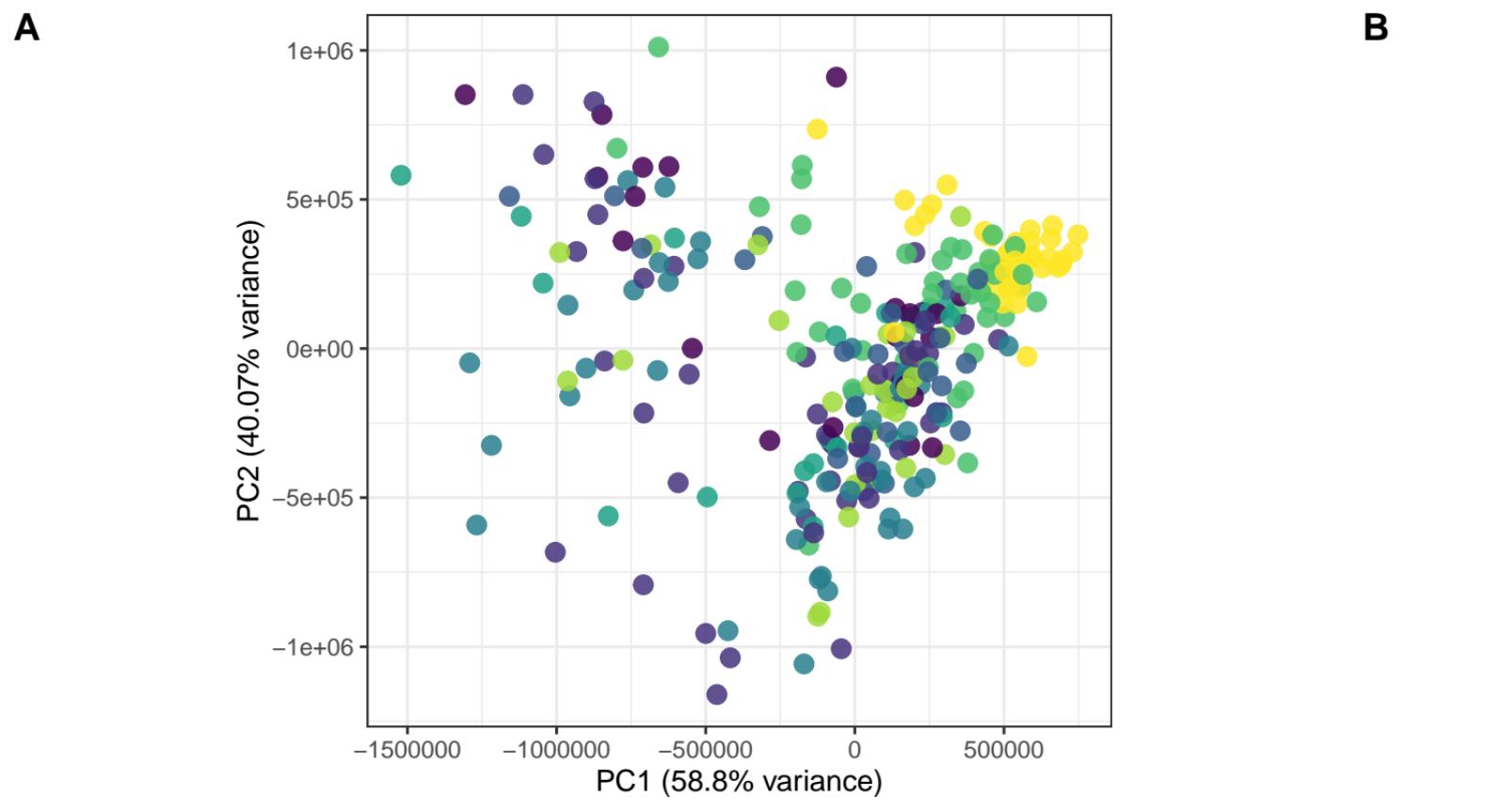
720 Volcano plots, showing  $\log_2$  fold-change (LFC) and  $-\log_{10}$  FDR-adjusted p-values from models  
721 of gene expression in the adult thorax based on correlation to morph (A), wing shape PC1 (B),  
722 and thorax shape PC1 (C). Greater log-fold change is associated with long-wing morphs (A),  
723 long-wing shapes (B) or wider thorax shapes (C). Transcripts are denoted by solid dots if they  
724 have an absolute LFC value  $> 1$  (two-fold difference) and an FDR-adjusted p-value  $< 10^{-3}$  (A-B)  
725 or 0.05 (C). Transcripts associated with transposons are labeled “r.t.”.

