






Genomics analysis of hexanoic acid exposure in *Drosophila* species

Zachary A. Drum , Stephen M. Lanno, Sara M. Gregory, Serena J. Shimshak, Mukshud Ahamed, Will Barr , Bethlehem Bekele, Alison Biester , Colleen Castro , Lauren Connolly, Nicole DelGaudio, William Humphrey, Helen Karimi, Sophie Karolczak, Tay-Shaun Lawrence, Andrew McCracken, Noah Miller-Medzon, Leah Murphy, Cameron Park, Sojeong Park, Chloe Qiu, Kevin Serra, Gigi Snyder, Alexa Strauss, Spencer Tang, Christina Vyzas, and Joseph D. Coolon *

Department of Biology, Wesleyan University, Middletown, CT 06457, USA

*Corresponding author: Wesleyan University, 123 Hall-Atwater, Middletown, CT 06459, USA. Email: jcoolon@wesleyan.edu

Abstract

Drosophila sechellia is a dietary specialist endemic to the Seychelles islands that has evolved to consume the fruit of *Morinda citrifolia*. When ripe, the fruit of *M. citrifolia* contains octanoic acid and hexanoic acid, two medium-chain fatty acid volatiles that deter and are toxic to generalist insects. *Drosophila sechellia* has evolved resistance to these volatiles allowing it to feed almost exclusively on this host plant. The genetic basis of octanoic acid resistance has been the focus of multiple recent studies, but the mechanisms that govern hexanoic acid resistance in *D. sechellia* remain unknown. To understand how *D. sechellia* has evolved to specialize on *M. citrifolia* fruit and avoid the toxic effects of hexanoic acid, we exposed adult *D. sechellia*, *D. melanogaster* and *D. simulans* to hexanoic acid and performed RNA sequencing comparing their transcriptional responses to identify *D. sechellia* specific responses. Our analysis identified many more genes responding transcriptionally to hexanoic acid in the susceptible generalist species than in the specialist *D. sechellia*. Interrogation of the sets of differentially expressed genes showed that generalists regulated the expression of many genes involved in metabolism and detoxification whereas the specialist primarily downregulated genes involved in the innate immunity. Using these data, we have identified interesting candidate genes that may be critically important in aspects of adaptation to their food source that contains high concentrations of HA. Understanding how gene expression evolves during dietary specialization is crucial for our understanding of how ecological communities are built and how evolution shapes trophic interactions.

Keywords: toxin resistance; host specialization; RNA-seq

Introduction

Insects have long been recognized as one of the most abundant and diverse groups of organisms on the planet, with a large fraction of them feeding on plants (Jaenike 1990; Stork 2018). Many of these phytophagous insects have evolved to be highly host plant specific. The evolution of such specialized interactions is often guided by specific plant chemistry, with most plants responding to increased insect herbivory by the production of toxic secondary metabolites (Jaenike 1990; Petschenka and Agrawal 2016). This leads to an evolutionary arms race as insects evolve resistance to these toxins (Heidel-Fischer and Vogel 2015). While host plant-insect adaptations are well studied in the literature, less is known about the underlying genetic mechanisms that contribute to the evolution of these complex ecological interactions (Ungerer et al. 2008).

Drosophila sechellia feeds and oviposits primarily on *Morinda citrifolia*, a fruit highly toxic to other *Drosophila* species (Legal et al. 1992). The plant produces ripe fruit year-round in the Seychelles island archipelago, the sole location where *D. sechellia* are found (Legal et al. 1992) providing abundant and consistent resources. Upon exposure to the ripe fruit, other *Drosophila* species display

frantic behavior and wing movements, reduction in locomotor activity, and death (Legal et al. 1994). *Drosophila sechellia* eggs are able to hatch and develop on *M. citrifolia* fruit but the embryos of other *Drosophila* species die (Amlou et al. 1998). Unlike its generalist sister species, *D. sechellia* prefer *M. citrifolia* to other hosts and are drawn to the fruit from a long distance (R'Kha et al. 1991). *Drosophila sechellia* evolved resistance to the toxins in *M. citrifolia* from an ancestral sensitive state (R'Kha et al., 1997) and this tolerance provides *D. sechellia* a temporal advantage over other fruit fly species that can only lay eggs in the fruit once it has rotten and the toxic volatiles are reduced.

The toxic properties of *Morinda* are attributed to the carboxylic acids present in the pulp of the ripe fruit (Legal et al. 1994). This pulp is largely characterized by carboxylic acids, primarily the fatty acids octanoic (OA) and hexanoic acid, (HA) which comprise 58% and 19% of the volatile compounds found the ripe fruit respectively (Farine et al. 1996). While the genetic basis of *D. sechellia* resistance to the most abundant and toxic compound in *M. citrifolia* fruit, OA, has been characterized in previous studies (Dworkin and Jones 2009; Andrade Lopez et al. 2017; Lanno et al. 2017; Peyser et al. 2017; Lanno and Coolon 2019; Lanno et al. 2019a,

Received: June 07, 2021. Accepted: September 10, 2021

© The Author(s) 2021. Published by Oxford University Press on behalf of Genetics Society of America.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

2019b), much less is known about the genes involved in HA resistance. In a study using fatty acid concentrations equivalent to 1.5 g of ripe *M. citrifolia* fruit, OA treatment alone killed all fruit fly species assayed except *D. sechellia*, while HA alone caused reversible knock-down in other *Drosophilids* (Farine et al. 1996). In a more recent study using higher concentrations of HA, mortality was observed in response to exposure to HA in *D. melanogaster* and *D. simulans* and *D. sechellia* is significantly more resistant to HA induced mortality than sister species (Peysers et al. 2017; Lanno and Coolon 2019). Surprisingly, tests of the three major detoxification gene families, cytochrome P450s (cyps), glutathione-S-transferases (GSTs) and esterases (Ests) found that none was involved in derived HA resistance in *D. sechellia* suggesting an alternative genetic mechanism must be involved in resistance. Studies have suggested that HA may be a more efficient *D. sechellia* attractant than OA, while OA is a more potent repellent of generalist species (Amlou et al. 1998). In a test using laboratory food medium supplemented with 0.5% of either OA or HA, *D. sechellia* exhibited oviposition preference for media supplemented with HA over OA (Amlou et al. 1998).

