UC Berkeley
UC Berkeley Electronic Theses and Dissertations

Title

The Evolution and Genetic Basis of Adaptations for Feeding and Egg-laying in Herbivorous Drosophilids

Permalink

https://escholarship.org/uc/item/5rx3t8wb

Author

Pelaez, Julianne Naomi

Publication Date

2022

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

The Evolution and Genetic Basis of Adaptations for Feeding and Egg-laying in Herbivorous Drosophilids

By

Julianne N. Peláez

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Integrative Biology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Noah Whiteman, Chair
Professor Doris Bachtrog
Professor Kristin Scott
Professor Peter Sudmant

Fall 2022

Abstract

The Evolution and Genetic Basis of Adaptations
for Feeding and Egg-laying in Herbivorous Drosophilids

by

Julianne N Peláez

Doctor of Philosophy in Integrative Biology

University of California, Berkeley

Professor Noah K. Whiteman, Chair

The evolution of herbivory in insect lineages has generated one of the most diverse groups of organisms on Earth. Adapting to an herbivorous lifestyle, however, comes with significant challenges for insects, as their host plants mount substantial defenses. Herbivorous insects must contend with the plant's first line of defense – a battery of physical barriers (i.e., spines, trichomes, waxy and thickened cuticles) and their second line of defense – the mass production of chemical cocktails aimed at killing and/or repelling them. The mechanisms by which insects evade and overcome these plant defenses have captivated the attention of researchers for over a century. But only recently have we begun to study the genetic and molecular mechanisms underlying these adaptations. In my dissertation, I investigate the evolution and genetic basis of key adaptations that facilitate the penetration of tough leaf surfaces, and adaptions that enable changes in chemosensory reception towards and metabolism of plant toxins.

My work focuses on the herbivorous lineage within *Scaptomyza*, a genus that is sister to Hawaiian *Drosophila* and nested within the paraphyletic genus of *Drosophila*. While close relatives feed on decaying plant tissues and microbes, most species of herbivorous *Scaptomyza* feed on plants of the order Brassicales (mustards and allies). *Scaptomyza* have become models for studying herbivory because of their recent evolution of herbivorous feeding (ca. 10-15 million years ago), in addition to their close evolutionary proximity to *Drosophila melanogaster*, which affords a wealth of genetic and genomic tools.

In Chapter 1, I investigate the evolution and genomic architecture of the plant-penetrating ovipositor in herbivorous *Scaptomyza*. Females use their dentate ovipositors to gain access to physically-defended leaves for feeding and egg-laying. This morphological trait has evolved several times in the Drosophilidae, in most cases involving an increase in hardened bristles along the margins. I tested the hypothesis that increased ovipositor bristle number was associated with herbivory in a comparative

phylogenetic framework and found a significant increase coinciding with the evolution of herbivory in *Scaptomyza*. We then conducted a genome wide association study on ovipositor bristle number in *S. flava.* We found that variation in ovipositor bristle number was most notably associated with genetic variants that were located near or in transcriptional repressors affecting bristle development. These findings suggests that increased ovipositor bristle number, a key trait for the cutting ovipositor, may have been driven by regulatory evolution on conserved developmental genes.

In Chapter 2, I examine the molecular evolution of major detoxification and chemosensory gene families. Because these genes directly mediate interactions with plant chemical defenses, I aimed to test the hypothesis that herbivory was associated with dramatic gene losses, duplications, and significant protein evolution, and to identify specific gene candidates involved in this dietary transition within *Scaptomyza.* We annotated these genes in three herbivorous and nine non-herbivorous drosophilid species. I then generated phylogenetic trees for each gene family and estimated evolutionary rates of gene loss and duplication. Indeed, I found that the branch leading to all three herbivorous *Scaptomyza* experienced significantly higher rates of gene loss in almost all chemosensory and detoxification gene families, and significantly higher rates of gene duplication in more than half. I then assessed changes in selective pressure, employing several codon-based models of protein evolution. I found that both changes in gene copy number and changes in selection pressure most strongly impacted chemosensory genes involved in the detection of yeast volatiles and bitter phytotoxins. These results add to the growing literature that indicates that herbivory is enabled by the losing the ability to detect attractive ancestral cues (e.g., yeast volatiles) and repellent host plant cues. Among detoxification genes, gene losses, duplications and positive selection were notably found in genes involved in oxidative stress responses, insecticide resistance, and phytotoxin metabolism. This study serves as a critical first step in identifying specific gene candidates that underlie behavioral adaptations that drive herbivorous insects towards appropriate food sources, and physiological adaptations that enable them to metabolize host plant toxins.

In Chapter 3, I investigate how the taste system of *S. flava* has evolved to deal with plant-derived bitter chemical defenses. Generalist insects are often repelled by or averse to bitter plant toxins. Indeed, the results from Chapter 2 indicated that herbivorous *Scaptomyza* may have lost their ability to detect bitter toxins, through the loss of gustatory receptors involved in bitter detection. This would enable them to feed undeterred on their new hosts. However, many other specialist herbivores use these toxins to identify their host plants. I therefore hypothesized that *S. flava* may have also evolved gustatory attraction towards mustard-specific defense compounds (specifically glucosinolates), while losing perception of bitter compounds derived from non-mustard plants. In feeding assays, I found that female *S. flava*, contrary to expectations, were less averse not only to a non-mustard compound (caffeine) but also to mustard-derived glucosinolates. This contrasted with non-herbivorous *Scaptomyza*, which were strongly averse to all bitter compounds. I then performed electrophysiological recordings on the labellar taste sensilla to test whether these behavioral changes in bitter preferences could be explained by changes in sensitivity in the peripheral nervous system. I found that one class of sensilla

in *S. flava* (I type) had reduced sensitivity to glucosinolates, compared to homologous sensilla in the non-herbivore *S. pallida*. Using existing knowledge of the expression patterns of gustatory receptors in *D. melanogaster*, coupled with the genetic changes in gustatory receptors identified in Chapter 2, these results suggest a critical role for losses of candidate gustatory receptors (paralogs of *Gr39aA* and *Gr28bA*) in driving overall changes in bitter detection. While it remains unclear the extent to which young herbivores evolve gustatory attraction towards plant defense compounds, the *S. flava* system showcases the importance of reduced aversion to bitter defense compounds, and how this may be achieved.

Dedication

To Cha, my mother,
for instilling in me an appreciation for the natural world.

And to George, my partner,
for his endless support.

Table of Contents

Dedication i
List of Figures, Tables, and Appendices iii
Acknowledgements vi

Chapter 1. Evolution and genomic basis of the plant-penetrating ovipositor: a key morphological trait in herbivorous Drosophilidae

Abstract 2
Introduction 2
Materials and Methods 5
Results 8
Discussion 12

Chapter 2. Comparative genomics of chemosensory and detoxification gene family evolution across herbivorous Drosophila

Abstract 16
Introduction 16
Materials and Methods 20
Results 24
Discussion 35

Chapter 3. Taste evolution in an herbivorous drosophilid

Abstract 43
Introduction 43
Materials and Methods 46
Results 48
Discussion 56

References 61

Appendices 81

List of Figures, Tables, and Appendices

Chapter 1

Table 1 11
Top SNPs significantly associated with variation in ovipositor bristle number are located in or near genes involved in the development of bristles, cuticles, and the nervous system.

