Title

The early embryonic transcriptome of a Hawaiian Drosophila picture-wing fly shows evidence of altered gene expression and novel gene evolution

Journal

Journal of Experimental Zoology Part B Molecular and Developmental Evolution, 338(5)

ISSN

1552-5007

Authors

Chenevert, Madeline
Miller, Bronwyn
Karkoutli, Ahmad
et al.

Publication Date

2022-07-01

DOI

10.1002/jez.b.23129

Peer reviewed

Received: 15 October 2021 | Revised: 14 January 2022 | Accepted: 13 February 2022

DOI: 10.1002/jez.b.23129

RESEARCH ARTICLE

JEZ-B MOLECULAR AND DEVELOPMENTAL EVOLUTION WILEY

The early embryonic transcriptome of a Hawaiian *Drosophila* picture-wing fly shows evidence of altered gene expression and novel gene evolution

Madeline Chenevert[1,2] | Bronwyn Miller[1] | Ahmad Karkoutli[1,3] | Anna Rusnak[1,4] | Susan E. Lott[5] | Joel Atallah[1]

[1]Department of Biological Sciences, University of New Orleans, New Orleans, Louisiana, USA

[2]Hayward Genetics Center, Tulane University School of Medicine, New Orleans, Louisiana, USA

[3]LSUHSC School of Medicine, New Orleans, Louisiana, USA

[4]Center for Biomedical Engineering, Brown University, Box A-2, Arnold Lab, Providence, Rhode Island, USA

[5]Department of Evolution & Ecology, University of California–Davis, Davis, California, USA

Correspondence
Joel Atallah, Department of Biological Sciences, University of New Orleans, New Orleans, LA 70148, USA.
Email: jatallah@uno.edu

Funding information
Louisiana Board of Regents, Grant/Award Number: LEQSF(2017-20)-RD-A-26

Abstract

A massive adaptive radiation on the Hawaiian archipelago has produced approximately one-quarter of the fly species in the family Drosophilidae. The Hawaiian *Drosophila* clade has long been recognized as a model system for the study of both the ecology of island endemics and the evolution of developmental mechanisms, but relatively few genomic and transcriptomic datasets are available for this group. We present here a differential expression analysis of the transcriptional profiles of two highly conserved embryonic stages in the Hawaiian picture-wing fly *Drosophila grimshawi*. When we compared our results to previously published datasets across the family Drosophilidae, we identified cases of both gains and losses of gene representation in *D. grimshawi*, including an apparent delay in Hox gene activation. We also found a high expression of unannotated genes. Most transcripts of unannotated genes with open reading frames do not have identified homologs in non-Hawaiian *Drosophila* species, although the vast majority have sequence matches in genomes of other Hawaiian picture-wing flies. Some of these unannotated genes may have arisen from noncoding sequence in the ancestor of Hawaiian flies or during the evolution of the clade. Our results suggest that both the modified use of ancestral genes and the evolution of new ones may occur in rapid radiations.

KEYWORDS
de novo genes, embryo, Hawaiian *Drosophila*, Hox genes, maternal-to-zygotic-transition, novel genes, transcriptomics

1 | INTRODUCTION

It is estimated that 1000 of the world's approximately 4000 Drosophilid fly species are endemic to the Hawaiian archipelago (O'Grady & DeSalle, 2018). Comprising the largest, and arguably the most diverse, radiation within the family, Hawaiian *Drosophila* flies (Figure 1) exhibit extensive morphological variation, and species with extreme adaptations are observed throughout the group (Craddock et al., 2018; Sarikaya et al., 2019). For example, species from the iconic Hawaiian picture-wing clade deviate dramatically in their body size (Magnacca & Price, 2015) with flies ranging from 20 to 50 times larger than *Drosophila melanogaster*. Additionally, eponymous with the group's name, Hawaiian picture-wing flies have spectacular wing pigmentation that is highly varied throughout the clade. Other clades

Disclosures: This paper is not being considered for publication elsewhere. This study formed part of Madeline Chenevert's M.S. thesis.

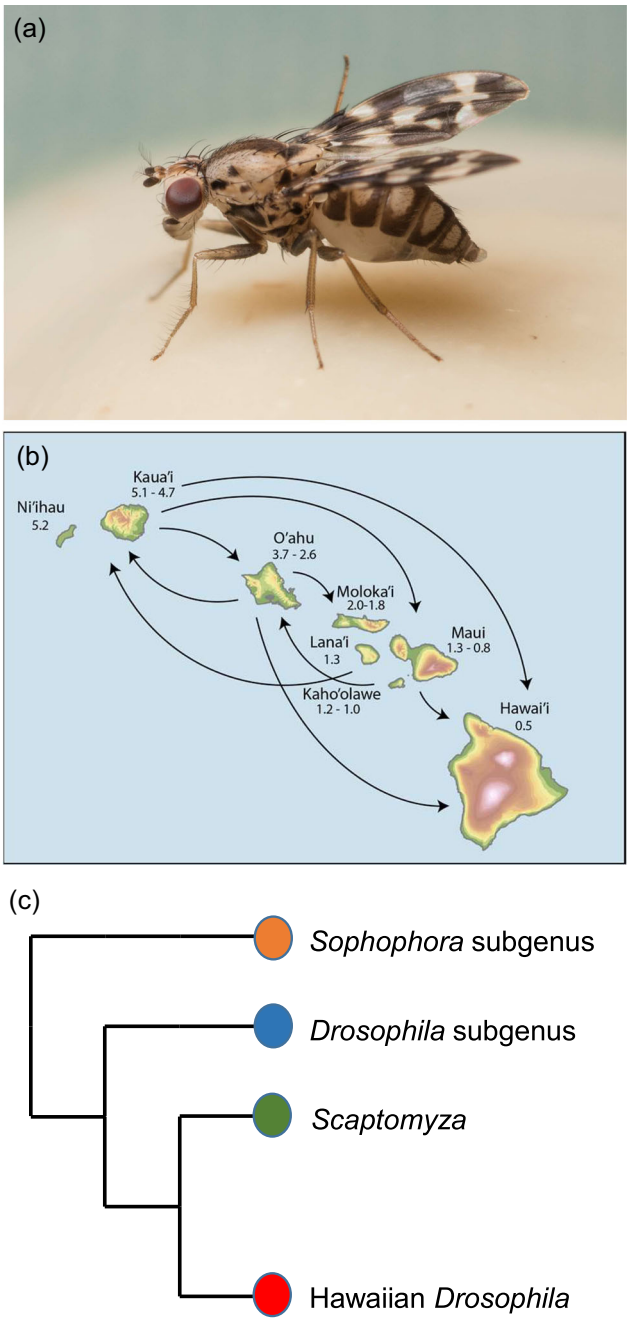


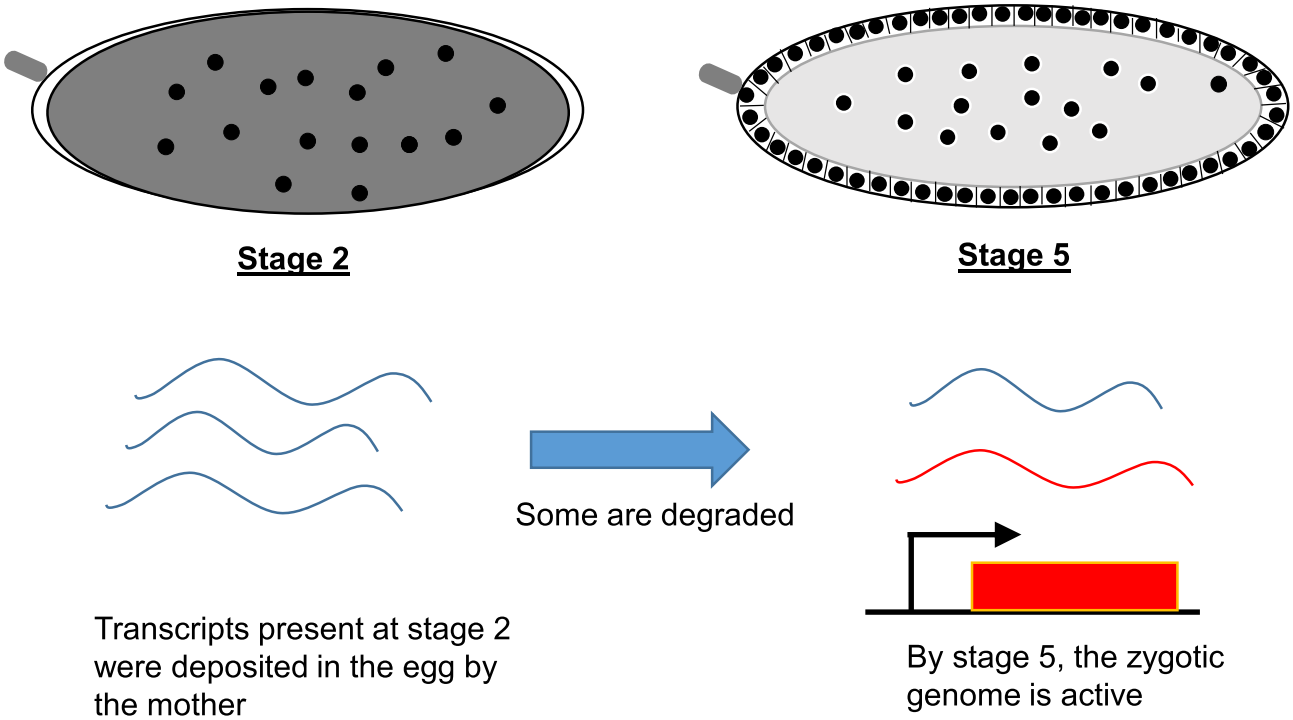
FIGURE 1 Hawaiian *Drosophila*. (a) *Drosophila grimshawi*. (b) A map of the Hawaiian islands, with arrows showing inferred colonization events in the Hawaiian picture-wing clade. The age of each island (in millions of years) is shown. Kaua'i is the oldest of the larger islands and Hawai'i the youngest. Most colonization events occurred from older (northern) to younger (southern) islands, although there were also cases of reverse colonization. (c) The Hawaiian *Drosophila* lineage is a sister-group to *Scaptomyza* (Katoh et al., 2017). The ancestor of both clades may have been Hawaiian (Lapoint et al., 2013; O'Grady & DeSalle, 2008), but *Scaptomyza* dispersed globally, while the Hawaiian *Drosophila* remained confined to the archipelago

are named after characteristic modifications to the tarsi (Lapoint et al., 2009, 2014) or mouth parts (Magnacca & Grady, 2009), both specific to male flies. In addition to changes to the adult body plan, the eggs display a vast range of variation in size, structure, and filament number (Kambysellis & Heed, 1971; Kambysellis et al., 1980). The morphological diversity exhibited by Hawaiian Drosophilidae raises the question of whether there is a similar variation in gene expression, and whether the evolution of novel genes (Van Oss & Carvunis, 2019), or the re-use of ancestral ones (Carroll, 2008), underlies the diversification of the clade.