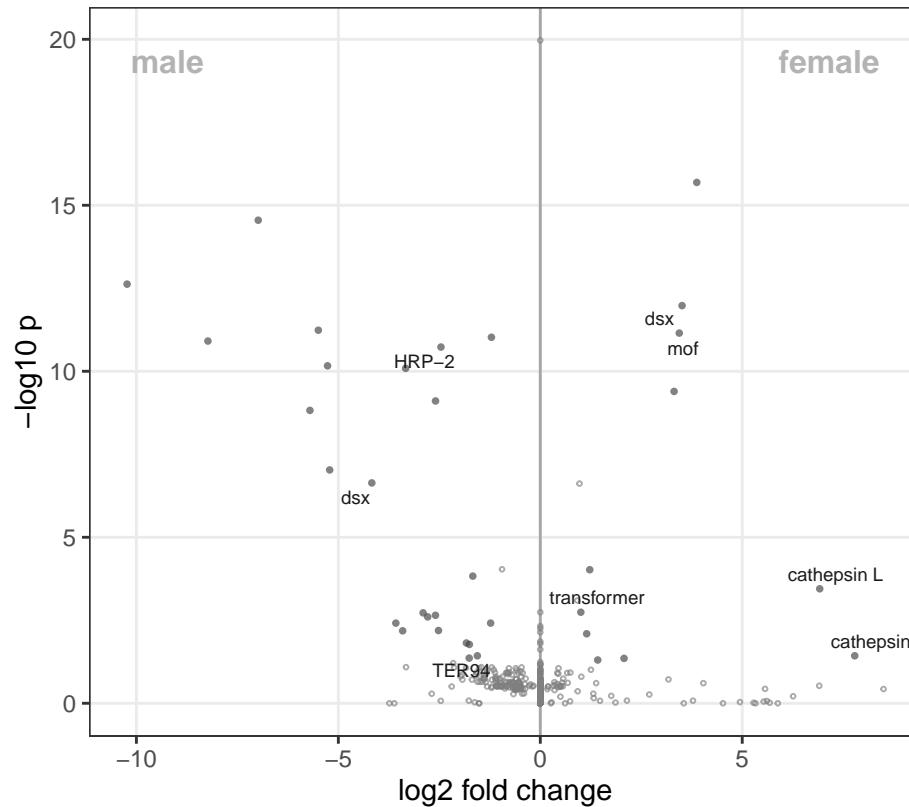
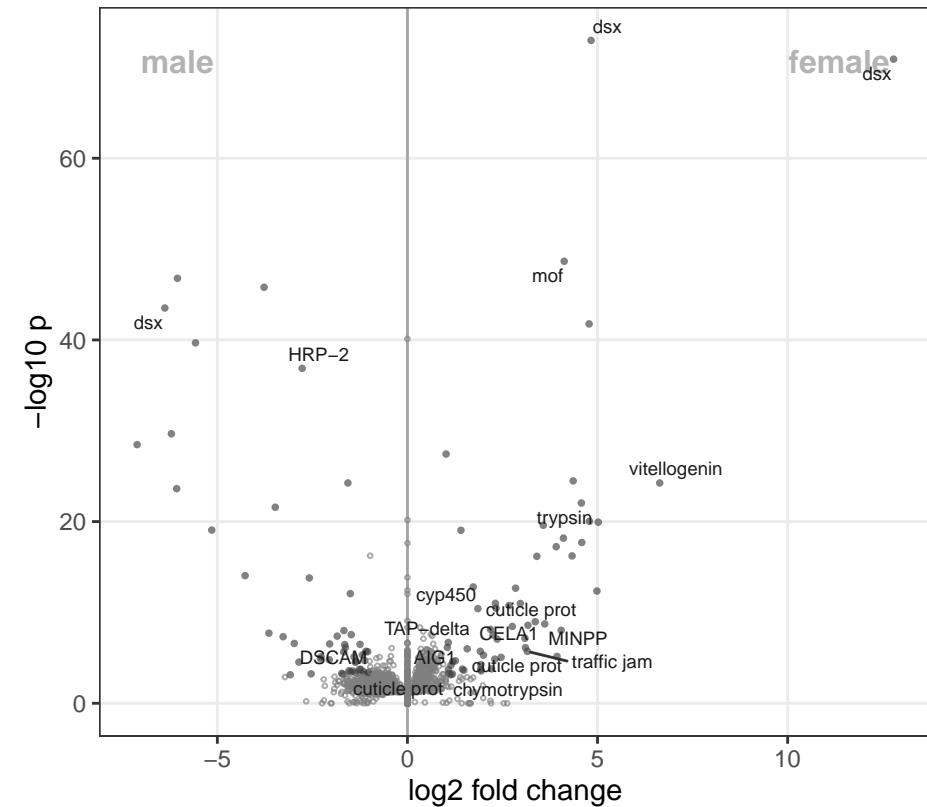
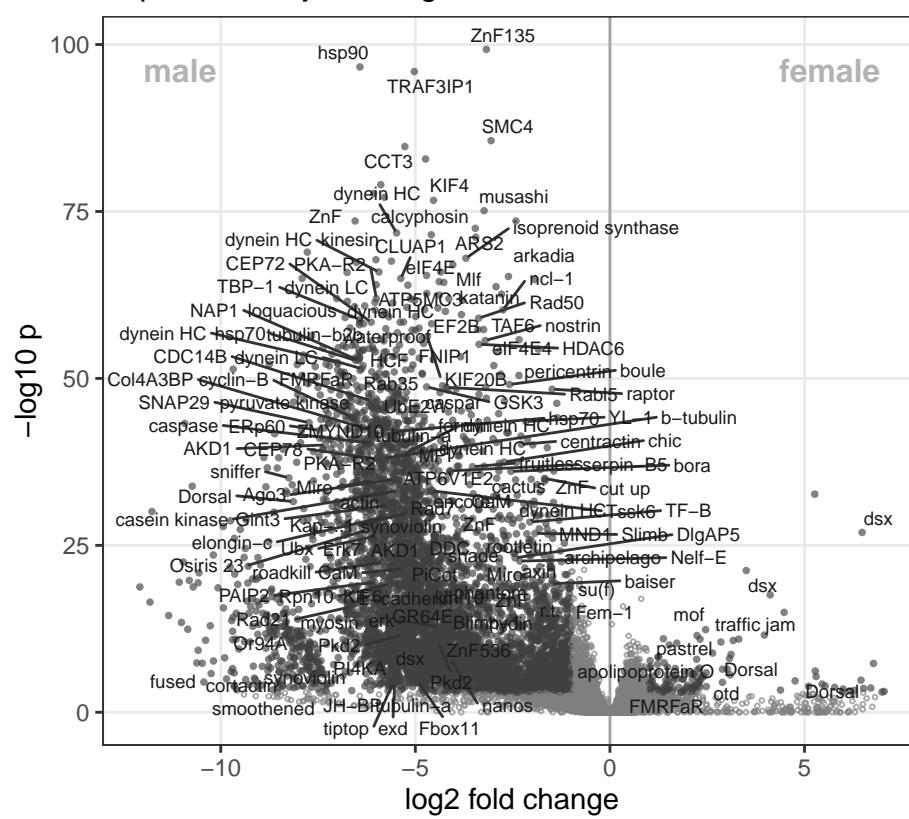