Table 2 12
Gene ontology (GO) terms enriched among genes intersecting the most significant GWAS SNPs are transcriptional repressors, phosphatidylinositol biosynthetic process and establishment of cell polarity.

Figure 1 4
Female ovipositor morphology of the herbivorous drosophilid *Scaptomyza flava* enables cutting into tough plant tissues for feeding and egg-laying. (a) The *S. flava* life cycle. (b) Light microscopy of drosophilid ovipositors. (c) Scanning electron micrographs of the ovipositor of *S. flava*.

Figure 2 8
The evolution of herbivory within *Scaptomyza* coincides with an increase in ovipositor sensilla number. (a) Time-calibrated phylogeny of herbivorous *Scaptomyza* and their non-herbivorous relatives. (b) Individual counts of sensilla numbers for extant species. (c) Scatterplot of ovipositor sensilla number as a function of ovipositor length.

Figure 3 9
Variation in the number of plant-cutting ovipositor bristles shows a continuous distribution and heritable in *S. flava*, enabling quantitative genetic dissection. (a) Diagrams of ovipositor bristle counting and wing chord length measurements. (b) Distribution of ovipositor bristle number. (c) Narrow sense heritability analysis from mother-daughter regression. (d, e) Bristle counts and ovipositor lengths of low and high bristle pools.

Figure 4 10
Pool-GWAS for variation in *S. flava* ovipositor bristle number implicates genes involved in nervous system development. (a) Distribution of *P*-values from pool-GWAS. (b) The relationship between physical distance and linkage disequilibrium in *S. flava*. (c) Manhattan plot of GWAS results. (d) Individual phenotyping and genotyping for the SNP near *Gαi*.

Appendix A: Supplemental methods from Chapter 1 81

Appendix B: Supplemental tables from Chapter 1 87

Appendix C: Supplemental figures from Chapter 1 107

Chapter 2

Table 1 29
Summary of selection analyses on chemosensory genes.

Table 2 31
Summary of selection analyses on detoxification genes.

Table 3 33
Gene functions of candidate chemosensory and detoxification genes duplicated, lost, or having experienced positive selection in the ancestral lineage of herbivorous *Scaptomyza*.

Figure 1 18
Gene families at the interface of plant-herbivore interactions. (a) Phylogeny of herbivorous *Scaptomyza* and *Drosophila*. (b) COI Phylogeny of North American *Scaptomyza.* (c) Detoxification gene families. (d) Chemosensory gene families.

Figure 2 26
Herbivorous drosophilid lineages exhibit elevated gene turnover within most chemosensory and detoxification gene families. (a) Gene counts per gene family. Estimates of evolutionary rates of (b) gene turnover, (c) gene loss, and (d) gene duplication.

Figure 3 28
Chemosensory and detoxification genes duplicated, lost, or that have experienced significant changes in selection pressure in the branch at the base of the herbivorous *Scaptomyza* lineage. (a) Summary of genetic changes affecting individual genes. (b) Localization map of chemosensory proteins that underwent or are undergoing significant evolutionary change.

Appendix D: Supplemental methods from Chapter 2 113

Appendix E: Supplemental tables from Chapter 2 118

Appendix F: Supplemental figures from Chapter 2 131

Chapter 3

Table 1 47
Chemical compounds used in feeding experiments and electrophysiology

Figure 1 50
Scaptomyza flava shows reduced bitter aversion to some plant-derived compounds. (a) Phylogeny of *S. flava* and microbe-feeding drosophilids. (b) Schematic of the two-choice feeding assay. Feeding preferences towards (c) non-mustard bitters and (d) glucosinolates.

Figure 2 51
Labellar sensilla types are conserved across *D. melanogaster, S. flava*, and *S. pallida*, with a few changes consistent in both non-herbivorous and herbivorous *Scaptomyza*. (a) Scanning electron microscopy (SEM) of labella. (b) Illustration of sensilla types based on SEM.

Figure 3 53
Identification of labellar sensilla classes based on response profiles. (a) Heatmap of spike responses to caffeine, lobeline, sucrose. (b) Illustration of representative sensilla recorded from in c. (c) Dose response curves for caffeine, lobeline, and sucrose. (d) Representative spike recordings from S. flava S5 sensilla.

Figure 4 55
Glucosinolate sensitivity across labellar sensilla class. (a) Spike response to increasing concentrations of three glucosinolates. (b) Responses to a mixed solution of sucrose and sinigrin. (c) Responses to six glucosinolates (10 mM). (d) Heatplot of spike responses towards all glucosinolates (10 mM).

Figure 5 59
Candidate changes in gustatory receptors in the herbivore Scaptomyza flava.

Appendix G: Supplemental methods from Chapter 3..... 144

Appendix H: Supplemental tables from Chapter 3..... 145

Appendix I: Supplemental figures from Chapter 3 154

Acknowledgements

First and foremost, I would like to thank Noah Whiteman for his invaluable mentorship and support throughout my dissertation. I appreciate that he allowed me the freedom to pursue my own research interests. This gave me the opportunity to stumble and waffle, and I appreciate how this encouraged me to recover better from failures and to stand more confidently in my work. All the while, I am grateful for his encouragement to collaborate within and outside the lab, as this taught me knew skills and provided me with many wonderful role models and mentors.

I thank my committee members, Kristin Scott, Peter Sudmant, and Doris Bachtrog. Their feedback and support have been especially valuable in shaping my research trajectory.

I am grateful to my lab mates for their scientific support and friendship throughout my PhD. I thank Kirsten Verster, Jessica Aguilar, Hiro Suzuki, Teru Matsunaga, Diler Haji, Rebecca Tarnopol, Nicolas Alexandre, Carolina Reisenman, Susan Bernstein, Cathy Rushworth, Tim O'Connor, Ben Goldman-Huertas, Erika LaPlante, Rebecca Duncan, and Marianna Karageorgi. I also thank the many undergraduate students I have had the privilege to work with. This includes Judith Okoro, Eunice Kim, Rachel Patterson, Malvika Singhal, Shirley Zhang, Esteban Rodas, Irene Liang, Anna Leipertz, Jiarui Wang, Giovani Pimentel-Solorio, and Haarini Sidhar.

I would like to thank my collaborators — Andy Gloss, Julie Ray, Danny Miller, Bernard Kim, Aya Takahashi, Frederic Marion-Poll, Beatrice Denis, and Penelope Tarnapacki — for their support on various projects. I especially thank Fred for the many scientific discussions that shaped my research and me as a scientist.

This work was supported by grants from the Society for the Study of Evolution (Rosemary Grant), Sigma Xi (Grant-In-Aid of Research), the UC Berkeley France-Berkeley Fund, and the UC Berkeley Integrative Biology department. Much of this research was also funded by a grant from the National Institute of General Medical Sciences of the National Institutes of Health to Noah Whiteman (R35GM119816). Several fellowships also provided financial support, including the National Science Foundation Graduate Research Fellowship, UC Berkeley Mentored Research Award, and the Chateaubriand Fellowship from the French Embassy in the US.