Transcriptional profiling of Hawaiian flies, however, has been limited. While recent studies have carried out functional genomics in adult stages of picture-wing flies (Eldon et al., 2019; Kang et al., 2016), gene expression levels in the embryo have not previously been analyzed. Hawaiian flies are notoriously difficult to culture (Montgomery, 1975), and the life cycle is extended relative to common laboratory species such as *Drosophila melanogaster*, complicating both the collection of samples and the interpretation of results from timed specimens.

Early embryogenesis, which constitutes the opening act in an organism's life (Tadros & Lipshitz, 2009), is the basis for all subsequent development, and perhaps the most critical period to first consider when analyzing a species that is difficult to culture with a dearth of transcriptomic data. Across Drosophilidae, early embryos pass through comparable developmental stages (Kuntz & Eisen, 2014). We have previously shown (Atallah & Lott, 2018) that gene transcript levels are highly concordant across *Drosophila* species at two early embryonic stages, Stage 2 (St2) and late Stage 5 (St5) (Figure 2). The first of these stages precedes the maternal-zygotic transition, while at Stage 5 zygotic expression is underway and many maternal transcripts have been degraded (Figure 2). While most genes are represented at both stages, those that are St5-only (zygotically transcribed with no maternal contribution) are strongly conserved across large phylogenetic distances (Atallah & Lott, 2018).

Here, we present results of a single-embryo RNA-Seq analysis of these critical stages in the Hawaiian picture-wing fly *D. grimshawi*. We compare the findings to our previously published data on other *Drosophila* flies (Atallah & Lott, 2018). We find that *D. grimshawi* shows extensive loss of gene transcript representation after zygotic genome activation (Stage 5) relative to non-Hawaiian species. Notably, in sharp contrast to outgroup clades, we find no evidence that the Hox genes, the downstream component of the antero-posterior segmentation cascade, are activated by St5. We further find instances of early embryonic mRNA representation of gene orthologs in this picture-wing fly which are not seen in other species until later stages in development. Finally, we conduct an analysis of unannotated genes in *D. grimshawi* that are represented in the early embryo and provide evidence that many of them are taxonomically restricted. Some of these genes appear to have been generated de novo from noncoding sequence, either in the Hawaiian *Drosophila* proper or in the ancestor of Hawaiian *Drosophila* and its sister-clade, *Scaptomyza* (Figure 1), which may have originated in Hawaii before diversifying globally (Lapoint et al., 2013; O'Grady & DeSalle, 2008).



Class I	Present at stage 2	Absent at stage 5	Maternally supplied transcripts that are entirely degraded
Class II	High at stage 2	Low at stage 5	Maternally supplied transcripts that are partially degraded
Class III	Low at stage 2	High at stage 5	Evidence of both maternal provision and zygotic transcription
Class IV	Absent at stage 2	Present at stage 5	No maternal provision, zygotic transcription

FIGURE 2 The maternal-to-zygotic transition (MZT). The zygotic genome is silent in the early embryo, which relies on transcripts that were deposited by the mother and translated after egg activation to jump-start development. Many of these transcripts are degraded. By stage 5, the embryonic transcriptome consists of both maternally deposited mRNAs that have not been degraded and zygotically transcribed mRNAs

2 | RESULTS

2.1 | Gene representation in the early *D. grimshawi* embryo

We found a total of 18,617 transcripts represented in the early embryo above our established threshold (see Materials and methods). These transcripts were derived from 8233 genes. Genes with one-to-one orthologs in *D. melanogaster* or other species are shown in Table S1. Using insights obtained from previous analyses, we identified four classes of genes (Figure 2), discussed below. We used the R package GOPlot (Walter et al., 2015) to graphically represent the results of our GO analysis as concentric circular plots (Figure 3 and Figure S1).

2.2 | Class 1: Stage 2-only

Class 1 genes are present only during Stage 2, indicating that they are maternally supplied transcripts which are completely degraded by Stage 5. We identified 241 Class 1 genes, and an analysis using DAVID did not show significant enrichment for any GO terms when compared to all embryonically-represented genes (Table S2). In other words, genes with transcripts that are completely degraded in early embryogenesis appear to be randomly distributed across functional categories and cellular components (relative to our background list). We have previously shown that the St2-only state is rarely conserved across species (Atallah & Lott, 2018) for a given gene. It is possible that maternal deposition, where transcripts are added to the egg by the nurse cells, sometimes introduces developmental "noise" (the unnecessary deposition of RNA transcripts), and that the embryo compensates by degrading those transcripts.

2.3 | Class 2: Stage 2 higher, Stage 5 lower

Class 2 genes are present at both stages, with lower expression at Stage 5 and have maternally supplied transcripts that are partially degraded by Stage 5. (It is impossible to rule out that some of these genes may be zygotically transcribed, with the transcription compensating for at least some of the degradation.) We found 728 Class 2 genes that were significantly differentially expressed. We found the identified *D. melanogaster* orthologs to be enriched in the GO terms "mitochondrion," "mitochondrial matrix," and "metabolic pathways," although there was no significant enrichment for any specific pathway (Table S2). Furthermore, there was no significant enrichment of any GO terms in the Biological Process (BP), Cellular Component (CC), or Molecular Function (MF) categories. As with Class 1, genes with partially degraded transcripts are not limited to a specific subset of categories.

2.4 | Class 3: Stage 2 lower, Stage 5 higher

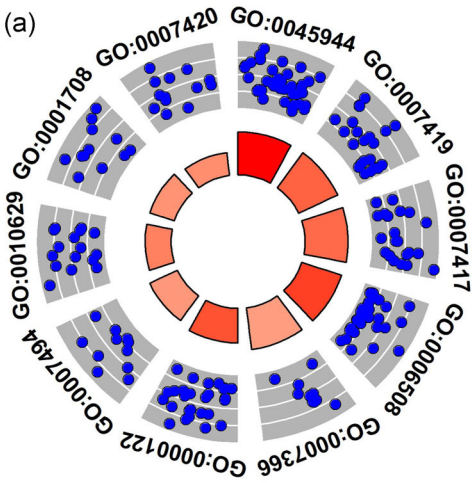
Class 3 genes include genes represented at both stages that exhibit increased expression levels at Stage 5 and are believed to be both maternally supplied and zygotically transcribed. We found 1364 Class 3 genes that were significantly differentially expressed between Stages 2 and 5. Our DAVID analysis indicated Class 3 genes are enriched in products that play roles in protein and mRNA binding (Figure S1). Cellular component enrichment was specific to the plasma membrane and cytoplasm. Class 3 genes were also enriched for components of the Hippo and Notch signaling pathways, RNA transport functions, and ubiquitin-mediated proteolysis. Many of these genes encode protein kinases. Although genes in the Wnt signaling pathway were not significantly enriched, enrichment was seen for the InterPro (Hunter et al., 2009) term "armadillo-type fold" (Table S2). Tandem armadillo repeats (Peifer et al., 1994) are found in many proteins involved in Wnt signaling (including β-catenin; McCrea et al., 1991), which plays a key role during cellularization (Stage 5). The provision of these components through two mechanisms (maternal deposition and zygotic transcription) could be a form of redundancy to increase developmental robustness (Mestek Boukhibar & Barkoulas, 2016).

2.5 | Class 4: Stage 5-only

We identified 880 Class 4 genes which were present only at St5 (zygotic-only genes). They represent a range of genes important in early development (Figure 3 and Figure S2), and include genes involved in patterning, development of the central nervous system (including the brain), and several other ontogenetic processes (Figure 3a). Since St5 is the period when cellularization occurs, it is not surprising that more genes map to the GO term "integral component of membrane" than to any other cellular component (Figure 3b and Table S2). They include genes encoding proteins localized to the plasma cell membrane and organelle membranes. Many of the gene products are also part of the extracellular matrix.

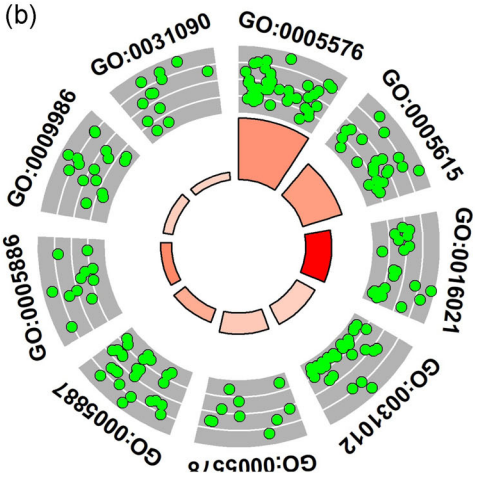
Transcription factors, which activate patterning cascades and jump-start development, are by far the most overrepresented molecular function class of St5-only genes, both in terms of the number of genes mapping to this category and the negative logarithm of the FDR (Figure 3c). Other zygotic-only genes have serine-type endopeptidase activity (GO term 0004252 in Figure 3c); these include *masquerade*, which has a known role axonal development, and *modular serine protease* (modSP), which is involved in the immune system (Table S2). St5-only genes also include heme-binding enzymes with oxidoreductive activity, including an array of cytochrome P450 enzymes (e.g., *phantom*, *disembodied*, *spookier*, and many others).