From a study analyzing the transcriptomic response of *D. sechellia* on OA, 104 genes were found to be differentially expressed in response to OA (Lanno et al. 2017). This study showed that several *Osiris* genes, including *Osi6* are upregulated in *D. sechellia* in response to OA. Another study showed that RNAi mediated knockdown of *Osi6* expression drastically decreased survival in response to OA (Andrade Lopez et al. 2017). Given that HA makes up about of fifth of the volatile compounds found in *M. citrifolia* and produces unique effects on behavior (Farine et al. 1996) and has an unknown and less common basis for toxin resistance (Peysers et al. 2017; Lanno and Coolon 2019), identifying the genes responding to HA is necessary to understand how *D. sechellia* has specialized on *M. citrifolia* fruit, and may help pinpoint genes that are involved in resistance to HA. By also analyzing the genes responding to HA exposure in generalist *Drosophila* species *D. melanogaster* and *D. simulans* we can identify derived gene expression responses specific to *D. sechellia* that may be critical for HA associated traits (Coolon et al. 2009). In this study, adult female *D. sechellia*, *D. melanogaster*, and *D. simulans* flies were fed either control food or food supplemented with 0.23% HA and significantly differentially expressed genes (DEGs) were identified using RNA-seq. Comparison of the identified genes with those found to respond to OA (Lanno et al. 2017) and L-DOPA (Lanno et al. 2019), another highly abundant compound found in *M. citrifolia* fruit, identified several genes common in response to OA, HA, and L-DOPA as well as genes unique to HA suggesting these genes may play an important role in the evolved resistance and specialization of *D. sechellia* to *M. citrifolia*.

Methods

Fly strains and culture

Drosophila sechellia (14021-0428.25), *D. simulans* (14021-0251.195), and *D. melanogaster* (14021-0231.36) flies were reared on standard cornmeal medium under a 16:8 light: dark cycle maintained at 20°C.

RNA extraction, library preparation, and sequencing

Zero to 3-day posteclosion adult female flies were fed control food (0.75 g *Drosophila* instant medium Formula 4-24, Carolina

Biological Supply Company) or identical food containing 0.23% hexanoic acid (HA). After 24 h, three replicates of ten whole flies per species and per treatment were homogenized and total RNA was extracted with a modified protocol of the Promega SV extraction system (Coolon et al. 2013; Figure 1). RNA quality and quantity was assessed using agarose gel electrophoresis and Nanodrop spectrophotometry. RNA was sent to the University of Michigan Sequencing Core Facility for poly-A selection, cDNA synthesis, bar-coded library preparation with TruSeq library preparation kits and sequencing on an Illumina platform, generating 405,166,795 single-end 65 nt sequencing reads for *D. sechellia* and 51 nt sequencing reads for *D. melanogaster* and *D. simulans* (Table 1).

BIOL310 genomics analysis

The genomics analysis of RNA-seq data presented in this manuscript was performed by 20 undergraduate and 6 graduate students as part of a semester-long course at Wesleyan University called Genomics Analysis (BIOL310). This is the third such manuscript (see Lanno et al. 2017, 2019a) made from this course where the aim is to provide undergraduate students an early opportunity with a course-based research experience with active participation in scientific discovery. Students in the course learn through engaging with never-before analyzed data using cutting edge genomics analysis techniques and bioinformatics tools through a discovery-based independent study. Every student in the course contributed to the quality control, analyses, write-up, and interpretation of the findings, providing their own unique perspective of the results and text written by each and every student was combined into this manuscript with very little modification.

After sequencing output files were obtained from the University of Michigan Sequencing Core (Table 1), fastq files containing raw sequencing reads were uploaded to the Galaxy platform (Afgan et al. 2016) and an RNA-seq analysis pipeline was performed (Figure 1) as previously described (Lanno et al. 2017, 2019a). Briefly, reads were assessed for quality using FASTQC (Andrews 2010), and any overrepresented sequences were analyzed using NCBI Blast (Altschul et al. 1990). Bowtie2 was used for mapping reads to the appropriate reference genome for each species with default parameters (Langmead and Salzberg 2012), with the most recent genomes for each species available at the time of analysis acquired from Ensembl (www.ensembl.org, Yates et al., 2020) (*D. sechellia*: *Drosophila_sechellia*.dsec_caf1.dna.toplevel.fa, *D. simulans*: *Drosophila_simulans*.ASM75419v3.dna.toplevel.fa and *D. melanogaster*: *Drosophila_melanogaster*.BDGP6.dna.toplevel.fa). The Bowtie2 output files were analyzed using Cuffdiff (Trapnell et al. 2010), which performs gene expression quantification and differential gene expression analysis using the aforementioned genome file along with the most recent annotated .gff3 file for each genome available at the time of analysis acquired from Ensembl (*D. sechellia*: *Drosophila_sechellia*.dsec_caf1.42.gff3, *D. simulans*: *Drosophila_simulans*.ASM75419v3.42.gff3 and *D. melanogaster*: *Drosophila_melanogaster*.BDGP6.95.gff3). In Cuffdiff, geometric normalization and library size correction was performed, along with bias correction using the reference genome, giving an output of DEGs for each species following false discovery rate multiple testing correction (Benjamini and Hochberg 1995, $q < 0.05$). Data were visualized using R (R Core Development Team 2010). In order to compare gene expression results across species, we obtained all 1:1:1 orthologs from *D. sechellia*, *D. simulans*, and *D. melanogaster* from Flybase (Thurmond