I thank the UC Berkeley research facilities that provided resources to carry out these research endeavors: the Electron Microscope Lab, the Berkeley Imaging Facility, the Molecular Imaging Center, Jacobs Hall, and the D-Lab. I am also grateful to Peter Oboyski and Igor Kligman for their help maintaining facilities for fly rearing and plant growing.

I thank my partner George Plummer, my sister Gina Pelaez, and my mother Hwa Cha Kwon. I cannot imagine how I would have fared without their support and encouragement. I also thank them for making many happy memories around the west coast with me. I would also like to express appreciation to my dear friend Paige Qin and the friends I made through unionizing, all of whom reminded me of the joys of life outside the lab.

In Chapter 1, I explore the evolution and genetic basis of a morphological trait that provides access to the physically defended interiors of the plant: the plant-penetrating ovipositor. Tough plant tissues pose a significant barrier to the evolution of herbivory, and different insect lineages overcome this hurdle through various means. Some herbivorous groups feed on tough plant tissue with chewing mouthparts in the larval stage, while others avoid these structural barriers with piercing/sucking mouthparts. Leaf mining enables herbivorous larvae to feed solely on the soft tissues in the interior of the plant, and additionally provides protection from predators/parasites and abiotic factors. It, however, first requires that the larvae or eggs gain access to the plant interior. Several independent herbivorous lineages in the Drosophilidae have evolved cutting ovipositors that females use to create wounds that they can deposit eggs into. Many of these lineages specifically share an increase in the number of sensilla lining the edge of the ovipositors. By studying the evolution and genetic basis of this increase in ovipositor bristle number, I sought to understand (1) whether variation in this trait *between* species was associated with herbivory, and (2) whether variation in this trait *within* species was associated with specific genetic variants. The intended outcome of this study was to understand how this could have evolved numerous times in independent herbivore lineages.

Chapter 1. Evolution and genomic basis of the plant-penetrating ovipositor: a key morphological trait in herbivorous Drosophilidae

This chapter has been accepted for publication and is reproduced here with permission:
Peláez J.N., Gloss A.D., Ray J.F., Chaturvedi S., Haji D., Charboneau J.L.M., Verster K.I., and Whiteman N.K. (in press). Evolution and genomic basis of the plant-penetrating ovipositor: a key morphological trait in herbivorous Drosophilidae. *Proceedings of the Royal Society B*.

Abstract

Herbivorous insects are extraordinarily diverse, yet are found in only one-third of insect orders. This skew may result from barriers to plant colonization, coupled with phylogenetic constraint on plant-colonizing adaptations. The plant-penetrating ovipositor, however, is one trait that surmounts physical barriers of entry into the plant and may be evolutionarily labile. Despite their significance, the evolution and genetic basis of such an innovation has not been well studied. Ovipositors densely lined with hard bristles have evolved repeatedly in herbivorous lineages within the Drosophilidae. Here, we focus on the evolution of this trait in *Scaptomyza*, a genus sister to Hawaiian *Drosophila*, which contains an herbivorous clade nested among microbe-feeders. Our phylogenetic approach revealed that ovipositor bristle number increased as herbivory evolved in the *Scaptomyza* lineage. We then dissected the genomic architecture of variation in ovipositor bristle number within *S. flava* through a genome wide association study. Top associated variants were enriched for transcriptional repressors, and the strongest associations included genes contributing to peripheral nervous system development. Genotyping individual flies supported the association at a variant upstream of *Gαi*, a neural development gene, contributing to a gain of 0.58 bristles/major allele. These results suggest that regulatory variation involving conserved developmental genes contributes to this key morphological trait involved in plant colonization.

Keywords: *Drosophila*, *Scaptomyza,* quantitative genetics, adaptation, innovation, herbivory

Introduction

Herbivorous insects are among the most successful animal radiations, representing approximately one quarter of animal species, yet are only found in one-third of extant insect orders (Mitter et al., 1988; Schoonhoven et al., 2005; Wiens et al., 2015), suggesting phylogenetic constraint on adaptations required for this transition. Indeed, herbivory requires multi-faceted adaptations, including: locating appropriate host plants, attachment to the host, resisting desiccation, and feeding on nutritionally unbalanced, chemically- and physically-defended plant tissues (Southwood, 1972). Despite the low number of insect orders with herbivorous species, herbivory has evolved many times independently within some orders (Wiens et al., 2015), including at least 25 times within

Diptera (Wiegmann et al., 2011). Identifying whether these clades share specific traits may help resolve the paradox of why herbivory has only evolved in some orders despite often leading to species radiations.

The plant-penetrating ovipositor is one such trait that facilitates entry into this new ecological niche and has driven species radiations. It evolved within major radiations of true fruit flies (Tephritidae), leaf-mining flies (Agromyzidae) and leafhoppers (Cicadellidae) – together comprising ~27,500 species – as well as within sawflies (Tenthredinidae), katydids (Tettigoniidae), and plant bugs (Miridae) (Eiseman et al., 2010). The insertion of eggs into plant tissue allows neonate larvae to bypass physical defenses and hatch directly into the leaf interior, avoiding desiccation and providing protection from the environment and enemies (Connor & Taverner, 1997; Southwood, 1972). Some insects with plant-penetrating ovipositors, like agromyzid flies, also consume leaf exudates from oviposition wounds (Marino & Cornell, 1992), providing a novel trophic resource to adults, even in the absence of chewing mouthparts.

The Drosophilidae is a compelling species radiation for studying plant-penetrating ovipositors as a key morphological trait for the evolution of herbivory. While most drosophilid species feed on decaying plant tissues and microbes, plant-penetrating ovipositors are found in all known lineages that evolved herbivory independently: (1) *D. suzukii*, a generalist pest of ripe fruit (Atallah et al., 2014), (2) leaf-miners within the genus *Scaptomyza* (Drosophilidae), which includes the model herbivore *Scaptomyza flava*, a specialized pest of Brassicaceae crops (Fig. 1a) (Gloss et al., 2019), (3) *Scaptodrosophila notha*, a specialist of living bracken fern fronds (*Pteridium* spp.) (Thomson et al., 1982), and (4) leaf-mining species of the subgenus *Lordiphosa* (Okada & Sasakawa, 1956). All four lineages bear sclerotized ovipositors, studded with sharp, enlarged bristles used to pierce or scrape into living plants. Drosophilid flies have already been in use as models for the evolution of herbivory (Groen & Whiteman, 2016), and genetic dissection of herbivore-specific traits is enabled by the availability of high-quality genome assemblies across the genus (Kim et al., 2021) functional genetic data from *Drosophila melanogaster*, and a strong phylogenetic framework for Drosophilidae (Finet et al., 2021).

Although ovipositors of herbivorous drosophilids differ in many morphological aspects, one shared feature is a row of supernumerary bristles along the ventral and posterior margins used for cutting (e.g., Fig. 1b, c). Drosophilids possess several ovipositor sensilla types, including one long hypogynial sensilla and three trichoid short hypogynial sensilla located on the inner surface of the oviscapt apex (McQueen et al., 2022). On the outer surface are hair- or peg-like bristles, located ventrally, posteriorly, and/or laterally (also termed “hypogynial teeth” but hereafter referred to as “ovipositor bristles”) (Fig. 1c). While the first two sensilla types on the inner surface do not vary widely in number across species, ovipositor bristles vary significantly in number, shape, and location (Craddock et al., 2018). The function of these bristles in *Drosophila* is not fully understood, but in *D. suzukii* and *D. melanogaster*, ovipositor bristles harbor mechanosensory neurons, and are used to sense substrate hardness during egg-laying (Crava et al., 2020; Taylor, 1989). We thus hypothesize that during the evolution of herbivory, these mechanosensory bristles were co-opted into a cutting tool through the

increase of bristles, facilitating access to turgid, fibrous plant cells. We therefore centered this study on increased bristle number, which has been well-studied from a quantitative genetics and developmental biology perspective (Norga et al., 2003).