(a)



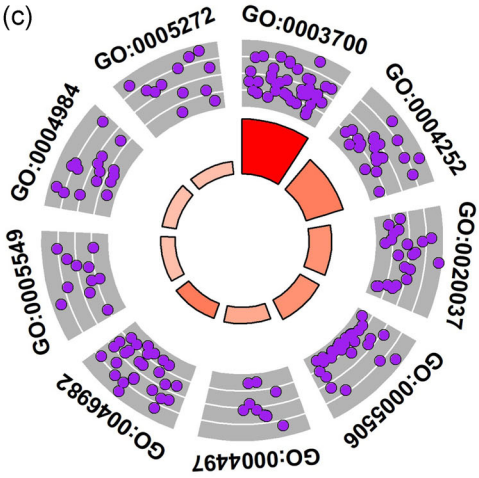
ID	Description
GO:0045944	positive regulation of transcription from RNA polymerase II promoter
GO:0007419	ventral cord development
GO:0007417	central nervous system development
GO:0006508	proteolysis
GO:0007366	periodic partitioning by pair rule gene
GO:0000122	negative regulation of transcription from RNA polymerase II promoter
GO:0007494	midgut development
GO:0010629	negative regulation of gene expression
GO:0001708	cell fate specification
GO:0007420	brain development

(b)



ID	Description
GO:0005576	extracellular region
GO:0005615	extracellular space
GO:0016021	integral component of membrane
GO:0031012	extracellular matrix
GO:0005578	proteinaceous extracellular matrix
GO:0005887	integral component of plasma membrane
GO:0005886	plasma membrane
GO:0009986	cell surface
GO:0031090	organelle membrane

(c)



ID	Description
GO:0003700	transcription factor activity, sequence-specific DNA binding
GO:0004252	serine-type endopeptidase activity
GO:0020037	heme binding
GO:0005506	iron ion binding
GO:0004497	monooxygenase activity
GO:0046982	protein heterodimerization activity
GO:0005549	odorant binding
GO:0004984	olfactory receptor activity
GO:0005272	sodium channel activity

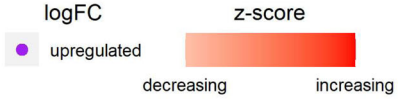


FIGURE 3 GO analysis of St5-only genes (Class 4). The height of the bars in the inner circle represents the Benjamini-corrected enrichment *p* value of the associated term, and the color represents the "z-score," in this case the square root of the number of genes mapping to the term. The outer circle shows the log of the fold enrichment (St5/St2) of genes mapping to each term. GOplot (Walter et al., 2015) was used to generate the plot. (a) Biological process GO terms. (b) Cellular component GO terms. (c) Molecular function GO terms

2.5.1 | Some gain, and widespread loss, of gene expression in the *D. grimshawi* Stage 5 embryo

The results of the differential expression analysis above were largely concordant with our previous findings for non-Hawaiian species (Atallah & Lott, 2018). We were curious, however, about any differences that might exist in gene expression or expression loss in *D. grimshawi*, the first Hawaiian species with embryonic transcriptomes. When considering genes with *D. melanogaster* orthologs, we found more cases (101) of unique losses of Stage 5 representation in *D. grimshawi* than in any other species (Figure 4).

Among the most interesting losses of St5 gene representation are the homeotic selector (Hox) genes, the downstream component of the anterior–posterior (AP) patterning pathway. As can be seen in Figure 5, multiple Hox genes (in particular, *Deformed*, *Antennapedia*, and *Abdominal-A*) are expressed by Stage 5 in all species except *D. grimshawi*. We confirmed that components of the AP cascade upstream of the Hox genes, including the segment polarity, pair-rule and gap genes, are all expressed by Stage 5 in *D. grimshawi* (Table S3).

We found a total of 79 genes with unique gains of representation (Figure 4). We took a closer look at the 35 genes with transcript abundance three times our threshold (FPKM > 3). Of these genes, eight have identified *D. melanogaster* orthologs (Table S4). They include *Lim3*, a zinc-finger transcription factor with a homeobox domain with a role in motor–neuron development. Another gene, *multiple wing hairs* (*mwh*), a downstream component of the planar cell polarity pathway, is represented at both St2 and St5 in *D. grimshawi*, but not in any other species we have examined.

2.5.2 | Strong expression of unannotated genes in *D. grimshawi*

RNA-Seq bioinformatics methods that include the discovery of new isoforms, such as the Tuxedo suite, often uncover previously unannotated genes. Although much faster alignment-free methods have become more popular in recent years, we chose the Tuxedo suite both to allow the identification of novel isoforms and genes and for straightforward comparison with our previous analyses of equivalent stages in non-Hawaiian species. In *D. grimshawi*, we find more unannotated transcripts with high Stage 5 embryonic mRNA levels (FPKM > 3) than in any other species we have previously studied (Figure 6a). This finding could be partially due to the relatively poor annotation of the genome of this Hawaiian species (although the *D. grimshawi* genome annotation has improved markedly in recent years; Yang et al., 2018). Indeed, it has long been known that poor genome annotation can lead to spurious claims of novel gene evolution (Schmid & Aquadro, 2001). We therefore decided to carry out additional analyses to determine whether some of these transcripts could belong to putative orphan genes.

Novel *Drosophila* genes, particularly those arising from noncoding DNA, are typically short (Hahn et al., 2007). A previous study (Heames et al., 2020) found unannotated genes in the Drosophilidae family to code for a median peptide length of only 81 amino acids. Using TransDecoder (Haas et al., 2013), we identified the longest complete open reading frame (ORF) (if any) in each of the 2365 unannotated transcripts, and found 969 transcripts with an ORF of at least 50 amino acids. We used BLASTP (Altschul et al., 1990) to determine whether these putative peptides had homologs in the annotated genomes or embryonic transcriptomes of other *Drosophila*

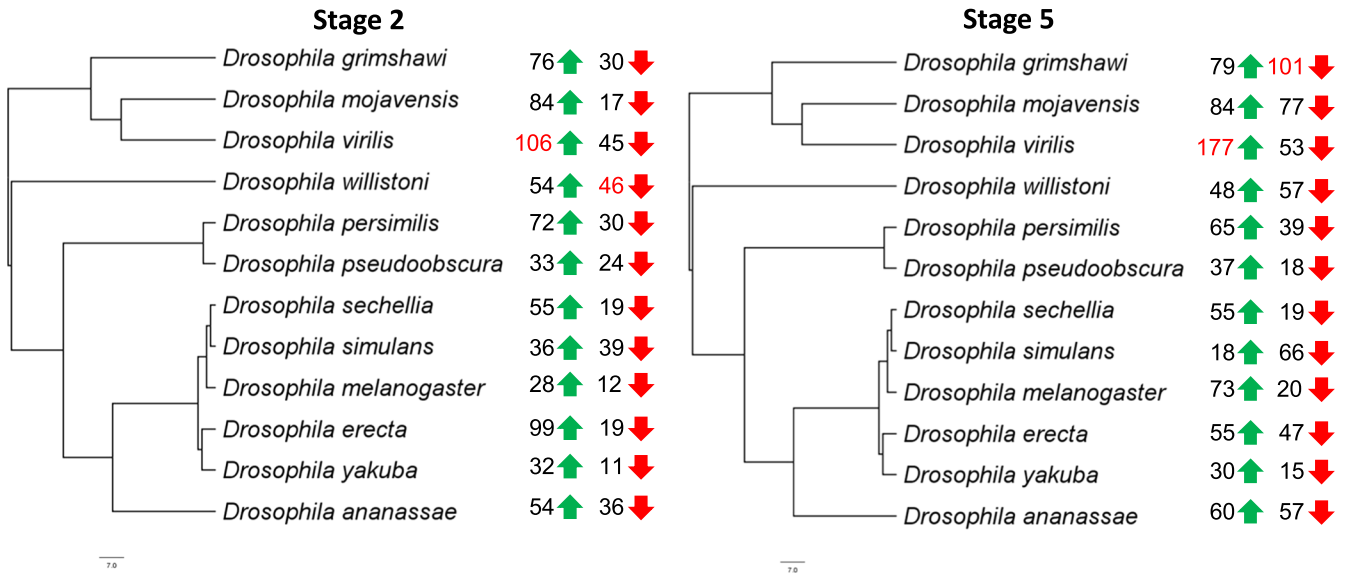


FIGURE 4 Numbers of unique gains (green arrows) and losses (red arrows) of St2 and St5 gene representation in each species. The analysis was conducted with 15 *Drosophila* species, but only the 12 originally sequenced species are shown (*D. mauritiana*, *D. santomea*, and *D. miranda* are not shown). *D. virilis* shows the largest number of gains at both St2 and St5, while *D. willistoni* shows the largest number of losses at Stage 2, and *D. grimshawi* shows the largest number of losses at Stage 5