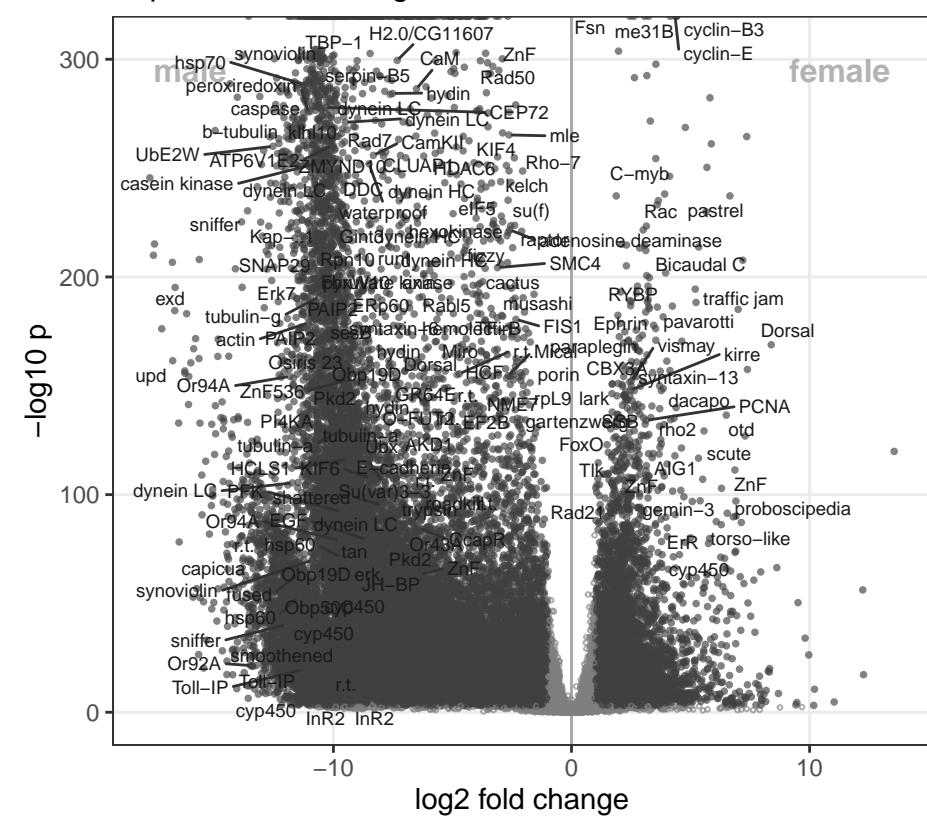
726 **Fig. 5 Expression comparisons by food density in juvenile testes**