To investigate whether increased ovipositor bristles were associated with this transition to herbivory and to explore underlying molecular mechanisms of this increase, we focused on the cutting ovipositor of herbivorous *Scaptomyza*, particularly *S. flava*. In addition to morphological changes, *S. flava* has acquired a stereotyped behavioral repertoire for feeding and egg-laying: females tap their ovipositors around the leaf searching for an ideal site, then scoop a hole by repeatedly opening their oviscapts laterally, and finally turn counterclockwise to imbibe the leaf exudates (Supplemental Videos S1, S2). Because *S. flava* females create hundreds of punctures over their lifespan (Seraj, 2000) and spend a lengthy amount of time cutting each hole – roughly one minute versus a few seconds in microbe-feeders, such as *D. melanogaster* (Yang et al., 2008) – adaptations that reduce energy expenditure could be favored by natural that hatch outside do not survive (Gloss et al., 2019).

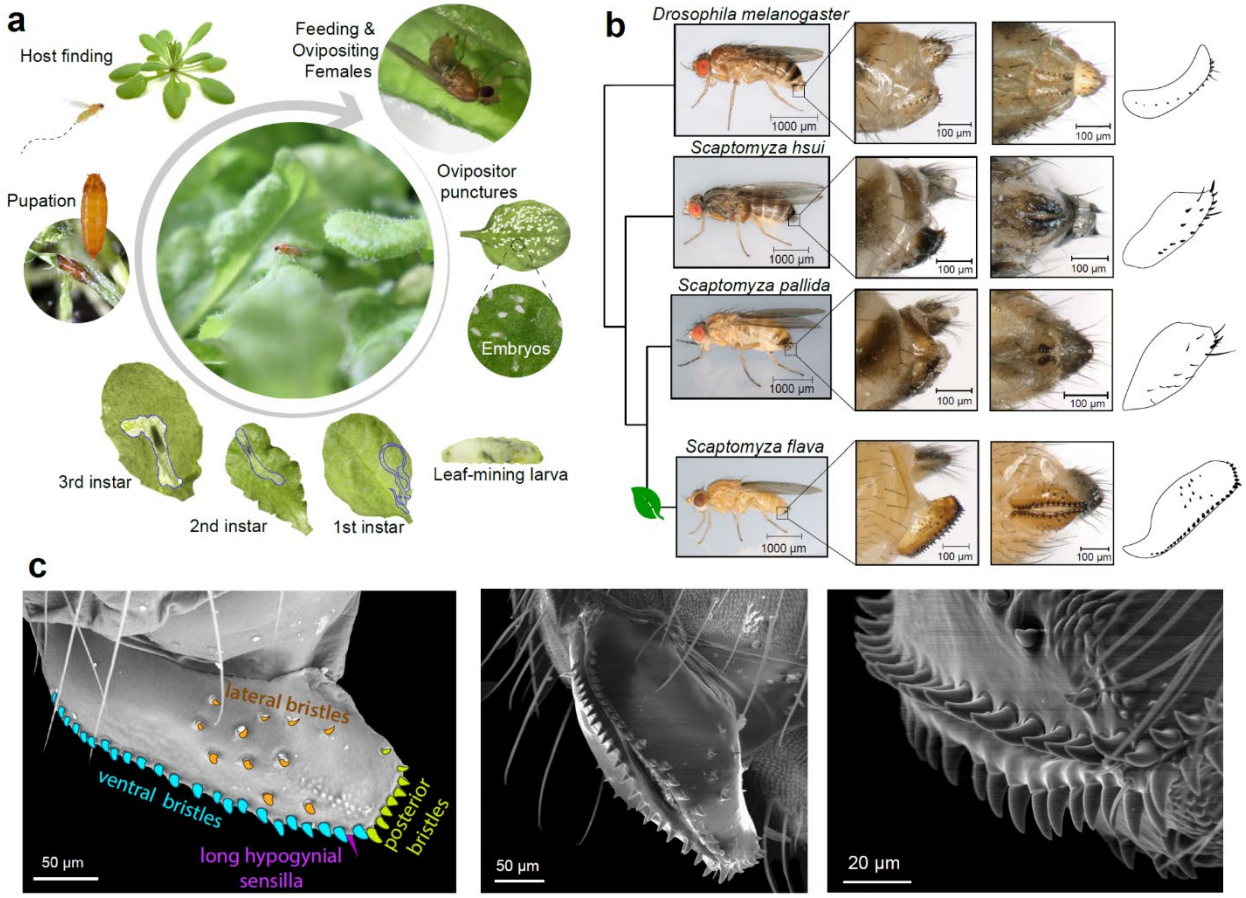


Figure 1. Female ovipositor morphology of the herbivorous drosophilid *Scaptomyza flava* enables cutting into tough plant tissues for feeding and egg-laying. (a) The *S. flava* life cycle is strongly dependent on accessing the leaf interior. On the underside of an *Arabidopsis thaliana* leaf, a female uses her serrated ovipositor to scoop a leaf puncture to drink from and/or to oviposit in. Larval mines outlined in blue. (b) Comparison of ovipositors (insets) of herbivorous and non-herbivorous drosophilid species. (c) Scanning electron micrographs of the ovipositor of *S. flava* (hypogynial short sensilla, not shown).

We first investigated whether ancestral increases in ovipositor bristle number paralleled the transition to herbivory in *Scaptomyza*, using phylogenetic generalized least squares (PGLS) methods and ancestral state reconstruction (ASR). Then, to uncover candidate genes and developmental pathways that underlie variation in ovipositor bristle number, we used pooled genome-wide association mapping (pool-GWAS) (Bastide et al., 2013; Sham et al., 2002) within the herbivorous species *S. flava*. Finally, we sought to confirm our pool-GWAS results by genotyping individuals and estimating the effect size of a single nucleotide polymorphism (SNP) that reached genome-wide significance.

Materials and Methods

(a) Phylogeny reconstruction.

We estimated a phylogeny of *Scaptomyza*, including the sister clade of Hawaiian *Drosophila*, using 11 genes and 95 taxa (Table S1). We expanded a previous phylogenetic dataset (Katoh et al., 2017) with five additional taxa: two with sequenced genetic markers (*S.* nr. *nigrita* (Nevada) and *S. montana* (Arizona) (Gloss et al., 2014)), and three obtained in this study from California (*S.* nr. *nigrita*, *S. montana*, and an undescribed species *S.* sp). DNA extraction and PCR methods have been described previously (Lapoint et al., 2013). PCR amplicons were cleaned, and Sanger sequenced in both directions, and trimmed and manually aligned to the other taxa using MAFFT v7.450. We estimated a species tree using the alignment of concatenated genes by maximum likelihood (ML) in RAxML (Stamatakis, 2006), and a time-calibrated tree by Bayesian inference using MrBayes v.3.2.4 (Ronquist & Huelsenbeck, 2003) and BEAST v.2.4.6 (Bouckaert et al., 2014). Alignment partitioning and model implementation are described in the Supplementary Methods. Complete phylogenies are reported in Fig. S1 and S2.