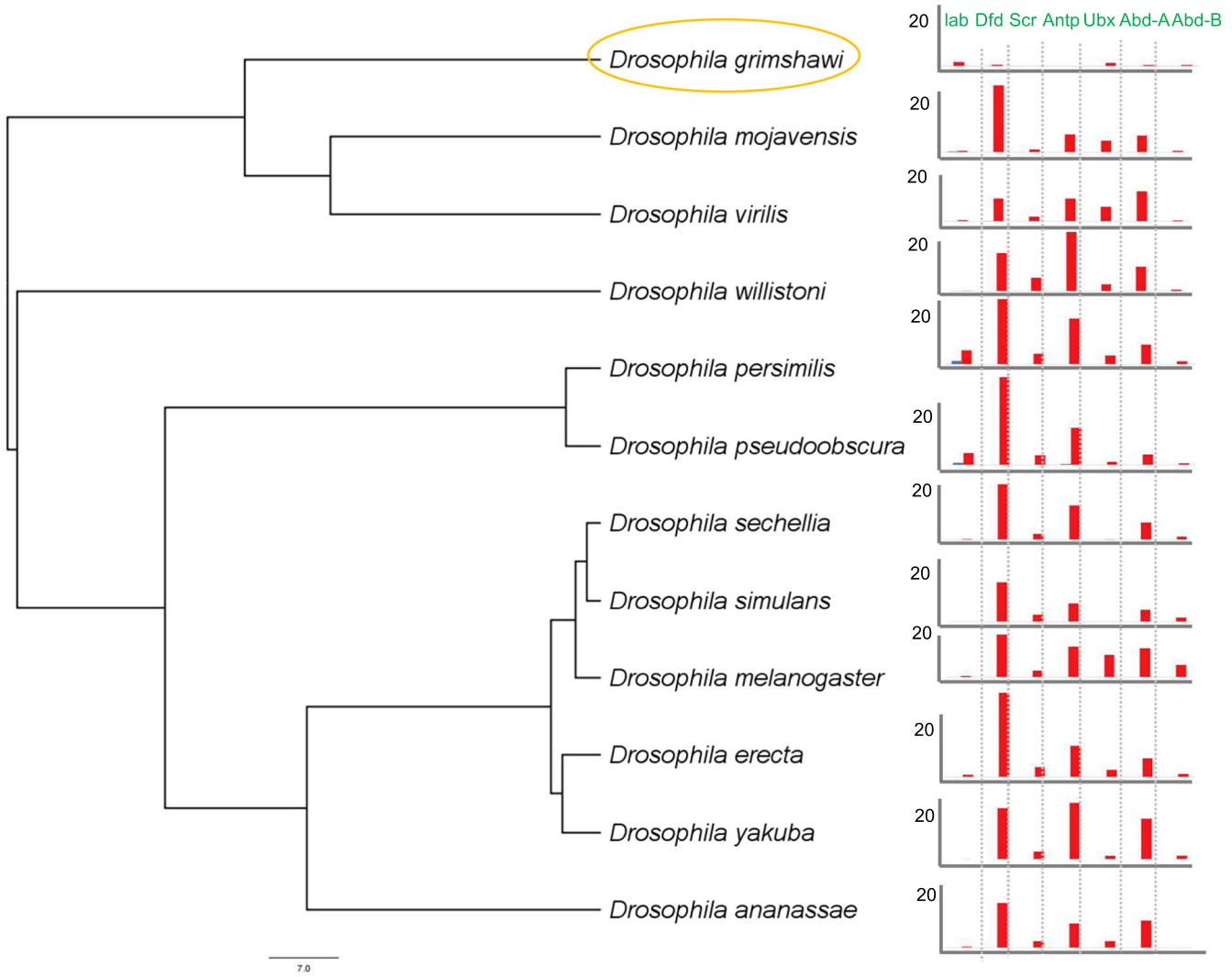


FIGURE 5 Loss of St5 representation of homeotic selector gene transcripts in *D. grimshawi*. St2 FPFM levels (close to 0 across all species) are shown in blue and St5 levels in red

species. A total of 854 of the 969 ORFs had no identifiable orthologs. Of these, we chose to focus on the 301 transcripts with an FPKM above our threshold of 1 at Stage 2, Stage 5, or both.

2.5.3 | A subset of unannotated genes may have been generated de novo from noncoding sequence

Taxonomically restricted genes may have either originated from ancestral genes (e.g., through divergence beyond recognition from an ortholog, or through gene fission or fusion) or emerged de novo from noncoding sequence (Van Oss & Carvunis, 2019). Examples of de novo gene evolution, which at one time was dismissed as highly improbable, have fascinated researchers in recent years (Klasberg et al., 2018; Neme & Tautz, 2016; Zhao et al., 2014). Researchers frequently use TBLASTN to identify putative intergenic or intronic regions in other species that orphan genes might have arisen from (Lu et al., 2017; Sun et al., 2015). We adopted a similar approach, aided by the recent publication (B.Y. Kim et al., 2021) of new *Drosophila* genome assemblies (as yet unannotated), generated through Oxford Nanopore long-read sequencing, that included two additional Hawaiian picture-wing species (*Drosophila murphyi* and *Drosophila sproati*) along with four *Scaptomyza* species (*Scaptomyza graminum*, *Scaptomyza hsui*, *Scaptomya Montana*, and *Scaptomyza pallida*).

Of the 114 transcripts with TBLASTN hits in one or more of the non-Hawaiian annotated genomes, most of the top hits (95, or 85%) are in intergenic or intronic regions in all species, suggesting that they may have arisen de novo from noncoding sequence (Figure 7a). Interestingly (Figure 6b), 97% of the larger set of 301 transcripts with FPKM > 1 (291) had TBLASTN hits in either *D. murphyi* or *D. sproati* (the other two picture-wing Hawaiian *Drosophila* species). Only 47% had hits in the genomes of one of the four species in the *Scaptomyza* lineage, and 35% in the *Drosophila* subgenus (*D. virilis* or *D. mojavensis*).

Of the transcripts with noncoding TBLASTN hits in annotated non-Hawaiian species genomes, 67 out of 95 (71%) matched regions

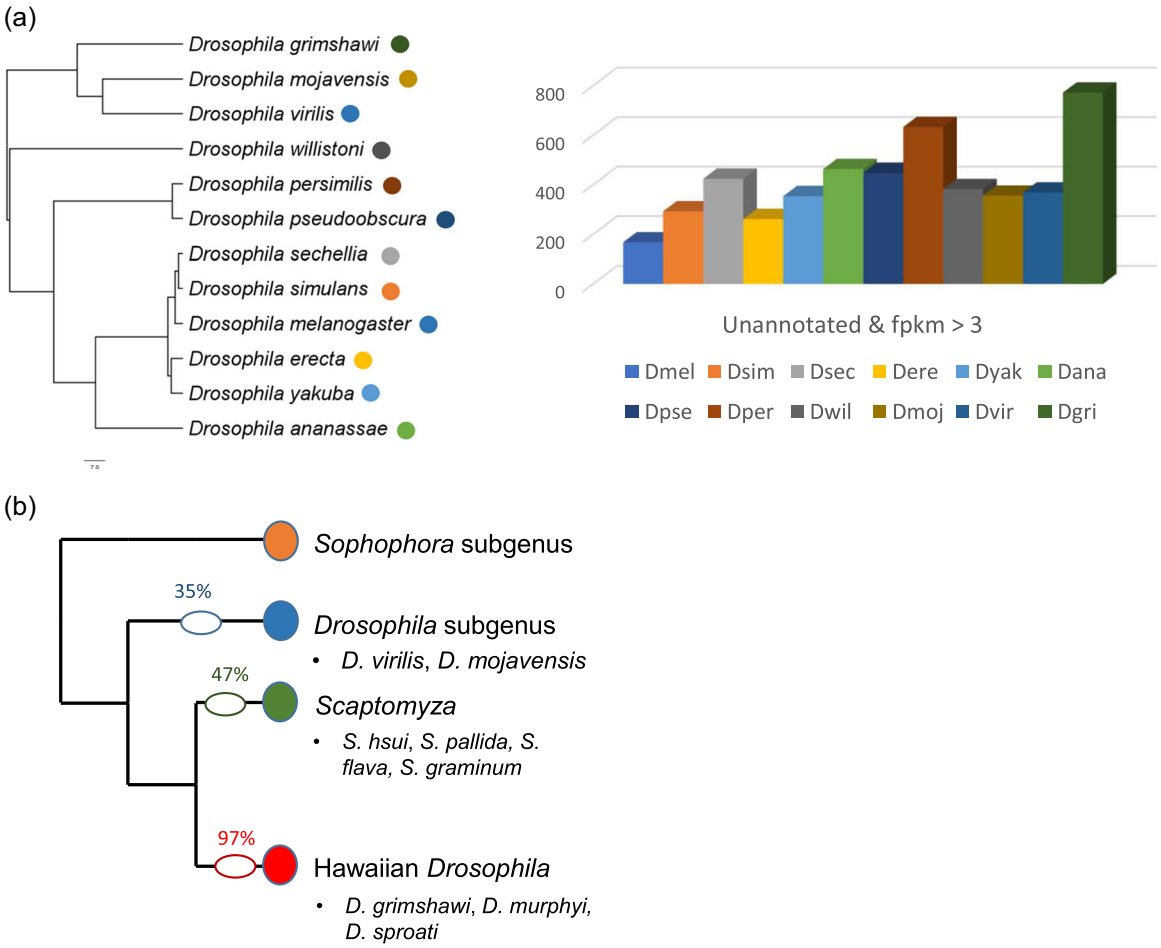


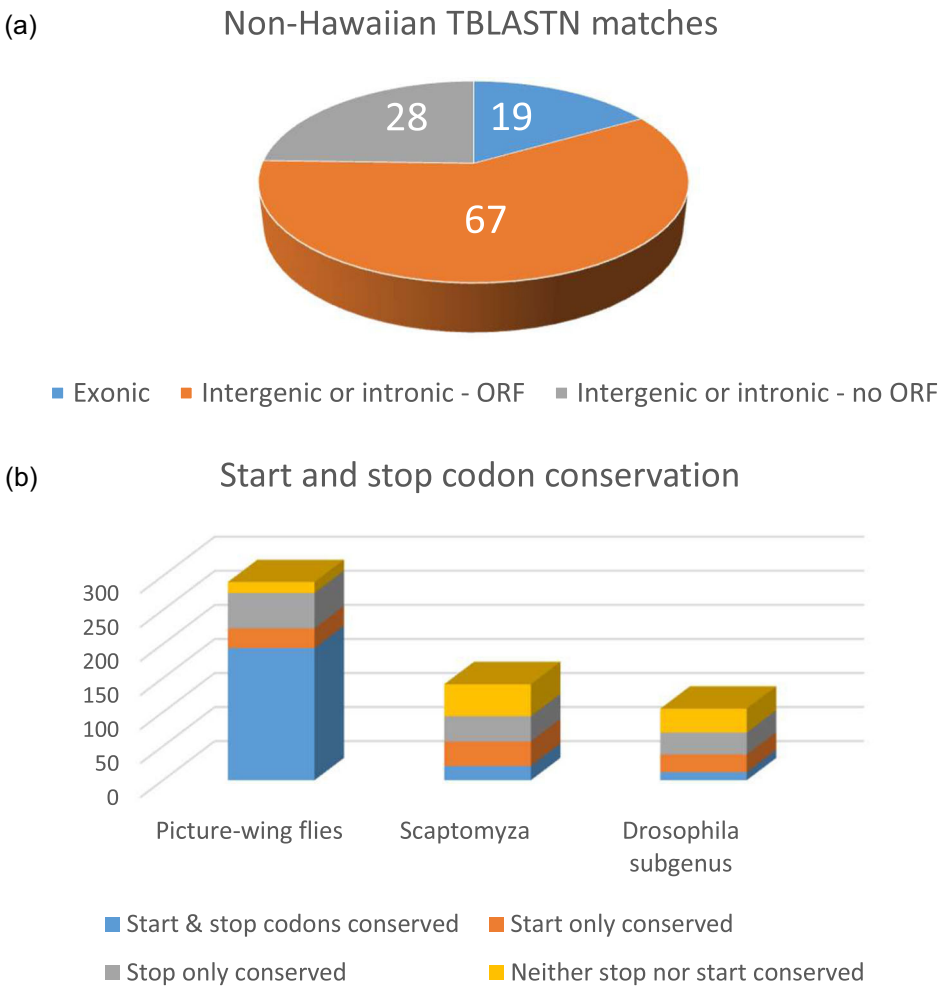
FIGURE 6 Unannotated genes in *D. grimshawi* show high expression levels, with a subset showing evidence of de novo evolution. (a) More *D. grimshawi* unannotated genes are represented at high levels at Stage 5 (FPKM > 3) than for any of the other species we have examined. (b) Almost all identified ORFs in our candidate set of 301 *D. grimshawi* unannotated transcripts have TBLASTN hits in genomes of at least one other picture-wing Hawaiian *Drosophila* species, while slightly under half have hits in a *Scaptomyza* genome, and about a third have hits in *D. virilis* or *D. mojavensis*