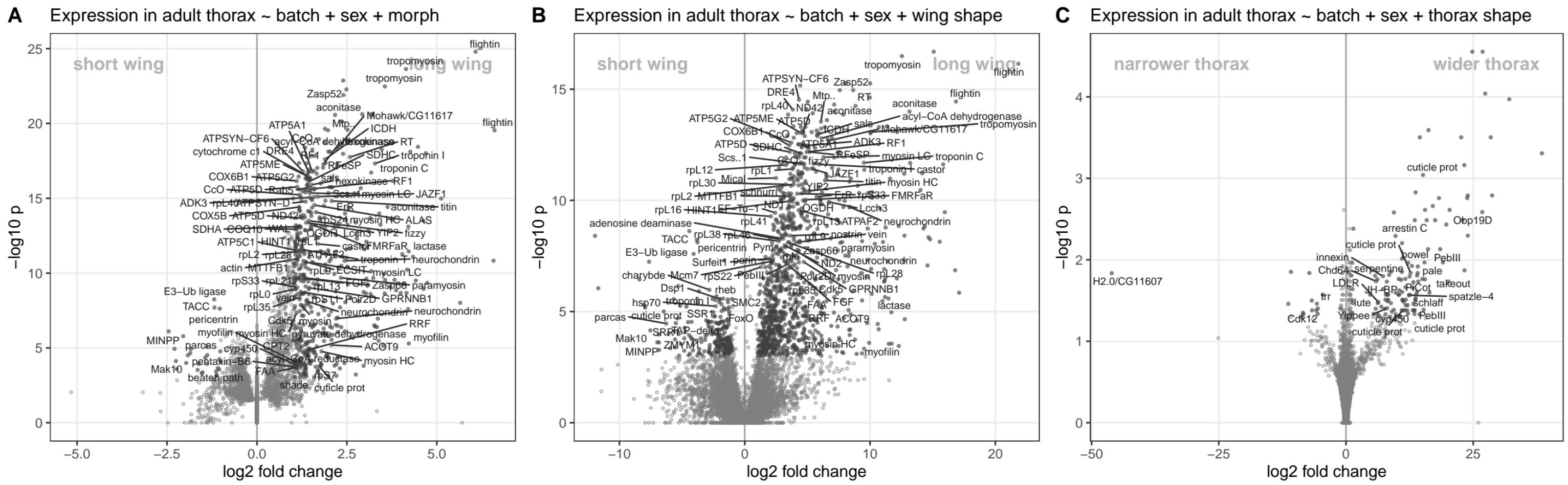
727 Volcano plots, showing  $\log_2$  fold-change (LFC) and  $-\log_{10}$  FDR-adjusted p-values from a model  
728 of gene expression in juvenile testes based on correlation to food density. Greater log-fold  
729 change is associated with increased food density. Transcripts are denoted by solid dots if they