(b) Ovipositor trait evolution.

To test whether ovipositor bristle number changed significantly during the evolution of herbivory, we performed phylogenetic least squares (PGLS) regression (Butler & King, 2004), including the following predictor variables: larval diet (herbivorous versus non-herbivorous) (Lapoint et al., 2013), ovipositor length, phylogenetic relatedness, and source of bristle counts (literature versus this study). We collected bristle counts from illustrations or images from the literature, or directly from wild or lab-reared individuals (Table S2). Where available, we averaged across multiple individuals and literature sources. Because distinguishing between sensilla types from the literature was not always clear, we counted all visible bristles and hypogynial long sensilla on one oviscapt, omitting the inconspicuous hypogynial short sensilla. We obtained ovipositor lengths from literature sources either from published measurements or using provided scale bars. Ovipositors of wild and lab-reared flies (n = 2-10 per species) were mounted on slides with Permount mounting medium (Fisher Scientific) and coverslips and photographed using an EOS Rebel T3i camera (Canon) mounted onto a Stemi 508 stereo microscope (Zeiss) with a 1000 μm scale bar. Ovipositor length was then measured using ImageJ.

We performed PGLS regression using ape (Paradis et al., 2004) and picante (Kembel et al., 2010) packages in R. Comparing models of trait evolution (Brownian motion, Ornstein–Uhlenbeck, Early Burst, and white noise) for bristle number using AICc in the geiger R package (Harmon et al., 2008), we selected Brownian motion as the best fit (Table S3). The degree of phylogenetic signal in the residuals was estimated using Pagel's lambda (λ) (Pagel, 1999). To visualize correlated evolutionary changes in diet and bristle number, we mapped onto the phylogeny estimated ancestral states of both traits by ML using phytools (Revell, 2012) and ape (Paradis et al., 2004). We compared models of trait evolution (equal rates, symmetric, and all rates different) for larval diet, and identified equal rates as the best fit (Table S4).

(c) Mapping population and measurements for pool-GWAS.

To identify genetic polymorphisms contributing to variation in bristle number, we used a pool-GWAS to detect allele frequency differences between pools of individuals with extreme phenotypes from the same population. Two *S. flava* outbred laboratory populations were founded from collections on mustard plants, one larva per plant, near Dover, New Hampshire, USA (both within 0.1 km of 43.10068, -70.81246): one population ("NH1") was founded from 79 larvae from *Turritis glabra* and the second ("NH2") from 58 larvae from *T. glabra* and *Barbarea vulgaris*. Within each population, newly eclosed adults were transferred to one mesh cage containing *Arabidopsis thaliana* (Col-0 accession) and allowed to mate randomly. In each population, over 1,200 offspring (G1) were reared on a mixture of *T. glabra* and *B. vulgaris*, allowed to mate randomly again, and adult female offspring (G2) were preserved in 95% ethanol and phenotyped for the GWAS (Fig. S3 illustrates mating and rearing scheme). (Flies were collected and bred on different hosts for a separate study on host adaptation.)

We mounted ovipositors on slides as described above, counting only ventral bristles (Fig. 1c, 3a), summed across both oviscapts. We excluded posterior bristles, which were largely invariable in number, and lateral bristles because we speculated that their involvement in leaf-cutting may be limited due to their smaller size. We quantified ovipositor length as described above, and wing chord (proxy for body size), measured from the wing base to the apex following the third longitudinal vein (Fig. 3a). Two independent measurements were averaged per specimen. Linear regression analyses in a pilot experiment (NH1/NH2 flies, N=100) revealed that bristle number was positively correlated with ovipositor length (B = 0.097 [S.E. = 0.025] pegs per μm length, R2 = 0.134, P = 0.0001), but not wing length (B = 0.001 [S.E. = 0.002], P = 0.25). We therefore quantified both ovipositor length and bristle number for all individuals (NH1, N=308 flies; NH2, N=422 flies). Narrow-sense heritabilities of ovipositor length and bristle number were quantified using mother-daughter regression; details are presented in the Supplemental Methods.

(d) Pooled genome sequencing.

Flies in the NH1 and NH2 populations were split into two phenotypically extreme pools per population (four pools: NH1-low, NH2-low, NH1-high, NH2-high), composed of

60-85 females in the upper or lower 20% tail of the distribution of residual bristle number. Residual bristle number was determined through a linear regression of ovipositor bristle number against ovipositor length using the lm function in R. Flies were homogenized with stainless-steel beads and TissueLyser (Qiagen). Genomic DNA was extracted using a DNeasy Blood and Tissue Kit (Qiagen). One Illumina library per pool was constructed with 100bp paired-end reads and a 350bp insert size. Each library was sequenced on one half lane on an Illumina HiSeq 2500 at Arizona State University.

(e) Read mapping, pool-GWAS and gene ontology enrichment analysis.

Illumina reads were mapped to the S. flava reference genome (GenBank accession no. GCA_003952975.1) and filtered following best practices for pooled genome sequencing (Schlötterer et al., 2014). Statistical significance of between-pool allele frequency differences per site was estimated using the Cochran-Mantel-Haenszel test (Kofler et al., 2011). We identified and conservatively sought to correct for a slight inflation of P-values (Devlin & Roeder, 1999). However, because the P-value distribution was non-uniform with an excess of higher and lower values, typical corrections based on the observed versus median test statistic gave unsuitable inflation factors. We therefore regressed observed against expected –log10(P) values with the intercept constrained to 0 and divided each –log10(P) value by the slope of the regression line (Thoen et al., 2017). Further details are presented in the Supplemental Methods.

To identify genes located in or near the top SNPs (ranked by P-value), we located the nearest annotated gene in either direction, using genome-wide annotations for *S. flava* (Gloss et al., 2019). We checked for unannotated genes between the SNP and closest annotated gene by comparing the spanning sequence against the *D. melanogaster* RefSeq protein database, using NCBI BLASTx with default settings. Gene functions were gathered from the Gene Summary, Gene Ontology Annotations, and linked publications in Flybase (release 2020_01) (Thurmond et al., 2019). To better interpret the pool-GWAS results, we profiled linkage disequilibrium (LD) and population structure in several wild populations of S. flava in Massachusetts and New Hampshire. Further details are presented in the Supplementary Methods.

To determine if any predicted functions were overrepresented among genes intersecting the top GWAS associations, we performed a Gene Ontology enrichment test using GOWINDA, which implements a permutation-based approach tailored to the properties of GWAS datasets (Kofler & Schlötterer, 2012). Full details, including orthology-based functional annotation and extension of gene models to capture regulatory regions, are described in the Supplemental Methods.

(f) Reproducing pool-GWAS association for a candidate SNP.