that contain an ORF in at least one of the fly genomes (Figure 7a). Alignments with CLUSTALW showed that in most cases both the start and stop codons were conserved in at least one of the other picture-wing *Drosophila* species (Figure 7b). The same was true far less frequently in *Scaptomyza* and the *Drosophila* subgenus species. We also found cases, however, where a start or stop codon may have been generated from noncoding sequence through a single-nucleotide substitution (Figure 7c). Our results suggest that a subset of unannotated genes with embryonic mRNA representation may have been generated de novo from noncoding sequence during early Hawaiian fruit fly diversification.

3 | DISCUSSION

Island endemics are excellent systems for exploring the evolution of novelty. Most famously, the flora and fauna of the Galapagos islands were critical to Charles Darwin's (1845) development of the theory of natural selection. More recently, the evolution of Anolis lizards on the Greater and Lesser Antilles has been analyzed rigorously (Corbett-Detig et al., 2020; Mahler et al., 2010, 2016). Islands such as the Hawaiian archipelago, which arose over the past 25 million years through volcanic eruptions, presented pristine, untouched environments for early colonists with untapped ecological opportunities (Whittaker et al., 2017). Evolution on these types of islands proceeds far more rapidly than on the mainland (Losos & Ricklefs, 2009), leading to diversity which often exceeds those of continental species.

As close relatives of the model organism *D. melanogaster*, the Hawaiian *Drosophila* clade is a particularly valuable model for studying the evolutionary genetics of island radiations. Research on this clade has slowed, however, in recent decades (O'Grady & DeSalle, 2018). We present here the first transcriptomic comparison in a Hawaiian fly of two stages in embryogenesis that we know to be highly conserved across Drosophilidae (Kuntz & Eisen, 2014). We find that while *D. grimshawi*'s early embryonic transcriptome at these stages is similar in many ways to other non-Hawaiian species, there are distinct differences, with numerous losses of orthologous gene representation. Most notably, while the Hox genes are



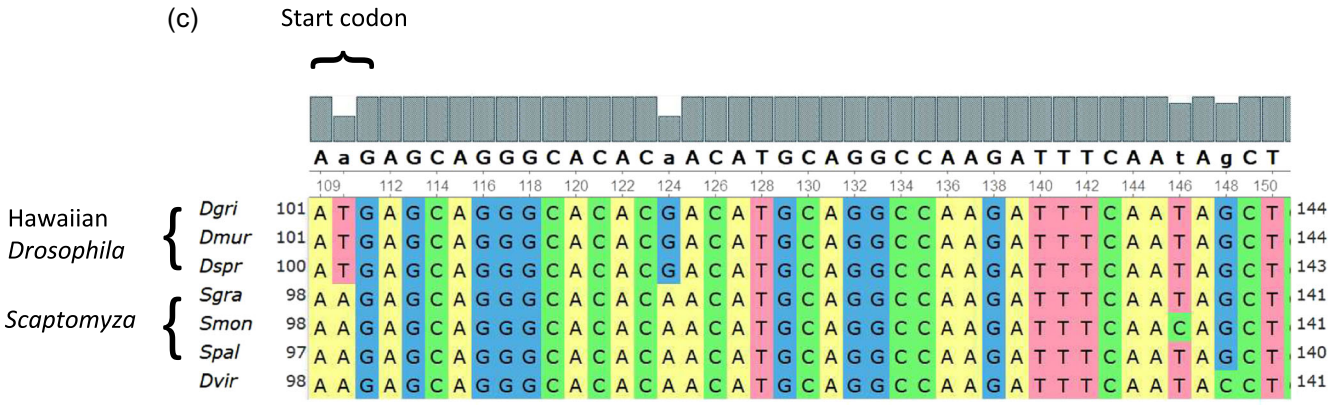


FIGURE 7 Putative de novo genes may have arisen from pre-existing ORFs or been generated from noncoding sequence. (a) Out of 114 transcripts with TBLASTN hits in non-Hawaiian species (from our unannotated candidate set of 301), the overwhelming majority are located in intergenic or intronic regions, with a pre-existing ORF identified in the majority of cases. (b) The start and stop codons of identified ORFs in *D. grimshawi* unannotated transcripts are highly conserved in other picture-wing species. Conservation is lower in *Scaptomyza* and the *Drosophila* subgenus. (c) An example of a new ORF in a picture-wing Hawaiian *Drosophila* species. In the unannotated *D. grimshawi* transcript TCONS_41457, a start codon may have evolved from a single nucleotide change in the ancestor of the picture-wing Hawaiian *Drosophila* species. The alignment was generated using CLUSTALW

represented by Stage 5 in every other species we have examined, we find no evidence of this in *D. grimshawi*. (Many of these genes, including the Hox genes, are probably expressed at later embryonic stages.) Future studies will be necessary to determine when homeotic selector genes are activated in this species and whether the apparent delay in Hox activation is also seen in other Hawaiian flies. The *cis* and *trans* regulation of these genes in Hawaiian species should also be investigated.

Genes with zygotic-only (i.e., Stage 5) expression in *D. grimshawi*, but not in other species (Table S4), are excellent candidates for functional knock-out experiments. We are currently developing protocols for targeted mutagenesis in *D. grimshawi* using the CRISPR-Cas9 system (Doudna & Charpentier, 2014; Jinek et al., 2012). The distinct advantage of examining the early embryo is that the effects of mutations can be assessed at the beginning of development. In the long-term, we would like to examine the activity and functions of genes active in larval and pupal discs of structures such as wings and legs, which have diverged markedly in Hawaiian species (Edwards et al., 2007; Stark & O'Grady, 2010). These goals will be more challenging to achieve because the role of a gene early in development may mask its function at a later stage (e.g., if the gene is embryonically lethal, it is more difficult to determine its later role in imaginal discs). Such research may require precise editing of specific enhancers.

Gains and losses of expression of genes with orthologs in other species, including the model fly *D. melanogaster*, could be examples of co-option of a conserved genetic toolkit (Carroll, 2008), although the functional experiments described above will be necessary for investigating this possibility. While the reuse of existing genes is thought to play an important role in developmental evolution (Stern, 2011), attention has turned more recently to the role of taxonomically restricted genes (Johnson, 2018; McLysaght & Guerzoni, 2015; Van Oss & Carvunis, 2019; Xia et al., 2021). Much of the work investigating de novo gene evolution has focused on transcripts expressed in the testis (Lange et al., 2021; Witt et al., 2019; Zhao et al., 2014), where evolutionary turnover occurs rapidly (Jagadeeshan & Singh, 2005; Mohammed et al., 2014), with fewer studies looking at genes expressed in relatively slowly evolving systems such as embryogenesis. In this study, we were interested in the fact that more unannotated genes were expressed in *D. grimshawi* at St5 than in any of the 14 other species we had previously examined. While we had speculated (Atallah & Lott, 2018) that unannotated genes might be taxonomically restricted, we had not examined this possibility further, and it is possible that many unannotated genes are simply a result of poor genome annotation (Schmid & Aquadro, 2001). In this study, we identified a total of 95 embryonic *D. grimshawi* transcripts with open reading frames that had homologous sequence matches in intergenic or intronic regions in a non-Hawaiian species (Figure 7a), suggesting that they may have originated de novo, either in the Hawaiian *Drosophila* lineage or the ancestor of Hawaiian *Drosophila* and *Scaptomyza*. Further verification will be possible as newly available (B.Y. Kim et al., 2021) Hawaiian fly genomes are annotated. The annotation of these genomes will allow us to conduct synteny analyses using whole-genome alignments to carefully analyze the positions of putative de novo genes, and the noncoding sequences they evolved from, relative to homologous annotated loci.

Our findings should be interpreted in light of other evidence showing the rapid evolution of novel genes in *D. grimshawi*. For example, a study that examined a different mechanism of novel gene evolution, gene family expansion (which is known to account for a large fraction of newly evolved essential genes in *Drosophila*; Chen et al., 2010), found more species-specific gene duplications in *D. grimshawi* than in 11 other sequenced *Drosophila* species (Zhong et al., 2013). It is likely that both family expansions and de novo gene evolution play important roles in adaptive radiations.

Two models exist to explain the emergence of de novo protein-coding genes (McLysaght & Guerzoni, 2015). In the "ORF-first" model, a pre-existing ORF is expressed through regulatory element evolution. In a second model, a noncoding RNA acquires an ORF (Xie et al., 2012). Analyzing embryonic transcriptomes from other Hawaiian *Drosophila* and *Scaptomyza* species, where we often found ORFs in regions that were homologous to the unannotated expressed sequences in *D. grimshawi*, will be necessary to distinguish between these models. As with co-opted genes, targeted mutagenesis of putative de novo genes will be important for determining whether they have acquired novel functions in the early embryo.