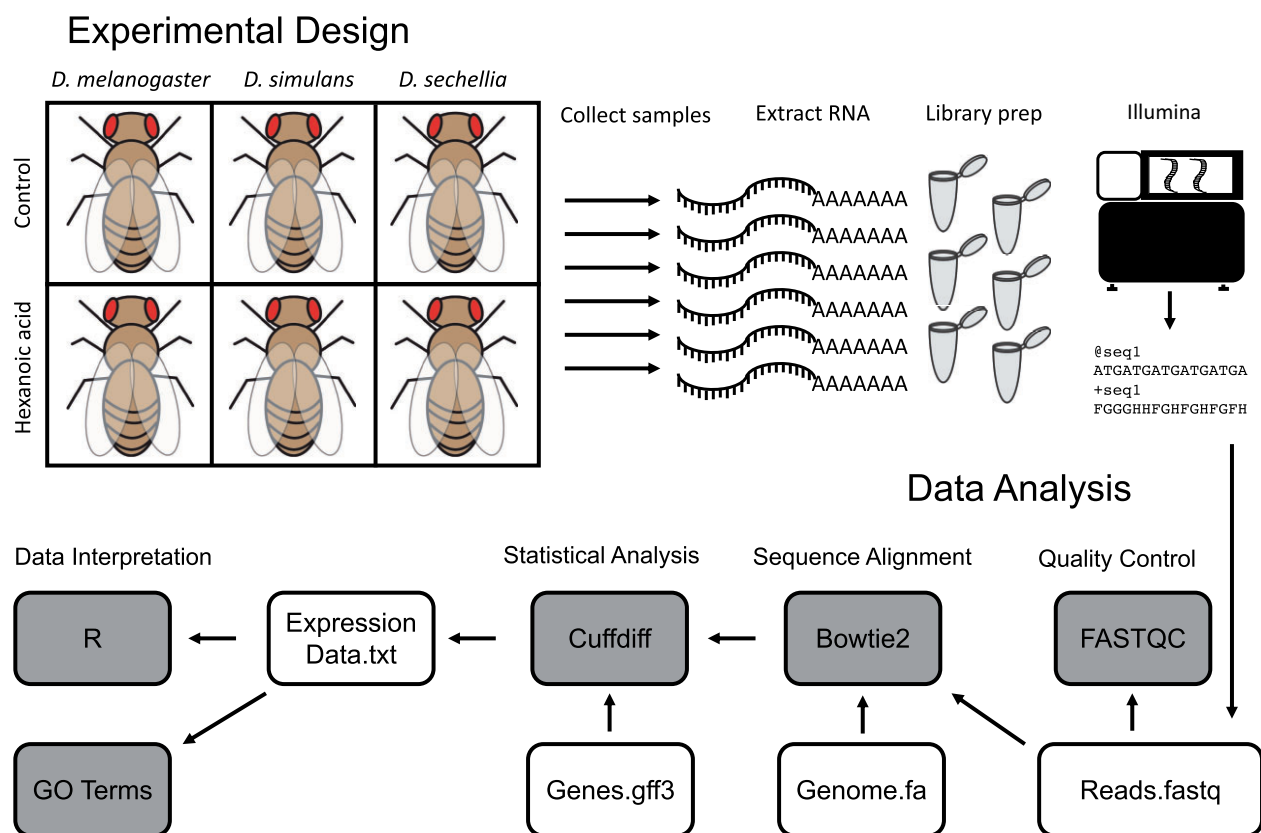


Figure 1 Experimental design and RNA-sequencing pipeline analysis. (A) Zero to 3-day-old adult female *D. sechellia*, *D. simulans*, and *D. melanogaster* flies were treated for 24 h on either control food or food supplemented with 0.23% hexanoic acid (HA). Flies were frozen in liquid nitrogen, RNA was extracted, library prep was performed using poly-A selection, and libraries were sequenced. (B) Raw sequencing reads were analyzed using FASTQC for quality control, and then aligned to the respective reference genome using Bowtie2. Differential expression was quantified using Cuffdiff using available respective genome annotation. Data were visualized using R, and significantly differentially expressed genes (DEGs) were used for downstream analysis.

Table 1 Sequencing results and mapping percentage

Sample	ID	No. of reads	No. of mapped reads	% Mapped	Read length (nt)
sim-C-1	105545	28,056,123	26,210,691	93.42	51
sim-C-2	105546	26,058,213	24,449,785	93.82	51
sim-C-3	105547	24,095,284	22,589,715	93.75	51
sim-HA-1	105557	16,844,466	15,844,735	96.06	51
sim-HA-2	105558	27,002,538	25,521,057	94.51	51
sim-HA-3	105559	23,222,911	21,786,539	93.81	51
mel-C-1	105542	21,999,530	20,633,866	93.79	51
mel-C-2	105543	20,950,464	19,779,953	94.41	51
mel-C-3	105544	22,157,160	20,919,514	94.41	51
mel-HA-1	105554	19,811,200	18,745,553	94.62	51
mel-HA-2	105555	23,577,339	22,322,344	94.68	51
mel-HA-3	105556	18,208,025	17,257,250	94.78	51
sec-C-1	76332	19,222,060	18,496,450	96.23	65
sec-C-2	76333	20,704,811	19,440,620	93.89	65
sec-C-3	76334	17,696,868	17,123,579	96.76	65
sec-HA-1	76338	30,612,710	29,079,271	94.99	65
sec-HA-2	76339	25,873,039	25,202,396	97.41	65
sec-HA-3	76340	19,074,054	18,031,453	94.53	65

et al. 2019). DEGs following *D. sechellia* exposure to OA or L-DOPA were downloaded from online databases (Lanno et al. 2017, 2019a). GO term enrichment was performed on *D. melanogaster* orthologs for each species using GeneOntology.org (www.geneontology.org,

Ashburner et al. 2000; Mi et al. 2019; Carbon et al. 2021). KEGG pathway analysis was performed using the *D. melanogaster* ortholog for each DEG from each species (https://www.kegg.jp/kegg/tool/map_pathway1.html, Kanehisa and Sato 2020).

Results

Differential gene expression in response to HA treatment

In order to identify candidate genes that are important in *D. sechellia* host specialization and evolved resistance to HA, we sought genes that have altered expression levels in response to HA exposure. Previous studies have shown that such environmentally plastic gene regulation can indicate the importance of that gene's function in that environment making identified genes good candidates for *D. sechellia* HA associated traits (Coolon et al. 2009; Lanno et al. 2017, 2019). To quantify gene expression response to HA we performed RNA-sequencing (RNA-seq) on adult female flies after exposure to control food and compared this to flies fed food containing 0.23% HA. Because many of the transcriptional responses to HA might be nonspecific, we measured gene expression responses in *D. sechellia*, *D. melanogaster*, and *D. simulans* to identify those responses (or loss of response) that are restricted to *D. sechellia* representing changes that might contribute to its unique phenotypes.

Using this approach, we identified 841 genes differentially expressed by *D. melanogaster* (Figure 2, A and D; Supplementary Table S1), 743 genes were differentially expressed by *D. simulans* (Figure 2, B and E; Supplementary Table S2) and only 93 genes were differentially expressed in *D. sechellia* (Figure 2, C and F; Supplementary Table S3) in response to HA. No significant difference in the number of upregulated genes (50/93) vs downregulated genes (43/93) was observed in *D. sechellia* in response to HA (Binomial Exact Test, $P=0.1066$). In *D. simulans*, there was a significant difference in the number of upregulated genes (69/743) compared to the number of downregulated genes (674/743) in response to HA (Binomial Exact Test, $P=2.2e-16$, Figure 2, B and E). In *D. melanogaster*, there was also a significant difference in the number of upregulated genes (171/841) compared to the number of downregulated genes (670/841) in response to HA (Binomial

Exact Test, $P=2.2e-16$, Figure 2, A and D). In *D. sechellia*, there were 39 DEGs identified that responded to HA that do not have annotated *D. melanogaster* orthologs. 27 of these genes were 5.8S rRNAs, two snoRNAs, and 7 genes of unknown function. Of the 27 5.8S rRNAs, all 27 were upregulated (Binomial Exact Test, $P=7.451e-09$). In *D. simulans*, of the 743 DEGs there were annotated *D. melanogaster* orthologs for 673 genes. For the remainder of the analysis, only genes with known *D. melanogaster* orthologs are considered to allow functional interpretations of DEGs, and for all subsequent analyses the *D. melanogaster* ortholog name was used. This filtering resulted in 673 differentially genes in *D. simulans*, 841 in *D. melanogaster* and 54 in *D. sechellia* used in subsequent analyses.

Identifying functional enrichment in DEGs

To identify the biological pathways that are involved in responses to HA in *D. sechellia*, *D. melanogaster*, and *D. simulans* KEGG analyses of the upregulated and downregulated genes in each species were performed. These analyses show that many different metabolic and detoxification pathways along with proteins that localize to the lysosome are changing in response to HA exposure in *D. melanogaster* and *D. simulans*, whereas very few genes in these pathways are responding in *D. sechellia* (Figure 3, A and B). In *D. sechellia*, many of the genes downregulated in response to HA are involved in immune processes whereas this was much less prominent in *D. melanogaster* and *D. simulans* DEGs (Figure 3C).