Pool-GWAS can be confounded by uneven contributions of individuals to pools and biases in sequencing and read mapping. To verify our pool-GWAS results using an approach robust to these confounding factors, we genotyped individual females at one of the top SNPs and estimated its effect size. The SNP was chosen because of its close

proximity to *G alpha i subunit* (*Gαi*), a gene involved in asymmetric cell division of sensory organ precursor (SOP) cells from which bristles are derived (Schaefer et al., 2001). Ovipositor bristle number and length were measured as described above. Genomic DNA was extracted from 74 females (NH1/NH2 (G2), Table S5), and a target region of 500bp around the SNP was Sanger-sequenced. Bristle number was modeled in a generalized linear model, assuming an additive effect of the major allele, using the lm function in R. Additional details are presented in the Supplementary Methods.

Results

(a) The evolution of herbivory coincided with an increase in ovipositor sensilla number within the *Scaptomyza* lineage.

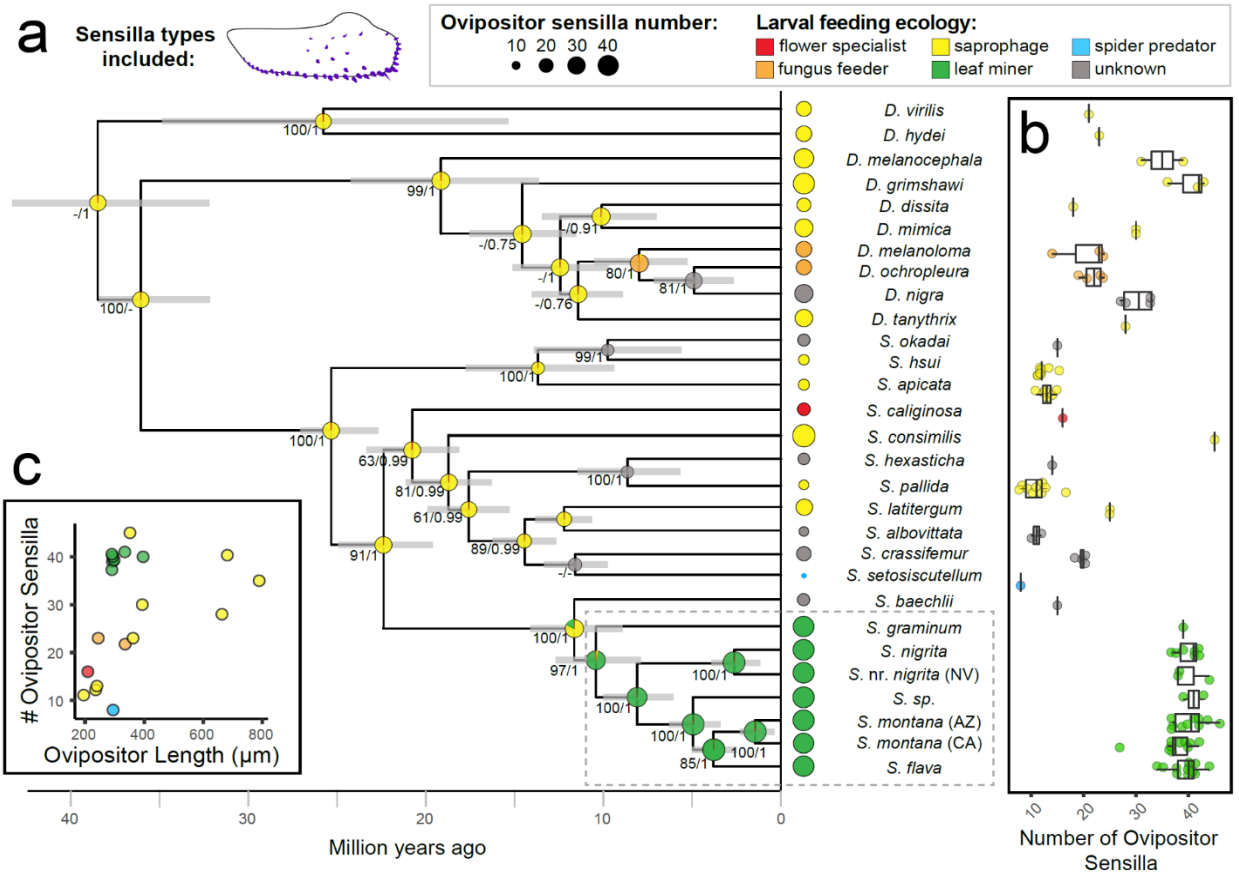


Figure 2.
The evolution of herbivory within *Scaptomyza* coincides with an increase in ovipositor sensilla number. (a) Time-calibrated phylogeny of herbivorous *Scaptomyza* and their non-herbivorous relatives, based on ML and Bayesian analyses, using 95 taxa, 11 genes and fossil and biogeographic time calibrations. Only taxa with data on ovipositor sensilla number and diet are shown (N=29; see Fig. S1-S2 for full phylogeny). Branch support is indicated by ML bootstrap values (≥50%) and Bayesian posterior probability (≥0.9). At nodes, bars indicate 95% highest posterior density interval around mean node age, pie graphs show probabilities of ancestral larval diets, and circle size represents ancestral ovipositor sensilla number per oviscapt estimated from ML ASR. Average sensilla numbers for extant species are shown at the tips, with individual counts shown in (b). (c) Scatterplot of ovipositor sensilla number as a function of ovipositor length.

PGLS methods revealed that ovipositor sensilla number is strongly influenced by larval diet (herbivorous versus non-herbivorous) ($F_{1,19}$ = 5.801, P = 0.028), phylogenetic relatedness (Pagel's λ = 1), and by ovipositor length (F1,19 = 4.655, P = 0.047), but not by source type (literature versus this study) ($F_{1,19}$ = 1.401, P = 0.254) (Table S6). Ancestral state reconstructions of ovipositor sensilla number and larval diet similarly suggest that ovipositor sensilla number increased coincident with the evolution of herbivory in *Scaptomyza*, estimated ~10.4 million years ago (mya) (8.2 -13 mya, 95% highest probability density) (Fig. 2a; Fig. S4). Relative to interspecific differences, variation within species is low (Fig. 2b).

(b) GWAS on ventral ovipositor bristle number in *Scaptomyza flava*.

Variation in ovipositor bristle number followed a continuous distribution in the NH1 and NH2 outbred laboratory populations of *S. flava* (Fig. 3a-b), typical of a quantitative trait controlled by multiple loci. Linear regression of ovipositor bristle number from mother-daughter pairs, controlling for the effect of ovipositor length, revealed that additive genetic variation accounted for half of this phenotypic variation (P = 0.034, h^2 = 0.50 ± 0.27 SE; Fig. 3c). By contrast, variation in ovipositor length was not heritable (P = 0.31).