4 | MATERIALS AND METHODS

4.1 | *Drosophila grimshawi* husbandry

Upon receiving the *D. grimshawi* stock (Stock Number 15287-2541.00) from the San Diego *Drosophila* Species Stock center (currently the National *Drosophila* Species Stock Center at Cornell University, https://www.drosophilaspecies.com/), the flies were maintained in accordance with a modified version of stock center protocols. The line we used was a replacement donated by Dr. Ken Kaneshiro from the same stock that was used in sequencing the *D. grimshawi* genome for the 12 Drosophila Genomes Project (Clark et al., 2007). The flies were kept in vials of Wheeler-Clayton food (Wheeler & Clayton, 1965). To counter the high humidity of New Orleans, which could have caused the flies to stick to the food, Wheeler-Clayton food was prepared using the maximum recommended amount of agar (14.5 g/L water). The cooked potato medium, used as the bottom layer of the Wheeler-Clayton food, was prepared with 456 g of filtered water for every 100 g of powdered mix to ensure a consistency that prevented food from falling during transfers. High doses of ethanol and propionic acid (6.5 ml/L water) were also used in the top layer recipe to prevent mold growth.

Vials were papered with Kimwipes, and flies were transferred to new vials twice per week. We found that maintaining stocks at a density of 12–16 flies per vial maximized fecundity while reducing death due to overcrowding. Vials containing third instar larvae were transferred to jars to pupate. The jars contained equal volumes of

oolite and aragonite sand, approximately 2 cm deep, and were covered with two layers of heavy-duty paper towels. The sand was moistened with water from a spray bottle once or twice per week, with the aim of keeping the sand and towels damp while avoiding puddles. Excess water was absorbed with cellulose acetate plugs ("flugs," Genesee Scientific), which were then disposed of.

After the first eclosions, the vials with larvae were moved to a new sand jar, and a small petri dish containing the top layer of the Wheeler-Clayton food was added to the jar to feed new flies as they emerged. Initially, we rapidly transferred the recently eclosed flies to vials to expand our population, but we observed that they often became stuck in the paper or condensation due to the high ambient humidity in New Orleans. To avoid losing adults before they reached sexual maturity, the flies were maintained in the jar for 2 weeks, and then moved to large egg collection cages. The cages were placed atop Petri dishes containing a base of apple-agar media, topped with a 2-ml drop of melted Wheeler-Clayton food, providing nutrient-rich media to encourage egg-laying. These Wheeler-Clayton pellets, containing the eggs, were transferred from the apple media with a spatula, washed with water or a 50% solution of apple cider vinegar to inhibit mold, cut into multiple pieces with a spatula and mixed into vials of Wheeler-Clayton food. Most vials prepared in this manner produced dozens of larvae and little to no mold.

4.2 | Egg collection

Single embryo collection and staging were carried out as described previously (Atallah & Lott, 2018), with a few modifications. Briefly, flies were maintained at room temperature (approximately 22°C) in an egg collection cage lined with a Petri dish of apple-agar media, supplemented with 2 ml of live yeast paste in the center of the plate. Flies were maintained on live yeast for an hour, after which plates were removed and replaced with new ones. Since *Drosophila* flies frequently retain fertilized eggs in their reproductive tracts (Horváth & Kalinka, 2018), only Stage 5 embryos were collected from these plates. After another hour, plates were removed and Stage 2 embryos were collected.

Eggs were dechorionated with a 50% bleach solution (Rothwell & Sullivan, 2007). Unlike smaller eggs of non-picture wing species, which could be dechorionated after less than 2 min in bleach, *D. grimshawi* eggs have an endochorion that is eight times thicker than *D. melanogaster* and dechorionation could take upwards of 3 min (Margaritis et al., 1983). The process was halted when the filaments were completely dissolved and no longer visible. Dechorionated eggs were rinsed with deionized water for at least 30 s and transferred to a drop of halocarbon oil (Sigma) on a microscope slide. Using a dissection scope, individual embryos were selected based on morphological characteristics outlined by Bownes (1975). St2 embryos exhibited empty poles at the tips of the anterior and posterior ends of the egg and lacked visible cell membranes. For St5, we collected late Stage 5 embryos, during the period after cellularization but before gastrulation. St5 embryos exhibited well-defined cell membranes and formation of the pole cells. Embryos showing damage or signs of gastrulation were rejected.

4.2.1 | Embryo processing and RNA extraction

A 1.5 ml microcentrifuge tube was filled with 800 µl of Trizol reagent (Ambion) and labeled with the sample number. Each embryo was placed on the corner of a clean microscope coverslip. Using a fine paintbrush, excess halocarbon oil was removed from the embryo. Three microliters of Trizol were transferred from the labeled 800 µl aliquot and placed on top of the embryo. Then the embryo was lysed using a sterile, 30-gauge medical lancet (ReliOn Ultra-Thin) and left to dissolve in the Trizol for 5 min. After the embryo had completely dissolved, an additional 3 µl of Trizol were added to the slide, and all 6 µl were pipetted and transferred to the tube containing the remaining 794 µl, which was mixed by pipetting. To ensure we had collected all embryonic tissue, 6 µl of Trizol were again removed from the tube and pipetted onto the place of the embryo and recollected. This rinsing step was repeated twice. Labeled samples were stored at −80°C until RNA isolation.

RNA was extracted per the manufacturer's instructions using a Trizol phenol–chloroform extraction (Invitrogen) in which the method was modified to accommodate an initial volume of 800 µl of Trizol. During RNA precipitation, 10 µl of 20 µg/µl glycogen (Invitrogen™ UltraPure™ Glycogen) were used, and the samples were spun in an Eppendorf 5424R centrifuge at 21130 RCF (maximum speed) for 60 min. If a sample showed no visible pellet, it was spun for an additional 30 min. After resuspension in 20 µl UltraPure water, 15.5 µl were kept for processing, and 3 aliquots of 1.5 µl each were taken for quality analysis.

4.2.2 | RNA quality analysis

RNA integrity was assessed on an Experion bioanalyzer using a Bio-Rad Experion RNA High-Sensitivity kit or an Agilent 2100 bioanalyzer using a Bioanalyzer RNA 6000 Pico assay (Agilent). RNA concentrations were measured using a Qubit 4 fluorometer with a Qubit high-sensitivity RNA assay.

4.2.3 | cDNA libraries

To remove any DNA contamination before library construction, a Turbo DNA-Free™ kit (Thermo Fisher Scientific) was used according to the manufacturer's directions. Then cDNA libraries for each sample were constructed using the NEBNext Ultra RNA Library Prep kit for Illumina. Fifteen PCR cycles were used in the final enrichment step of library generation. cDNA quality was assessed using an Agilent 2100 Bioanalyzer with a High-Sensitivity DNA Kit. The samples showed main peaks around 300 base pairs, as expected.

4.2.4 | Sequencing

The four highest quality libraries (all from single embryos) from each stage were sent to Novogene for sequencing. Only libraries from samples that had shown minimal RNA degradation were selected. In our bioinformatic analysis only three libraries from each stage were used. One library from St2 with low overall read mapping rate (<65%) was not used while another library from St5 appeared to have been mislabeled. The files containing the reads are publicly available at NCBI's Sequence Read Archive (BioProject Accession Number PRJNA771180).

4.2.5 | Transcriptome mapping, assembly, and differential expression

Bioinformatic analysis was carried out as described previously (Atallah & Lott, 2018). Briefly, adapters were removed using Cutadapt (Martin, 2011) and reads were trimmed and filtered for quality. We used the Tuxedo suite (Trapnell et al., 2012) for transcriptomic analysis. We aligned the reads, using Tophat2 (D. Kim et al., 2013), to the GCF_000005155.2_dgri_caf1 *D. grimshawi* NCBI reference genome. Cufflinks (Trapnell et al., 2013) was used with the -N upper-quartile normalization option. Gene expression levels were determined by combining the expression of all gene isoforms. Differential expression was determined using Cuffdiff 2 (Trapnell et al., 2013), a method that accounts for count overdispersion (a common problem in RNA-Seq data) relative to what would be expected under a Poisson model. Differentially expressed genes or isoforms were those with an adjusted *p* value (referred to as a *q* value or false discovery rate [FDR]) of less than 0.05. *D. melanogaster* orthologs were assigned using the Flybase (Larkin et al., 2021) orthology table. An FPKM (fragments per kilobase of transcript per million mapped reads) threshold of 1 was used, as employed previously (Atallah & Lott, 2018).

4.2.6 | Gene ontology (GO) enrichment analysis

GO enrichment of genes with identified one-to-one *D. melanogaster* orthologs was assessed using DAVID (Huang et al., 2009a, 2009b). In cases where Cufflinks combined two annotated genes, both genes were considered differentially expressed in the GO analysis. The background list in all cases was the set of all genes represented in the *D. grimshawi* embryo, unless otherwise specified. The R package GOPlot (Walter et al., 2015) was used to generate graphical representations of gene enrichment.

4.2.7 | Unannotated gene analysis

Transcripts of genes that were identified by Cufflinks but had no annotation in the NCBI genome were searched for open reading frames using TransDecoder (Haas et al., 2013). Transcripts with complete open reading frames of at least 50 amino acids were then compared to the genomes of 11 other *Drosophila* species as well as their previously generated embryonic transcriptome assemblies (Atallah & Lott, 2018; Combs & Eisen, 2013; Lott et al., 2014; Paris et al., 2015). In cases where no ortholog was identified using BLASTP (Altschul et al., 1990), we used TBLASTN to search for matches in the other *Drosophila* genomes, then analyzed these matches for ORFs using TransDecoder. Alignments were generated using CLUSTALW (Thompson et al., 2003).

ACKNOWLEDGMENTS

We would like to thank Delmy Urbina, Jacob Michilak, and Bianca Canizares for assistance with the experiments. Logan Crees captured the *D. grimshawi* fly image for Figure 1A. Funding for this study was provided to Joel Atallah from the Louisiana Board of Regents (grant number LEQSF(2017-20)-RD-A-26).

CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

The data supporting this study are publicly available in the text and supplementary files and on NCBI's Sequence Read Archive (BioProject Accession Number PRJNA771180). Data from previous studies (Atallah & Lott, 2018; Lott et al., 2014; Paris et al., 2015) used in the analysis are archived\referenced under the NCBI GEO accession number GSE112858. Processed data and scripts are available at J.A.'s github (https://github.com/joelatallah/D_grimshawi_transcriptome_analysis) and upon request from the corresponding author.