To understand which biological and cellular processes are being altered in *D. sechellia*, *D. melanogaster*, and *D. simulans* in response to HA, Gene Ontology (GO) term enrichment analysis was performed (Supplementary Tables S16–S18). In *D. sechellia*, GO term enrichment analysis of DEGs for cellular component showed a significant enrichment for extracellular region genes ($P=4.83e-05$), suggesting that several of the genes responding

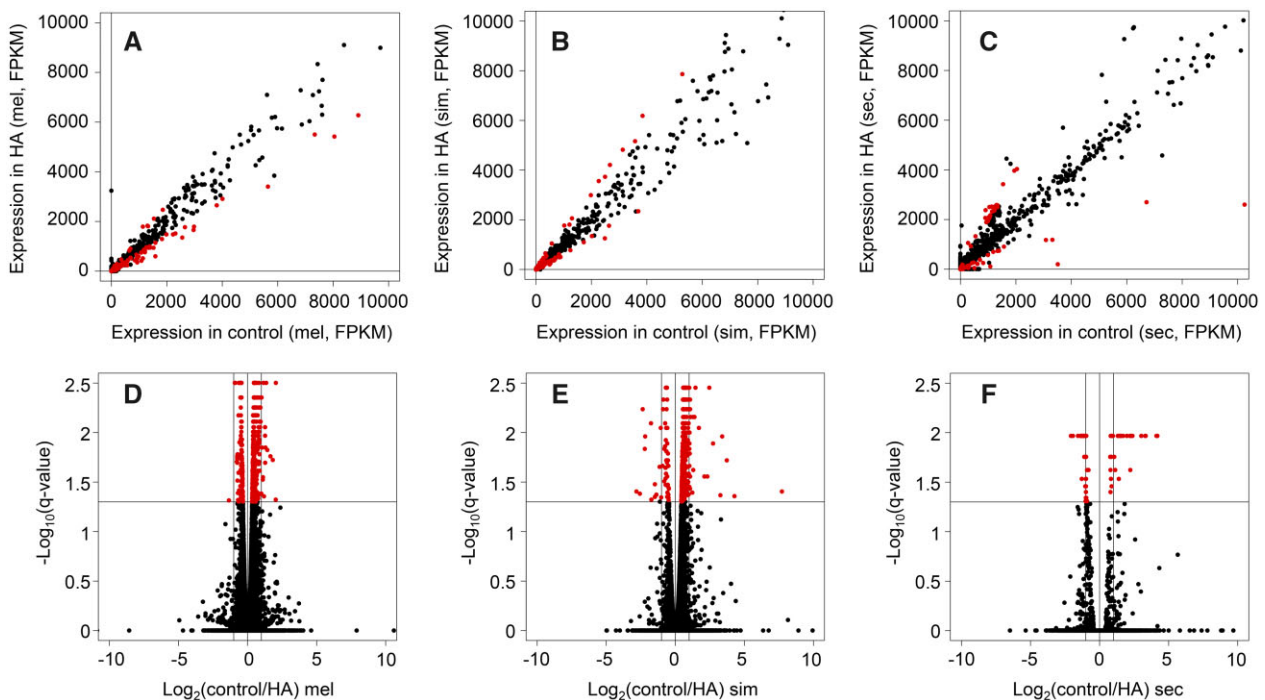


Figure 2 DEGs in adult *D. melanogaster*, *D. simulans*, and *D. sechellia* in response to HA. (A–C) Plots showing DEG expression in control vs expression in HA treatment in (A) *D. melanogaster*, (B) *D. simulans*, and (C) *D. sechellia*. Statistically significant genes are shown in red. (D–F) Plots showing $\log_2(\text{control/HA})$ fold change in (D) *D. melanogaster*, (E) *D. simulans*, and (F) *D. sechellia*. Statistically significant genes are shown in red.

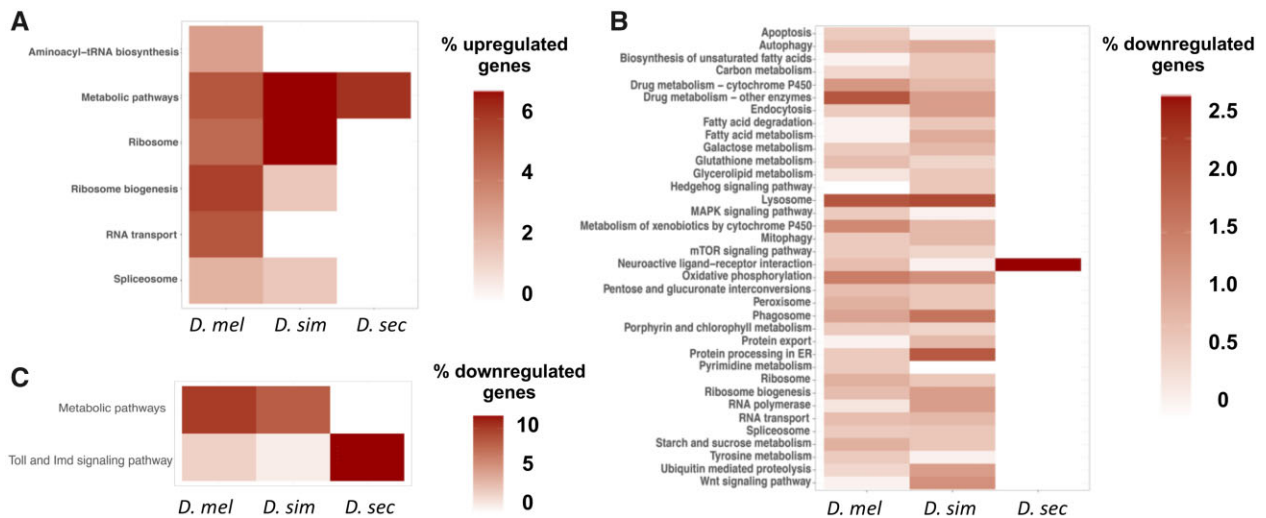


Figure 3 KEGG and GO analysis of DEGs in response to HA. (A) Significantly upregulated genes with *D. melanogaster* orthologs for each species are shown by the percentage of genes of the total number upregulated genes that fall into each KEGG pathway. (B) Significantly downregulated genes with *D. melanogaster* orthologs for each species are shown by the percentage of genes out of the total number of downregulated genes that fall into each KEGG pathway. The KEGG pathways Metabolic pathways and Toll and Imd signaling were excluded. (C) Significantly downregulated genes with *D. melanogaster* orthologs for each species are shown by the percentage of genes out of the total number of downregulated genes that fall into the Metabolic pathways and Toll and Imd signaling pathway KEGG pathways.