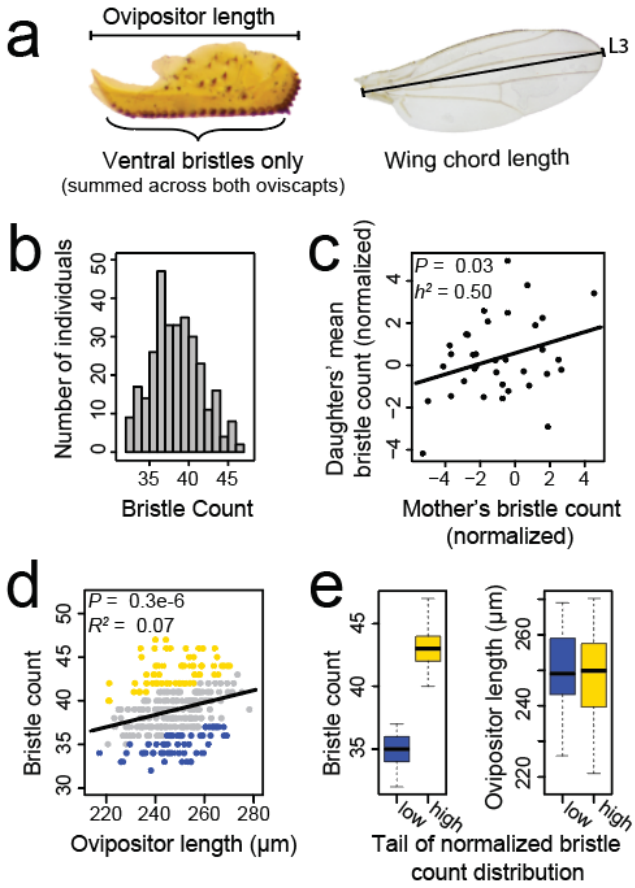


Figure 3.
Ovipositor bristle number is continuously distributed and heritable in *S. flava*, enabling quantitative genetic dissection. (a) Ovipositor bristle counts for this analysis included only ventral bristles. Wing chord length was measured along the third longitudinal vein (L3). Panels b, d, and e show phenotype distributions of the pool-GWAS outbred mapping population (NH1). (b) Ovipositor bristle number follows a continuous distribution. (c) Ovipositor bristle count, expressed as residuals from a linear regression of bristle count against ovipositor length, is heritable in the narrow sense (h2=0.50) from mother-daughter regression analysis (N=35). (d) After regressing out the effect of ovipositor length on bristle count, pools of phenotypically extreme individuals included either the upper (yellow) or lower (blue) 20% distribution tails. (e) Low pool individuals had ~20% fewer bristles than those in the high pool, but not statistically different ovipositor lengths.

We next sought to characterize the genomic architecture underlying this variation using a pool-GWAS. Because ovipositor length was positively correlated with bristle number (Fig. 3d), pools were constructed by adjusting bristle number relative to that expected from ovipositor length (Fig. 3e). This approach should interrogate bristle number independently of ovipositor size and minimize noise from non-heritable variation in

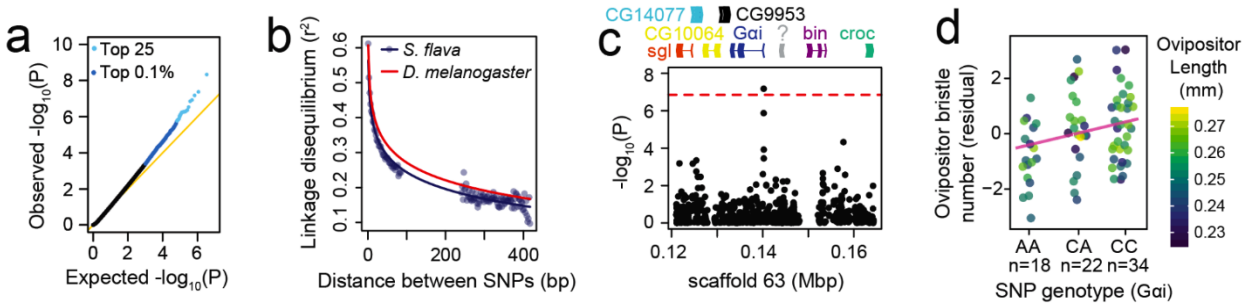


Figure 4.
Pool-GWAS for variation in *S. flava* ovipositor bristle number implicates genes involved in nervous system development. (a) An excess of strong P-values (conservatively adjusted for genomic inflation) suggests an enrichment of true associations among the top scoring SNPs. (b) The relationship between physical distance and linkage disequilibrium, inferred from pooled sequencing of wild *S. flava*, is similar to that seen in *D. melanogaster*. The distribution gap reflects the difference in read insert size (350bp) and sequenced read length (100bp). (c) Manhattan plot centered on a top SNP upstream of *G-alpha i subunit* (*Gαi*), a gene involved in nervous system development (Schaefer et al., 2001). The red line indicates the 5% FDR cutoff for genome-wide significance. Annotated genes are plotted above. (d) Genotyping individuals for the SNP near *Gαi* recapitulate pool-GWAS association. Bristle number, expressed as residuals (observed-predicted based on covariates), increases additively and independently of ovipositor length, shown by color scale. Regression line shown in pink.

ovipositor length. Mapped reads from whole genome re-sequencing of the four pools had a mean experiment-wide coverage depth of 166X per polymorphic site (per pool: NH1-low: 31x; NH1-high: 38x; NH2-low: 23x; NH2-high: 53x).

After excluding low frequency variants (1.6 million SNPs remaining), we found an excess of SNPs with significantly differentiated allele frequencies among high and low bristle number pools (Fig. 4a), with 5 and 19 significant SNPs at 5% and 10% false discovery rate (FDR) cutoffs, respectively (Tables 1 and S7). Because LD decays in *S. flava* at a rapid rate similar to that seen in *D. melanogaster* (Fig. 4b), SNPs showing strong associations are likely in close proximity to causal polymorphisms or are causal themselves. LD decay rates were similar across several populations in Belmont, MA and Portsmouth, NH, and across two host species (*T. glabra* and *B. vulgaris*) (Fig. S5). We cannot fully discount the possibility that long-range LD caused by undetected population structure could cause false associations between ovipositor bristle number and candidate SNPs. However, we did not find evidence for genetic population structure (F_{st}) across these populations (Fig. S6).

Many of the top SNPs (Table S7), including those reaching genome-wide significance (FDR$\leqslant$0.05, Table 1), were located near genes involved in neural development or neural cell fate specification (i.e., *G protein alpha i subunit, sloppy paired 2, tenascin accessory*), cytoskeleton organization (i.e. *muscle-specific protein*), and cuticle development (i.e. *cuticular protein 11B*).

Table 1.
Top SNPs associated with variation in ovipositor bristle number are located in or near genes involved in the development of bristles, cuticles, and the nervous system. SNPs reaching genome-wide significance (FDR⩽0.05) from the pool-GWAS are shown in descending P-value ranking.

P-value ranking	Scaffold	Position on scaffold	P-value	FDR q value	Nearby gene(s)	SNP location (relative to gene)	Gene product function (known from *D. melanogaster*)
1	scaffold00465	184	4.66E-09	0.008	**Muscle-specific protein**	453 bp downstream	Actin binding; cytoskeleton organization; required for proper positioning of muscle nuclei, mitochondria, and neuromuscular junction
2	scaffold00015	186,025	4.26E-08	0.036	**heavyweight**	8,623 bp downstream	Predicted to have phosphotyrosine residue binding activity; polymorphisms associated with body mass and starvation resistance
					cuticular protein 11B	1,625 bp downstream	Chitin-based cuticle development
3	scaffold00063	140,206	6.69E-08	0.037	**G protein alpha i subunit**	2 bp upstream upstream	Asymmetric neuroblast division and asymmetric protein localization involved in cell fate determination; cytoskeleton organization; and nervous system development
					Don juan-like [D. grimshawi]	3,446 bp downstream downstream	Unknown
4	scaffold00053	135,098	1.20E-07	0.048	**sloppy paired 2**	560 bp upstream	Transcription factor that regulates embryonic segment polarity and neural fate specification by temporal patterning of medulla neuroblasts
					CG11018	1,809 bp upstream	Unknown
5	scaffold00071	290,082	1.42E-07	0.048	**CG32655**	9,932 bp downstream	Unknown
					tenascin accessory	42,315 bp downstream	Nervous system development; regulation of cell-cell adhesion; and synapse organization

* Inferences about gene function were based on orthologous function in D. melanogaster from Flybase FB2020_01.[38]

(c) Gene ontology enrichment analysis on candidate SNPs.