ORCID

Susan E. Lott https://orcid.org/0000-0002-6942-2231
Joel Atallah http://orcid.org/0000-0003-2425-2395

REFERENCES

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Atallah, J., & Lott, S. E. (2018). Evolution of maternal and zygotic mRNA complements in the early Drosophila embryo. *PLoS Genetics*, *14*(12), e1007838. https://doi.org/10.1371/journal.pgen.1007838

Bownes, M. (1975). A photographic study of development in the living embryo of *Drosophila melanogaster*. *Journal of Embryology and Experimental Morphology*, *33*(3), 789–801.

Carroll, S. B. (2008). Evo-devo and an expanding evolutionary synthesis: A genetic theory of morphological evolution. *Cell*, *134*(1), 25–36. https://doi.org/10.1016/j.cell.2008.06.030

Chen, S., Zhang, Y. E., & Long, M. (2010). New genes in Drosophila quickly become essential. *Science*, *330*(6011), 1682–1685. https://doi.org/10.1126/science.1196380

Clark, A. G., Eisen, M. B., Smith, D. R., Bergman, C. M., Oliver, B., Markow, T. A., Kaufman, T. C., Kellis, M., Gelbart, W., Iyer, V. N., Pollard, D. A., Sackton, T. B., Larracuente, A. M., Singh, N. D., Abad, J. P., Abt, D. N., Adryan, B., Aguade, M., … Newfeld, S. (2007). Evolution of genes and genomes on the Drosophila phylogeny. *Nature*, *450*(7167), 203–218. https://doi.org/10.1038/nature06341

Combs, P. A., & Eisen, M. B. (2013). Sequencing mRNA from cryo-sliced Drosophila embryos to determine genome-wide spatial patterns of

gene expression. *PLoS One*, *8*(8), e71820. https://doi.org/10.1371/journal.pone.0071820

Corbett-Detig, R. B., Russell, S. L., Nielsen, R., & Losos, J. (2020). Phenotypic convergence is not mirrored at the protein level in a lizard adaptive radiation. *Molecular Biology and Evolution*, *37*(6), 1604–1614. https://doi.org/10.1093/molbev/msaa028

Craddock, E. M., Kambysellis, M. P., Franchi, L., Francisco, P., Grey, M., Hutchinson, A., Nanhoo, S., & Antar, S. (2018). Ultrastructural variation and adaptive evolution of the ovipositor in the endemic Hawaiian Drosophilidae. *Journal of Morphology*, *279*(12), 1725–1752. https://doi.org/10.1002/jmor.20884

Darwin, C. (1845). *Journal of researches into the natural history and geology of the countries visited during the voyage of H.M.S. Beagle round the world*. John Murray.

Doudna, J. A., & Charpentier, E. (2014). The new frontier of genome engineering with CRISPR-Cas9. *Science*, *346*(6213), 1258096. https://doi.org/10.1126/science.1258096

Edwards, K. A., Doescher, L. T., Kaneshiro, K. Y., & Yamamoto, D. (2007). A database of wing diversity in the Hawaiian Drosophila. *PLoS One*, *2*(5), e487. https://doi.org/10.1371/journal.pone.0000487

Eldon, J., Bellinger, M. R., & Price, D. K. (2019). Hawaiian picture-winged Drosophila exhibit adaptive population divergence along a narrow climatic gradient on Hawaii Island. *Ecology and Evolution*, *9*(5), 2436–2448. https://doi.org/10.1002/ece3.4844

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., Couger, M. B., Eccles, D., Li, B., Lieber, M., MacManes, M. D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C. N., … Regev, A. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols*, *8*(8), 1494–1512. https://doi.org/10.1038/nprot.2013.084

Hahn, M. W., Han, M. V., & Han, S.-G. (2007). gene family evolution across 12 Drosophila genomes. *PLoS Genetics*, *3*(11), 197. https://doi.org/10.1371/journal.pgen.0030197

Heames, B., Schmitz, J., & Bornberg-Bauer, E. (2020). A continuum of evolving de novo genes drives protein-coding novelty in Drosophila. *Journal of Molecular Evolution*, *88*(4), 382–398. https://doi.org/10.1007/s00239-020-09939-z

Horváth, B., & Kalinka, A. T. (2018). The genetics of egg retention and fertilization success in Drosophila: One step closer to understanding the transition from facultative to obligate viviparity. *Evolution*, *72*(2), 318–336. https://doi.org/10.1111/evo.13411

Huang, D. W., Sherman, B. T., & Lempicki, R. A. (2009a). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols*, *4*(1), 44–57. https://doi.org/10.1038/nprot.2008.211

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

Hunter, S., Apweiler, R., Attwood, T. K., Bairoch, A., Bateman, A., Binns, D., Bork, P., Das, U., Daugherty, L., Duquenne, L., Finn, R. D., Gough, J., Haft, D., Hulo, N., Kahn, D., Kelly, E., Laugraud, A., Letunic, I., Lonsdale, D., … Yeats, C. (2009). InterPro: The integrative protein signature database. *Nucleic Acids Research*, *37*(suppl_1), D211–D215. https://doi.org/10.1093/nar/gkn785

Jagadeeshan, S., & Singh, R. S. (2005). Rapidly evolving genes of Drosophila: Differing levels of selective pressure in testis, ovary, and head tissues between sibling species. *Molecular Biology and Evolution*, *22*(9), 1793–1801. https://doi.org/10.1093/molbev/msi175

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., & Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*, *337*(6096), 816–821. https://doi.org/10.1126/science.1225829

Johnson, B. R. (2018). Taxonomically restricted genes are fundamental to biology and evolution. *Frontiers in Genetics*, *9*, 407. https://doi.org/10.3389/fgene.2018.00407

Kambysellis, M. P., & Heed, W. B. (1971). Studies of oogenesis in natural populations of Drosophilidae. I. Relation of ovarian development and ecological habitats of the Hawaiian species. *The American Naturalist*, *105*(941), 31–49.

Kambysellis, M. P., Starmer, T., Smathers, G., & Heed, W. B. (1980). Studies of oogenesis in natural populations of Drosophilidae. II. Significance of microclimatic changes on oogenesis of Drosophila mimica. *The American Naturalist*, *115*(1), 67–91.

Kang, L., Settlage, R., McMahon, W., Michalak, K., Tae, H., Garner, H. R., Stacy, E. A., Price, D. K., & Michalak, P. (2016). Genomic signatures of speciation in sympatric and allopatric Hawaiian picture-winged Drosophila. *Genome Biology and Evolution*, *8*(5), 1482–1488. https://doi.org/10.1093/gbe/evw095

Katoh, T., Izumitani, H. F., Yamashita, S., & Watada, M. (2017). Multiple origins of Hawaiian Drosophilids: Phylogeography of Scaptomyza Hardy (Diptera: Drosophilidae). *Entomological Science*, *20*(1), 33–44. https://doi.org/10.1111/ens.12222

Kim, B. Y., Wang, J. R., Miller, D. E., Barmina, O., Delaney, E., Thompson, A., Comeault, A. A., Peede, D., D'agostino, E. R., Pelaez, J., Aguilar, J. M., Haji, D., Matsunaga, T., Armstrong, E. E., Zych, M., Ogawa, Y., Stamenković-Radak, M., Jelić, M., Veselinović, M. S., … Petrov, D. A. (2021). Highly contiguous assemblies of 101 drosophilid genomes. *eLife*, *10*, e66405. https://doi.org/10.7554/eLife.66405

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., & Salzberg, S. L. (2013). TopHat2: Accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biology*, *14*, R36. https://doi.org/10.1186/gb-2013-14-4-r36

Klasberg, S., Bitard-Feildel, T., Callebaut, I., & Bornberg-Bauer, E. (2018). Origins and structural properties of novel and de novo protein domains during insect evolution. *The FEBS journal*, *285*(14), 2605–2625. https://doi.org/10.1111/febs.14504

Kuntz, S. G., & Eisen, M. B. (2014). Drosophila embryogenesis scales uniformly across temperature in developmentally diverse species. *PLoS Genetics*, *10*(4), e1004293. https://doi.org/10.1371/journal.pgen.1004293

Lange, A., Patel, P. H., Heames, B., Damry, A. M., Saenger, T., Jackson, C. J., Findlay, G. D., & Bornberg-Bauer, E. (2021). Structural and functional characterization of a putative de novo gene in Drosophila. *Nature Communications*, *12*(1), 1667. https://doi.org/10.1038/s41467-021-21667-6

Lapoint, R. T., Magnacca, K. N., & O'Grady, P. M. (2009). Review of the spoon tarsus subgroup of Hawaiian Drosophila (Drosophilidae: Diptera), with a description of one new species. *Zootaxa*, *2003*, 53–68.

Lapoint, R. T., Magnacca, K. N., & O'Grady, P. M. (2014). Phylogenetics of the Antopocerus—Modified Tarsus Clade of Hawaiian Drosophila: Diversification across the Hawaiian Islands. *PLoS One*, *9*(11), e113227. https://doi.org/10.1371/journal.pone.0113227

Lapoint, R. T., O'Grady, P. M., & Whiteman, N. K. (2013). Diversification and dispersal of the Hawaiian Drosophilidae: The evolution of Scaptomyza. *Molecular Phylogenetics and Evolution*, *69*(1), 95–108. https://doi.org/10.1016/j.ympev.2013.04.032

Larkin, A., Marygold, S. J., Antonazzo, G., Attrill, H., dos Santos, G., Garapati, P. V., Goodman, J. L., Gramates, L. S., Millburn, G., Strelets, V. B., Tabone, C. J., & Thurmond, J., FlyBase Consortium. (2021). FlyBase: Updates to the Drosophila melanogaster knowledge base. *Nucleic Acids Research*, *49*(D1), D899–D907. https://doi.org/10.1093/nar/gkaa1026

Losos, J. B., & Ricklefs, R. E. (2009). Adaptation and diversification on islands. *Nature*, *457*(7231), 830–836. https://doi.org/10.1038/nature07893

Lott, S. E., Villalta, J. E., Zhou, Q., Bachtrog, D., & Eisen, M. B. (2014). Sex-Specific Embryonic gene expression in species with newly evolved

sex chromosomes. *PLoS Genetics*, *10*(2), e1004159. https://doi.org/10.1371/journal.pgen.1004159

Lu, T.-C., Leu, J.-Y., & Lin, W.-C. (2017). A comprehensive analysis of transcript-supported de novo genes in Saccharomyces sensu stricto yeasts. *Molecular Biology and Evolution*, *34*(11), 2823–2838. https://doi.org/10.1093/molbev/msx210

Magnacca, K. N., & Grady, P. M. (2009). *Revision of the Modified Mouthparts Species Group of Hawaiian Drosophila (Diptera: Drosophilidae): The Ceratostoma, Freycinetiae, Semifuscata, and Setiger Subgroups, and Unplaced Species* (Vol. 130). University of California Press.