to HA exposure have proteins that are secreted. For biological process GO term enrichment analysis, processes involved in the antibacterial humoral response were significantly enriched ($P=1.75e-08$). In *D. sechellia*, no molecular function processes were significantly enriched. In *D. melanogaster*, there was also significant enrichment for GO terms for extracellular region genes ($P=4.42e-04$), along with genes found inside the nucleolus ($P=8.40e-04$). For biological process, GO terms associated with genes involved in the antibacterial humoral response were significantly enriched as in *D. sechellia* ($P=3.02e-02$), as were other terms involved in the *Drosophila* immunity. The *D. melanogaster* HA response also was enriched for genes involved in ribosome biogenesis ($P=3.66e-02$). In *D. simulans*, genes found inside the nucleolus were significantly enriched in the set of HA responsive genes ($P=2.65e-02$). In an analysis of DEGs in *D. simulans* that are upregulated, genes involved in Notch signaling were significantly enriched ($P=4.11e-04$) alongside genes involved in vitelline membrane and chorion formation ($P=2.56e-02$). In *D. sechellia*, upregulated genes were significantly enriched for the larval serum protein complex ($P=1.96e-02$). Downregulated genes were significantly enriched with antibacterial humoral response GO terms ($P=3.24e-10$) along with the response to hyperoxia ($P=1.52e-03$). These downregulated genes were also enriched for the extracellular region ($P=4.17e-05$). In *D. melanogaster*, upregulated genes were significantly enriched to be involved in the larval serum protein complex ($P=1.23e-03$) and were enriched intracellularly ($P=4.02e-16$) and within intracellular organelles ($1.32e-13$). These upregulated genes were significantly enriched in many biological processes, including chromatin silencing ($P=3.25e-03$), ecdysone receptor-mediated signaling ($P=1.89e-02$), and chorion assembly ($P=2.25e-02$). In *D. melanogaster*, downregulated genes were significantly enriched for the antibacterial humoral response ($P=4.52e-03$) and for the defense response to Gram-positive bacteria ($P=5.07e-04$). These downregulated genes were significantly enriched to be localized to the extracellular region ($P=1.25e-04$).

Comparing DEGs identified in response to HA in *D. melanogaster*, *D. simulans*, and *D. sechellia*

In order to identify genes with *D. sechellia* specific responses to HA, we compared the DEGs from the three species when exposed to HA (Figure 4A). We identified 33 genes that were differentially expressed by *D. sechellia* flies exposed to HA that were not responsive to HA exposure in both *D. melanogaster* and *D. simulans* flies (Figure 4A, Supplementary Table S13). Gene ontology enrichment analysis was performed for the 33 DEGs that are unique to the response to HA in *D. sechellia* and the downregulated DEGs are significantly enriched for immune response GO terms and response to stress (GO:0006950, Supplementary Table S23). There is no significant enrichment of unique upregulated DEGs in *D. sechellia*. In order to identify those genes where *D. sechellia* specific loss of response to HA was observed we selected those genes with significant changes in response to HA in both *D. melanogaster* and *D. simulans* that were not significantly differentially expressed by *D. sechellia* in response to HA. This analysis yields a total of 213 genes with this expression pattern in our data (Supplementary Table S10). Of these 213 DEGs shared between *D. melanogaster* and *D. simulans*, 205 of these genes are downregulated in both species in response to HA (Supplementary Table S10). GO analysis of these DEGs shared between *D. melanogaster* and *D. simulans* gives no significant enrichment. There are only 4 DEGs unique to *D. melanogaster* and *D. simulans* that are upregulated in both species (CG11835, *Cpr62Bc*, *cype*, and *RpL6*). Interestingly, only 2 genes were identified, CG13114 and *Fbp1*, that had significant response to HA in all three species, with the expression of both increasing in *D. simulans* and decreasing in *D. sechellia* and *D. melanogaster* (Figure 4A, Supplementary Table S9).

Common transcriptional responses of *D. sechellia* exposed to HA, OA, and L-DOPA suggests overlapping regulatory mechanisms

To assess the overlap in transcriptional responses of *D. sechellia* to both OA and HA, the predominant fatty acid volatiles in *M. citrifolia*, along with 3,4-dihydroxyphenylalanine (L-DOPA), which is found in *M. citrifolia* fruit and is important for the specialization

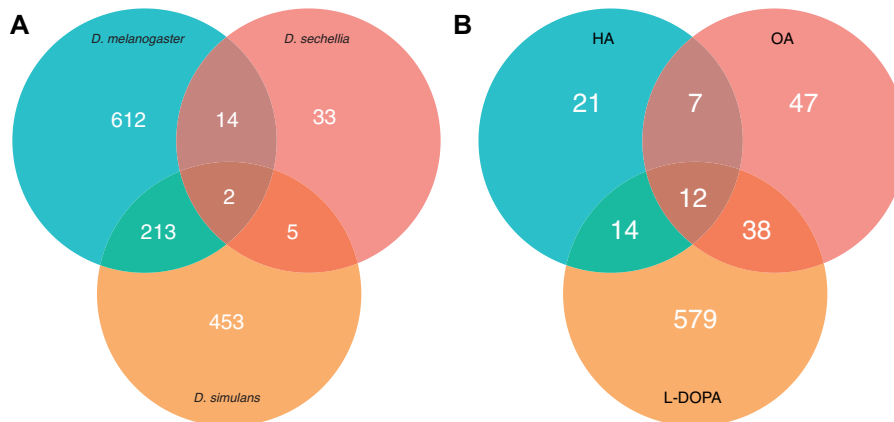


Figure 4 DEG overlap between species and treatments. (A) The number of differentially expressed genes in response to HA after RNA-seq are shown for each species, *D. sechellia*, *D. simulans*, and *D. melanogaster*. Overlap and species-specific number of DEGs are indicated. (B) The number of differentially expressed genes in *D. sechellia* when it is exposed to OA, HA, or L-DOPA treatment as well as the number of specific and overlapping genes are shown.

of *D. sechellia* to this fruit (Lavista-Llanos et al. 2014; Lanno et al. 2019a), we compared DEGs between our HA treatment, DEGs identified in a previous study using a 0.7% OA treatment (Lanno et al. 2017), and DEGs identified in a previous study of responses to 10 mg/mL of L-DOPA added to the fly food (Lanno et al. 2019a). Treatment with OA treatment yielded 103 *D. sechellia* genes with *D. melanogaster* orthologs that were significantly differentially expressed. Treatment with L-DOPA yielded 643 *D. sechellia* genes with *D. melanogaster* orthologs that were significantly differentially expressed. Comparison of genes responsive to HA, OA, and L-DOPA identified 12 DEGs that respond to HA, OA, and L-DOPA treatment in *D. sechellia* (Figure 4B, Supplementary Table S6). Interestingly, of the 19 shared DEGs with *D. melanogaster* orthologs between OA and HA treatments, all genes are downregulated except for E(spl)mgamma-HLH, which is upregulated upon both OA and HA treatment (Supplementary Table S4). To compare metabolic and cellular pathways involved in the conserved response between OA and HA treatment in *D. sechellia*, GO term enrichment analysis was performed on shared DEGs. Genes involved with biological processes related to the humoral immune response were overrepresented ($P=2.95e-08$, Supplementary Table S22). No significantly enriched processes were found for molecular function or cellular component. A KEGG pathway analysis of DEGs in *D. sechellia* similarly found that upon OA, *D. sechellia* downregulated genes involved in both the Toll and Imd signaling pathways as well as genes involved in metabolic processes (Supplementary Table S21 and Figure S1). Interestingly, in response to OA and HA, *D. sechellia* downregulated genes involved the humoral immune response (AttC, CecA2, Def, DptB, Dro, edin, and PGRP-SB1 Supplementary Table S4).