- 730 have an absolute LFC value > 1 (two-fold difference) and an FDR-adjusted p-value < 0.05.  
731 Transcripts associated with transposons are labeled “r.t.”.





**A Expression in juvenile thorax ~ batch + sex****B Expression in adult thorax ~ batch + sex****C Expression in juvenile gonad ~ batch + sex****D Expression in adult gonad ~ batch + sex**

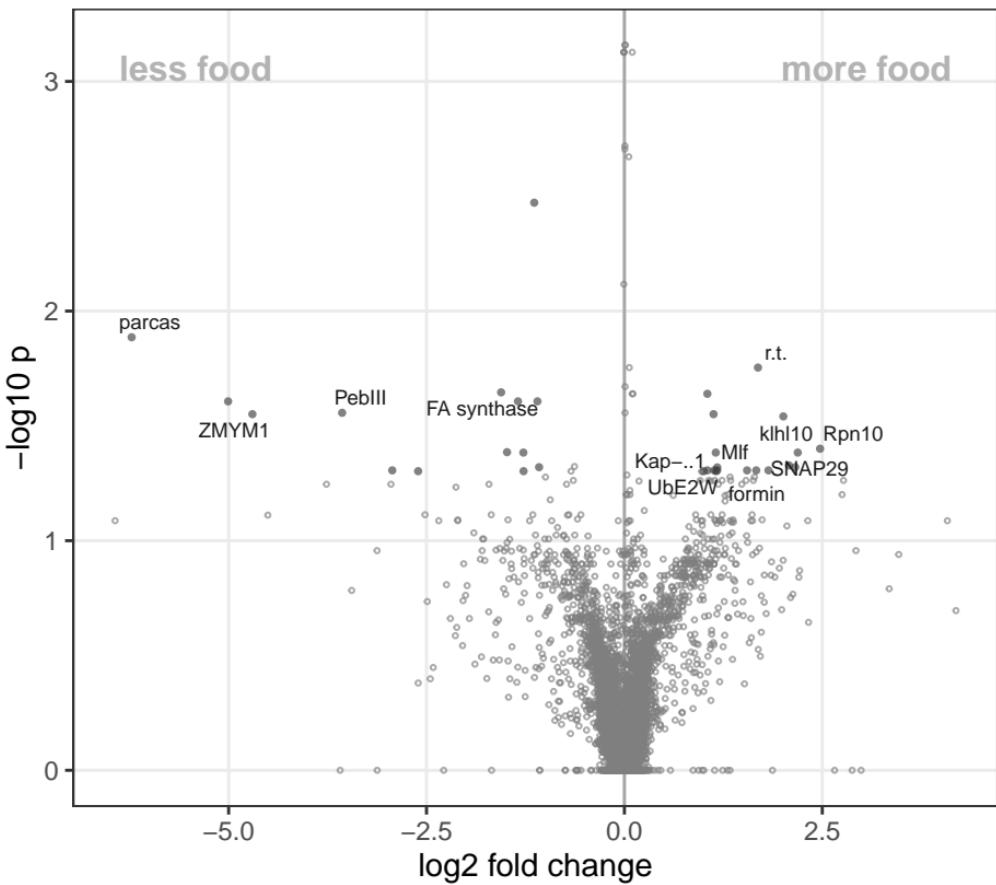


Samples: 91; Transcripts: 32368  
short wing-biased: 209 (0.65%), long wing-biased: 972 (3%), at p-value cutoff 0.001

Samples: 90; Transcripts: 32236  
short wing-biased: 236 (0.73%), long wing-biased: 969 (3%), at p-value cutoff 0.001

Samples: 90; Transcripts: 32236  
narrower thorax-biased: 37 (0.11%), wider thorax-biased: 169 (0.52%), at p-value cutoff 0.05

# Expression in juvenile testes ~ batch + food density



Samples: 27; Transcripts: 24508

less food-biased: 20 (0.082%), more food-biased: 29 (0.12%), at  $p$ -value cutoff 0.05