Magnacca, K. N., & Price, D. K. (2015). Rapid adaptive radiation and host plant conservation in the Hawaiian picture wing Drosophila (Diptera: Drosophilidae). *Molecular Phylogenetics and Evolution*, *92*, 226–242. https://doi.org/10.1016/j.ympev.2015.06.014

Mahler, D. L., Lambert, S. M., Geneva, A. J., Ng, J., Hedges, S. B., Losos, J. B., & Glor, R. E. (2016). Discovery of a Giant Chameleon-Like Lizard (Anolis) on Hispaniola and its significance to understanding replicated adaptive radiations. *The American Naturalist*, *188*(3), 357–364. https://doi.org/10.1086/687566

Mahler, D. L., Revell, L. J., Glor, R. E., & Losos, J. B. (2010). Ecological opportunity and the rate of morphological evolution in the diversification of Greater Antillean Anoles. *Evolution*, *64*(9), 2731–2745. https://doi.org/10.1111/j.1558-5646.2010.01026.x

Margaritis, L. H., Dellas, K., Kalantzi, M. C., & Kambysellis, M. P. (1983). Eggshell of Hawaiian Drosophila: Structural and biochemical studies in Drosophila grimshawi and comparison to Drosophila melanogaster. *Wilhelm Roux's Archives of Developmental Biology*, *192*, 303–316. https://scholar.google.com/scholar_lookup?title=eggshell%2Bof%2BHawaiian%2BDrosophila%3A%2Bstructural%2Band%2Bbiochemical%2Bstudies%2Bin%2BDrosophila%2Bgrimshawi%2Band%2Bcomparison%2Bto%2BDrosophila%2Bmelanogaster%26author=Margaritis%2C%2BL.H.&publication_year=1983

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal*, *17*(1), 10–12. https://doi.org/10.14806/ej.17.1.200

McCrea, P. D., Turck, C. W., & Gumbiner, B. (1991). A homolog of the armadillo protein in Drosophila (plakoglobin) associated with E-cadherin. *Science*, *254*(5036), 1359–1361.

McLysaght, A., & Guerzoni, D. (2015). New genes from non-coding sequence: The role of de novo protein-coding genes in eukaryotic evolutionary innovation. *Philosophical Transactions of the Royal Society, B: Biological Sciences*, *370*(1678), 20140332. https://doi.org/10.1098/rstb.2014.0332

Mestek Boukhibar, L., & Barkoulas, M. (2016). The developmental genetics of biological robustness. *Annals of Botany*, *117*(5), 699–707. https://doi.org/10.1093/aob/mcv128

Mohammed, J., Bortolamiol-Becet, D., Flynt, A. S., Gronau, I., Siepel, A., & Lai, E. C. (2014). Adaptive evolution of testis-specific, recently evolved, clustered miRNAs in Drosophila. *RNA*, *20*(8), 1195–1209. https://doi.org/10.1261/rna.044644.114

Montgomery, S. L. (1975). Comparative breeding site ecology and the adaptive radiation of picture-winged Drosophila (Diptera: Drosophilidae) in Hawaii. *Proceedings, Hawaiian Entomological Society*, *22*(1).

Neme, R., & Tautz, D. (2016). Fast turnover of genome transcription across evolutionary time exposes entire non-coding DNA to de novo gene emergence. *eLife*, *5*, e09977. https://doi.org/10.7554/eLife.09977

O'Grady, P., & DeSalle, R. (2008). Out of Hawaii: The origin and biogeography of the genus Scaptomyza (Diptera: Drosophilidae). *Biology Letters*, *4*(2), 195–199. https://doi.org/10.1098/rsbl.2007.0575

O'Grady, P., & DeSalle, R. (2018). Hawaiian Drosophila as an evolutionary model clade: Days of future past. *BioEssays*, *40*(5), 1700246. https://doi.org/10.1002/bies.201700246

Paris, M., Villalta, J. E., Eisen, M. B., & Lott, S. E. (2015). Sex bias and maternal contribution to gene expression divergence in Drosophila blastoderm embryos. *PLoS Genetics*, *11*(10), e1005592. https://doi.org/10.1371/journal.pgen.1005592

Peifer, M., Berg, S., & Reynolds, A. B. (1994). A repeating amino acid motif shared by proteins with diverse cellular roles. *Cell*, *76*(5), 789–791. https://doi.org/10.1016/0092-8674(94)90353-0

Rothwell, W. F., & Sullivan, W. (2007). Fixation of Drosophila embryos. *Cold Spring Harbor Protocols*, *2007*(9), pdb.prot4827. https://doi.org/10.1101/pdb.prot4827

Sarikaya, D. P., Church, S. H., Lagomarsino, L. P., Magnacca, K. N., Montgomery, S. L., Price, D. K., Kaneshiro, K. Y., & Extavour, C. G. (2019). Reproductive capacity evolves in response to ecology through common changes in cell number in Hawaiian Drosophila. *Current Biology*, *29*(11), 1877–1884 e6. https://doi.org/10.1016/j.cub.2019.04.063

Schmid, K. J., & Aquadro, C. F. (2001). The evolutionary analysis of "Orphans" from the Drosophila genome identifies rapidly diverging and incorrectly annotated genes. *Genetics*, *159*(2), 589–598. https://doi.org/10.1093/genetics/159.2.589

Stark, J. B., & O'Grady, P. M. (2010). Morphological variation in the forelegs of the Hawaiian Drosophilidae. I. The AMC clade. *Journal of Morphology*, *271*(1), 86–103. https://doi.org/10.1002/jmor.10783

Stern, D. L. (2011). Evolution, development, & the predictable genome. Roberts and Co. Publishers. http://agris.fao.org/agris-search/search.do?recordID=US201300151702

Sun, W., Zhao, X.-W., & Zhang, Z. (2015). Identification and evolution of the orphan genes in the domestic silkworm, Bombyx mori. *FEBS Letters*, *589*(19, Part B), 2731–2738. https://doi.org/10.1016/j.febslet.2015.08.008

Tadros, W., & Lipshitz, H. D. (2009). The maternal-to-zygotic transition: A play in two acts. *Development*, *136*(18), 3033–3042. https://doi.org/10.1242/dev.033183

Thompson, J. D., Gibson, Toby J., & Higgins, D. G. (2003). Multiple sequence alignment using ClustalW and ClustalX. *Current Protocols in Bioinformatics*, *00*(1), 2.3.1–2.3.22. https://doi.org/10.1002/0471250953.bi0203s00

Trapnell, C., Hendrickson, D. G., Sauvageau, M., Goff, L., Rinn, J. L., & Pachter, L. (2013). Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nature Biotechnology*, *31*(1), 46–53. https://doi.org/10.1038/nbt.2450

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., Pimentel, H., Salzberg, S. L., Rinn, J. L., & Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature Protocols*, *7*(3), 562–578. https://doi.org/10.1038/nprot.2012.016

Van Oss, S. B., & Carvunis, A.-R. (2019). De novo gene birth. *PLoS Genetics*, *15*(5), e1008160. https://doi.org/10.1371/journal.pgen.1008160

Walter, W., Sánchez-Cabo, F., & Ricote, M. (2015). GOplot: An R package for visually combining expression data with functional analysis. *Bioinformatics*, *31*(17), 2912–2914. https://doi.org/10.1093/bioinformatics/btv300

Wheeler, M. R., & Clayton, F. (1965). A new Drosophila culture technique. *Drosophila Information Service*, *40*, 98.

Whittaker, R. J., Fernández-Palacios, J. M., Matthews, T. J., Borregaard, M. K., & Triantis, K. A. (2017). Island biogeography: Taking the long view of nature's laboratories. *Science*, *357*(6354), https://doi.org/10.1126/science.aam8326

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

Xia, S., VanKuren, N. W., Chen, C., Zhang, L., Kemkemer, C., Shao, Y., Jia, H., Lee, U., Advani, A. S., Gschwend, A., Vibranovski, M. D.,

Chen, S., Zhang, Y. E., & Long, M. (2021). Genomic analyses of new genes and their phenotypic effects reveal rapid evolution of essential functions in Drosophila development. *PLoS Genetics*, *17*(7), e1009654. https://doi.org/10.1371/journal.pgen.1009654

Xie, C., Zhang, Y. E., Chen, J.-Y., Liu, C.-J., Zhou, W.-Z., Li, Y., Zhang, M., Zhang, R., Wei, L., & Li, C.-Y. (2012). Hominoid-specific de novo protein-coding genes originating from long non-coding RNAs. *PLoS Genetics*, *8*(9), e1002942. https://doi.org/10.1371/journal.pgen.1002942

Yang, H., Jaime, M., Polihronakis, M., Kanegawa, K., Markow, T., Kaneshiro, K., & Oliver, B. (2018). Re-annotation of eight Drosophila genomes. *Life Science Alliance*, *1*(6), 201800156. https://doi.org/10.26508/lsa.201800156

Zhao, L., Saelao, P., Jones, C. D., & Begun, D. J. (2014). Origin and spread of de novo genes in Drosophila melanogaster populations. *Science*, *343*(6172), 769–772. https://doi.org/10.1126/science.1248286

Zhong, Y., Jia, Y., Gao, Y., Tian, D., Yang, S., & Zhang, X. (2013). Functional requirements driving the gene duplication in 12 Drosophila species. *BMC Genomics*, *14*(1), 555. https://doi.org/10.1186/1471-2164-14-555

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Chenevert, M., Miller, B., Karkoutli, A., Rusnak, A., Lott, S., & Atallah, J. (2022). The early embryonic transcriptome of a Hawaiian *Drosophila* picture-wing fly shows evidence of altered gene expression and novel gene evolution. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, 1–15. https://doi.org/10.1002/jez.b.23129