Discussion

Understanding the genetic basis of how organisms evolve to occupy different ecological niches and adapt to their environments is crucial to understanding the evolution of plant and animal interactions. Insect-host plant specialization is an excellent example of the evolution of such interactions and has been the subject of numerous ecological studies. While the phenomenon is well documented, the genetic basis of evolved host specialization is still not widely understood. Here, we focus on the specialization of *D. sechellia* to feed almost exclusively on one host plant, *M. citrifolia* because it is an excellent model to understand the genetic basis of dietary specialization. This is in part because it is

has evolved recently and very closely related to the genetic model generalist species *D. melanogaster*. Fortunately, we can take advantage of the wealth of genetic tools and information about *D. melanogaster* and sister species to understand the evolution of dietary specialization in this group (Groen and Whiteman 2016).

Drosophila sechellia upregulates a single cytochrome P450 (*Cyp4e1*) in response to HA whereas *D. simulans* downregulates this same gene (Supplementary Table S2), but previous work has shown that cytochrome P450s are not the evolved mechanism by which *D. sechellia* is able to survive OA or HA exposure (Peysers et al. 2017). Previous work has shown that HA induces a “reversible coma” in generalist *D. melanogaster*, *D. simulans*, and *D. mauritiana* flies (Farine et al. 1996), and *D. sechellia* prefers to oviposit on HA compared to OA and control food sources (Amlou et al. 1998). *Drosophila sechellia* has a premature stop codon in *Obp56e* as well as *Obp57d* and *Obp57e* alleles that reduce their avoidance to noni volatiles (Matsuo et al. 2007; Dworkin and Jones 2009). The mechanisms that drive this attraction may be through changes in gene expression and may be reflected in these predicted regulatory networks. *Drosophila sechellia* downregulates many genes involved in the immune response when exposed to both HA and OA, suggesting that somehow these two medium-chain fatty acids are interacting with negative effectors of Imd or Toll signaling. As these two immune pathways have similar and overlapping target genes, further analysis is needed to determine which immune pathways are involved in these interactions (Hanson and Lemaitre 2020). This interaction between *D. sechellia* and the volatiles from *M. citrifolia* reduces the immune response, which could cause alterations in gut microbiota composition that aids in the detoxification of relevant plant secondary metabolites. E(spl)mgamma-HLH is a Notch responsive Myc-like transcription factor that has been shown to interact with Relish, the main regulator of Imd immune signaling (Dushay et al. 1996) through recent yeast two-hybrid assays (Shokri et al. 2019) and is upregulated in *D. sechellia* in response to both OA and HA exposure (Figure 4B, Supplementary Table S4). The plastic response of insect immune systems allows them to fend off pathogens when needed, but also allow for the management of endosymbionts (Login et al. 2011; Vilcinskis 2013). Unlike its generalist sister species, *Drosophila sechellia* was previously shown to lack an immune response when confronted with parasitic wasps, hinting that immune system responses in this species may be unlike its generalist sister species (Salazar-Jaramillo et al. 2017). In addition, prior work showed that free fatty acids are involved in regulating

immune responses in mammals (Alvarez-Curto and Milligan 2016), and hexanoic acid priming of plants can activate the jasmonic acid pathway to increase plant resistance to fungal pathogens (García-Robles et al. 2013; Aranega-Bou et al. 2014). Using functional studies to alter the expression of known immune effectors and test if this alters resistance or preference for noni volatiles would be useful in understanding the interaction between immunity and these chemicals. Likewise, further study comparing the microbiomes between *D. sechellia* and *D. simulans* and how they may change from feeding on *M. citrifolia* fruit as well as how each responds to pathogen challenge in this context would help to understand why *D. sechellia* is downregulating its immune response when there could be a serious fitness cost of this action.

Recent studies have shown that *D. melanogaster* uses sweet-tasting gustatory receptor neurons to sense OA and HA (Masek and Keene 2013; Chen and Amrein 2017; Tauber et al. 2017). Our data show that upon exposure to HA, only *D. sechellia* significantly upregulates another odorant-binding protein, *Obp56a*. In contrast to HA treatment, in response to OA *D. sechellia* adults do not change expression of any odorant-binding proteins (Lanno et al. 2017). Both *D. simulans* and *D. melanogaster* downregulate the expression of *Ovp57a* in response to HA (Supplementary Table S10), and *D. sechellia* upregulates the expression of *Obp56a* in response to HA (Supplementary Table S13). Recently, studies examining the gustatory and behavioral basis of the attraction of *D. sechellia* to noni have shown that *Or22a* neurons in the fly brain are involved in the attraction to noni (Auer et al. 2020), as are *Or85b/c* and *Ir75b* neurons (Prieto-Godino et al., 2017). *Orco* mutant *D. sechellia* flies lose olfactory responses to both 2-heptanone and 1-hexanol, two compounds found in ripe noni fruit (Auer et al. 2020). Further work on the role of these genes in *D. sechellia* and the regulatory mechanisms responsible for the change of odorant-binding protein expression in response to HA may help to elucidate how *D. sechellia* has evolved to specialize on *M. citrifolia*.

Drosophila sechellia is altering the expression of far fewer genes in response to HA than in the generalist species *D. melanogaster* and *D. simulans* (Supplementary Tables S1–S3). *Drosophila sechellia* appears to be downregulating many genes involved in the immune response whereas *D. melanogaster* and *D. simulans* alter the expression of genes involved in many different metabolic pathways (Figure 4, A–C). As generalist insects feed on many different plants which produce different secondary metabolites to defend themselves from predators, perhaps generalist species have a more plastic regulatory response to subvert toxicity whereas the specialist *D. sechellia* is resistant to the toxicity of its host through a specific constitutive mechanism. A recent study comparing the fitness of *D. melanogaster* to *D. sechellia* larvae fed different food sources showed a loss of carbohydrate metabolic responses in *D. sechellia*, as they have specialized on a fruit with a relatively low sugar content, *M. citrifolia* (Watanabe et al. 2019). Of the many significant genes differentially expressed in *D. melanogaster* and *D. simulans*, most DEGs are not drastically differentially expressed in HA treatment compared to controls (Figure 2, D and E). *Drosophila sechellia* conversely alters the expression of relatively far fewer genes, but many of the DEGs in response to HA are drastically differentially expressed (Figure 2F). Similarly, another study found that when adapted to a grass diet, *Spodoptera littoralis* had a smaller transcriptional response when fed maize compared to more generalist *S. littoralis* (Roy et al. 2016). Determining the scale of these responsive regulatory effects and their role in toxin

resistance will help elucidate how *D. sechellia* has evolved to avoid the toxicity of *M. citrifolia*.

In response to OA, *D. sechellia* increases its expression of several *Osiris* genes (Lanno et al. 2017). Previous work using RNAi in *D. melanogaster* to knock-down the expression of individual genes and examine survival in these flies when exposed to OA showed that the reduction of *Osi6*, *Osi7*, and *Osi8* expression decreased survival (Andrade Lopez et al. 2017). In response to HA, neither *D. sechellia* nor *D. melanogaster* significantly alter the expression of any *Osiris* genes, but *D. simulans* downregulates the expression of *Osi6*, *Osi7*, and *Osi15* (Supplementary Table S2). The cellular and physiological function of these genes is unknown, so understanding what these genes are doing and how they may be helping to shape the interactions between these insects and their toxic hosts will be useful to better understand how these interactions evolve (Coolon et al. 2019).

Examining and comparing the changes in transcriptional output of insects when exposed to these different plant chemicals in specialized vs generalist species provides a framework to understand how these interactions have evolved. Pathway analyses of these genes are useful in determining the physiological function of altered expression in response to these chemicals, but functional analyses of transcription factors that alter expression of these genes are necessary to better understand the regulatory mechanisms involved in the dietary specialization. Comparing transcription factors responding to plant chemicals may help elucidate regulatory mechanisms involved in these responses and shed light on how insects use changes in the transcription of target genes in order to compete against plants in this evolutionary arms race to adapt to toxic food sources.

Data availability

All RNA-seq data generated in this manuscript have been submitted to the NCBI Gene Expression Omnibus under accession number GSE185176. Supplementary tables for this manuscript have been uploaded to GSA figshare: <https://doi.org/10.25387/g3.14745684>.

Funding

Research reported in this publication was supported by Wesleyan University (Startup to J.D.C., Department of Biology funds to JDC, College of the Environment funds to J.D.C.), and the National Institute of General Medical Sciences of the National Institutes of Health under Award Number R15GM135901 (award to J.D.C.). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Conflicts of interest

The authors declare that there is no conflict of interest.

Literature cited

Afgan E, Baker D, van den Beek M, Blankenberg D, Bouvier D, et al. 2016. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic Acids Res.* 44:W3–W10. doi:10.1093/nar/gkw343.

- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. BLAST_article.pdf. *J Mol Biol.* 215:403–410. doi:10.1016/S0022-2836(05)80360-2.
- Alvarez-Curto E, Milligan G. 2016. Metabolism meets immunity: the role of free fatty acid receptors in the immune system. *Biochem Pharmacol.* 114:3–13. doi:10.1016/j.bcp.2016.03.017.
- Amlou M, Moreteau B, David JR. 1998. Genetic analysis of *Drosophila sechellia* specialization: Oviposition behavior toward the major aliphatic acids of its host plant. *Behav Genet.* 28:455–464. doi:10.1023/A:1021689312582.
- Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. <http://www.Bioinformatics.Babraham.Ac.Uk/Projects/Fastqc>.
- Aranega-Bou P, de la O Leyva M, Finiti I, García-Agustín P, González-Bosch C. 2014. Priming of plant resistance by natural compounds. Hexanoic acid as a model. *Front Plant Sci.* 5:488. doi:10.3389/fpls.2014.00488.
- Ashburner M, Ball CA, Blake J, Botstein D, Butler H, et al. 2000. Gene ontology: tool for the unification of biology. *Nat Genet.* 25:25–29. doi:10.1038/75556.
- Auer TO, Khallaf MA, Silbering AF, Zappia G, Ellis K, et al. 2020. Olfactory receptor and circuit evolution promote host specialization. *Nature.* 579:402–408. doi:10.1038/s41586-020-2073-7.
- Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Statist Soc.* 57:289–300.
- Carbon S, Douglass E, Good BM, Unni DR, Harris NL, et al. 2021. The gene ontology resource: enriching a GOLD mine. *Nucleic Acids Res.* 49:D325–D334. doi:10.1093/nar/gkaa1113.
- Chen Y, Amrein H. 2017. Ionotropic receptors mediate *Drosophila* oviposition preference through sour gustatory receptor neurons. *Curr Biol.* 27:2741–2750.e4. doi:10.1016/j.cub.2017.08.003.
- Coolon JD, Drum Z, Lanno SM, Smith CR. 2019. Osiris genes and insect adaptive evolution. In: eLS. Chichester: John Wiley & Sons, Ltd. 1–14. doi:10.1002/9780470015902.a0028763.
- Coolon JD, Jones KL, Todd TC, Carr BC, Herman MA. 2009. *Caenorhabditis elegans* genomic response to soil bacteria predicts environment-specific genetic effects on life history traits. *PLoS Genet.* 5:e1000503. doi:10.1371/journal.pgen.1000503.
- Coolon JD, Webb W, Wittkopp PJ. 2013. Sex-specific effects of cis-regulatory variants in *Drosophila melanogaster*. *Genetics.* 195:1419–1422. doi:10.1534/genetics.113.156331.
- Dushay MS, Åsling B, Hultmark D. 1996. Origins of immunity: relish, a compound rel-like gene in the antibacterial defense of *Drosophila*. *Proc Natl Acad Sci USA.* 93:10343–10347. doi:10.1073/pnas.93.19.10343.
- Dworkin I, Jones CD. 2009. Genetic changes accompanying the evolution of host specialization in *Drosophila sechellia*. *Genetics.* 181:721–736. doi:10.1534/genetics.108.093419.
- Farine J, Legal LUC, Moreteau B, I, JLEQ, Laboratoire I, Recherches, et al. 1996. Volatile components of ripe fruits of *Morinda citrifolia* and their effects on *Drosophila*. *Phytochemistry.* 4:433–438. doi:10.1016/0031-9422(95)00455-6.
- García-Robles I, Ochoa-Campuzano C, Fernández-Crespo E, Camañes G, Martínez-Ramírez AC, et al. 2013. Combining hexanoic acid plant priming with *Bacillus thuringiensis* insecticidal activity against Colorado potato beetle. *Int J Mol Sci.* 14:12138–12156. doi:10.3390/ijms140612138.
- Groen SC, Whiteman NK. 2016. Using *Drosophila* to study the evolution of herbivory and diet specialization. *Curr Opin Insect Sci.* 14:66–72. doi:10.1016/j.cois.2016.01.004.
- Hanson MA, Lemaitre B. 2020. New insights on *Drosophila* antimicrobial peptide function in host defense and beyond. *Curr Opin Immunol.* 62:22–30. doi:10.1016/j.coi.2019.11.008.
- Heidel-Fischer HM, Vogel H. 2015. Molecular mechanisms of insect adaptation to plant secondary compounds. *Curr Opin Insect Sci.* 8:8–14. doi:10.1016/j.cois.2015.02.004.
- Jaenike J. 1990. Host specialization in phytophagous insects. *Ann Rev Ecol Syst.* 21:243–273. doi:10.1146/annurev.es.21.110190.001331.
- Kanehisa M, Sato Y. 2020. KEGG Mapper for inferring cellular functions from protein sequences. *Protein Sci.* 29:28–35. doi:10.1002/pro.3711.
- Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods.* 9:357–359. doi:10.1038/nmeth.1923.
- Lanno SM, Coolon JD. 2019. Derived esterase activity in *Drosophila sechellia* contributes to evolved octanoic acid resistance. *Insect Mol Biol.* 28:798–806. doi:10.1111/imb.12587.
- Lanno SM, Gregory SM, Shimshak SJ, Alverson MK, Chiu K, et al. 2017. Transcriptomic Analysis of octanoic acid response in *Drosophila sechellia* using RNA-sequencing. *G3 (Bethesda).* 7:g3.300297.2017. doi:10.1534/g3.117.300297.
- Lanno SM, Linde SC, Peyser RD, Shimshak SJ, Coolon JD. 2019. Investigating the role of Osiris genes in *Drosophila sechellia* larval resistance to a host plant toxin. *Ecol Evol.* 9:1922–1933. doi:10.1002/ece3.4885.
- Lavista-Llanos S, Svatoš A, Kai M, Riemensperger T, Birman S, et al. 2014. Dopamine drives *Drosophila sechellia* adaptation to its toxic host. *Elife.* 3:e03785. doi:10.7554/eLife.03785.
- Legal L, Chappe B, Jallon JM. 1994. Molecular basis of *Morinda citrifolia* (L.): toxicity on *Drosophila*. *J Chem Ecol.* 20:1931–1943.
- Legal L, David JR, Jallon JM. 1992. Toxicity and attraction effects produced by *Morinda citrifolia* fruits on the *Drosophila melanogaster* complex of species. *Chemoecology.* 3:125–129.
- Login FH, Login FH, Balmand S, Balmand S, Vallier A, et al. 2011. Antimicrobial peptides keep insect Endosymbionts under control. *Science.* 334:362–365. <http://www.sciencemag.org/cgi/doi/10.1126/science.1209728%5Cnpapers2://publication/doi/10.1126/science.1209728>.
- Lopez A, Lanno JM, Auerbach SM, Moskowitz JM, Sligar EC, et al. 2017. Genetic basis of octanoic acid resistance in *Drosophila sechellia*: functional analysis of a fine-mapped region. *Mol Ecol.* 26:1148–1160. doi:10.1111/mec.14001.
- Masek P, Keene AC. 2013. *Drosophila* fatty acid taste signals through the PLC pathway in sugar-sensing neurons. *PLoS Genet.* 9:e1003710. doi:10.1371/journal.pgen.1003710.
- Matsuo T, Sugaya S, Yasukawa J, Aigaki T, Fuyama Y. 2007. Odorant-binding proteins OBP57d and OBP57e affect taste perception and host-plant preference in *Drosophila sechellia*. *PLoS Biol.* 5:0985–0996. doi:10.1371/journal.pbio.0050118.
- Mi H, Muruganujan A, Ebert D, Huang X, Thomas PD. 2019. PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Res.* 47:D419–D426. doi:10.1093/nar/gky1038.
- Petschenka G, Agrawal AA. 2016. How herbivores coopt plant defenses: Natural selection, specialization, and sequestration. *Curr Opin Insect Sci.* 14:17–24. doi:10.1016/j.cois.2015.12.004.
- Peyser RD, Lanno SM, Shimshak SJ, Coolon JD. 2017. Analysis of cytochrome P450 contribution to evolved plant toxin resistance in *Drosophila sechellia*. *Insect Mol Biol.* 26:715–720. doi:10.1111/imb.12329.
- Prieto-Godino LL, Rytz R, Cruchet S, Bargeton B, Abuin L, et al. 2017. Evolution of Acid-Sensing Olfactory Circuits in *Drosophilids*. *Neuron.* 93:661–676.e6. doi:10.1016/j.neuron.2016.12.024.28111079.

- R Core Development Team. 2010. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. ISBN:3-900051-07-0.
- R’Kha S, Moreteau B, Coyne JA, David JR. 1997. Evolution of a lesser fitness trait: egg production in the specialist *Drosophila sechellia*. *Genet Res.* 69:17–23. doi:10.1017/s0016672396002546 9164172
- R’Kha S, Capy P, David JR. 1991. Host-plant specialization in the *Drosophila melanogaster* species complex: a physiological, behavioral, and genetical analysis. *Proc Natl Acad Sci USA.* 88:1835–1839. doi:10.1073/pnas.88.5.1835.
- Roy A, Walker WB, Vogel H, Chattington S, Larsson MC, et al. 2016. Diet dependent metabolic responses in three generalist insect herbivores Spodoptera spp. *Insect Biochem Mol Biol.* 71:91–105. doi:10.1016/j.ibmb.2016.02.006.
- Salazar-Jaramillo L, Jalvingh KM, de Haan A, Kraaijeveld K, Buermans H, et al. 2017. Inter- and intra-species variation in genome-wide gene expression of *Drosophila* in response to parasitoid wasp attack. *BMC Genomics.* 18:1–14. doi:10.1186/s12864-017-3697-3.
- Shokri L, Inukai S, Hafner A, Weinand K, Hens K, et al. 2019. A comprehensive *Drosophila melanogaster* transcription factor interactome. *Cell Rep.* 27:955–970.e7. doi:10.1016/j.celrep.2019.03.071.
- Stork NE. 2018. How many species of insects and other terrestrial arthropods are there on earth? *Annu Rev Entomol.* 63:31–45. doi:10.1146/annurev-ento-020117-043348.
- Tauber JM, Brown EB, Li Y, Yurgel ME, Masek P, et al. 2017. A subset of sweet-sensing neurons identified by IR56d are necessary and sufficient for fatty acid taste. *PLoS Genet.* 13:1–18. doi:10.1371/journal.pgen.1007059.
- Thurmond J, Goodman JL, Strelets VB, Attrill H, Gramates LS, et al.; FlyBase Consortium. 2019. FlyBase 2.0: the next generation. *Nucleic Acids Res.* 47:D759–D765. doi:10.1093/nar/gky1003.
- Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, et al. 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol.* 28:511–515. doi:10.1038/nbt.1621.
- Ungerer MC, Johnson LC, Herman MA. 2008. Ecological genomics: understanding gene and genome function in the natural environment. *Heredity (Edinb).* 100:178–183. doi:10.1038/sj.hdy.6800992.
- Vilcinskas A. 2013. Evolutionary plasticity of insect immunity. *J Insect Physiol.* 59:123–129. doi:10.1016/j.jinsphys.2012.08.018
- Watanabe K, Kanaoka Y, Mizutani S, Uchiyama H, Yajima S, et al. 2019. Interspecies comparative analyses reveal distinct carbohydrate-responsive systems among *Drosophila* species. *Cell Rep.* 28:2594–2607.e7. doi:10.1016/j.celrep.2019.08.030.
- Yates AD, Achuthan P, Akanni W, Allen J, Allen J, et al. 2020. Ensembl 2020. *Nucleic Acids Res.* 48:D682–D688. doi:10.1093/nar/gkz966 31691826.

Communicating editor: J. Tennesen