To gain insight into developmental mechanisms that may contribute to variation in ovipositor bristle number, we tested for enriched gene ontology (GO) annotations among genes intersecting SNPs with the strongest GWAS associations (top 0.1% and 0.005% of *P*-values genome-wide). Using a restricted set of GO terms to minimize redundancy, we uncovered a single enriched term: RNA polymerase II-specific DNA-binding transcription repressor activity (GO:0001227; Tables 2 and S8). Several of the transcriptional repressors identified fine-tune gene expression levels during the specification of cell fate during neural development. Notably, the strongest GWAS association among transcriptional repressors falls in the gene *hairy* (*h*), which in *D. melanogaster* functions in the establishment of bristle precursor positioning from within proneural clusters (Usui et al., 2008).

We further tested for enrichment using the exhaustive list of all GO terms. Although no terms were enriched after applying a strict Bonferroni correction, two terms surpassed a nominal cutoff of $P < 0.001$, and both reflect broadly conserved developmental functions: phosphatidylinositol (PI) biosynthetic process and establishment of cell polarity (Tables 2 and S8). Many of the candidate genes annotated with PI biosynthetic process (GO:0006661) are kinases and transferases involved in the production of PI derivatives, which act as signaling molecules that regulate cellular growth and patterning (Balakrishnan et al., 2015; Hassan et al., 1998). Notably, establishment of cell polarity (GO:0030010) precedes the differentiation of sensory organ precursors into distinct neural cell types through asymmetric cell division (Schweisguth, 2015). *G protein α i*

subunit (*Gαi*) was one of the cell polarity genes identified and was also one of the strongest pool-GWAS associations (Fig. 4c).

Table 2.
Gene ontology terms enriched among genes intersecting the most significant pool-GWAS SNPs.

Candidate SNPs	GO Terms Investigated	Enriched GO Term	Genes intersecting candidate SNPs: Observed	Expected	Possible[a]	Fold Enrichment	*P*	*P* (Bonf.)	Genes intersecting candidate SNPs (named by orthology to *D. mel.*)
Top 0.1%	Non-redundant subset	DNA-binding transcription repressor activity, RNA polymerase II-specific (Molecular Function, level 4; GO:0001227)	16	5.469	63	2.93	0.00004	0.0092	*aop, chn, dpn, E(spl)mβ-HLH, E5, h, Hey, HHEX, l(2)gd1, lms, Mad, Med, Rbf2, CG12299, CG1233, CG7987*
Top 0.1%	Full set	Phosphatidylinositol biosynthetic process (Biological Process, level 6; GO:0006661)	8	2.013	43	3.97	0.00067	1.00	*GAA1, Pi3K68D, PIG-O, PIG-S, PIG-Z, PIP5K59B (x2), CG5342*
Top 0.005%	Full set	Establishment of cell polarity (Biological Process, level 3; GO:0030010)	5	0.751	92	6.66	0.00075	1.00	*Gαi, Khc-73, scrib, sktl, CG5964*

DNA-binding transcription repressor activity, RNA polymerase II-specific (GO:0001227)

Genes intersecting SNPs	Adjusted *P*-value	FDR q-value	Genome-wide Rank	SNP location	Transcriptional repressor function (known from *D. melanogaster*)	Nervous system dev.	Notch signalling pathway
aop	3.83E-04	0.41	1531	120 bp upstream	Regulates cell fate transitions during development of the nervous system, heart, trachea and eye.	✓	
chn	6.46E-05	0.28	379	1255 bp upstream	Regulates *emc, h, ac* and *Dl*. Functions in sensory neurons, photoreceptors, blood cells, and muscle and intestinal precursors.	✓	
dpn	9.49E-05	0.3	509	within 3rd exon	Regulates genes requiring bHLH protein for transcription. Maintains self-renewal and identity of type II neuroblasts in larval brain.	✓	✓
E(spl)mβ-HLH	6.07E-05	0.27	369	973 bp downstream	Regulates genes requiring bHLH protein for transcription, including Achaete-scute complex (AS-C) genes. Contributes to neural-epidermal lineage decision during neurogenesis.	✓	✓
E5	3.73E-05	0.25	237	within 1st intron	Predicted to be involved in brain development and neuron differentiation.	✓	
h	1.73E-06	0.11	23	within 1st exon	bHLH transcriptional repressor that recruits Gro corepressor to target promoters. Contributes to embryonic segmentation and peripheral neurogenesis.	✓	
Hey	3.01E-04	0.39	1269	within 2nd intron	*Hairy/E(spl)-related with YRPW motif* (*Hey*) encodes a bHLH transcription factor involved in neuron fate determination.	✓	✓
HHEX	6.96E-05	0.28	405	within 1st intron	Involved in response to sucrose. One allele shown to affect trichogen cells, which generates the sensillum shaft.	✓	
l(2)gd1	3.80E-04	0.41	1520	within 3rd exon	Involved in female germ-line stem cell asymmetric division; wing vein morphogenesis; and sensory organ precursor cell division.	✓	✓
lms	5.60E-05	0.26	351	1107 bp downstream	Homeodomain transcription factor.		
Mad	4.16E-05	0.25	266	within 1st intron	Mediates response to BMP ligands encoded by *dpp, scw* and *gbb* (proteins involved in growth regulation, patterning and stem cell fate).	✓	
Med	2.01E-04	0.35	943	2496 bp downstream	Binds to Mad or Smox to facilitate signal transduction for *dpp* or Activin ligands. Involved in dorsal-ventral patterning, patterning/proliferation of wing disc and gene expression in larval brain mushroom body.	✓	
Rbf2	3.82E-04	0.41	1528	184 bp upstream	Cell cycle and developmental regulator. Represses the expression of differentiation markers in ovaries and embryos.		
CG12299	3.61E-04	0.4	1466	901 bp downstream	Predicted to enable DNA-binding transcription repressor activity.		
CG1233	3.85E-05	0.25	244	within 3rd exon	Predicted to enable DNA-binding transcription repressor activity.		
CG7987	4.33E-06	0.17	39	957 bp downstream	Predicted to enable DNA-binding transcription repressor activity.		

[a] The maximum possible number of intersections equals the number of genes assigned to the GO category that have at least one genotyped SNP passing quality control filters.

(d) Reproducing pool-GWAS association at a top candidate SNP (*Gαi*).

To confirm that the pool-GWAS adequately estimated allele frequencies, we focused on a SNP in the 5' UTR of *Gαi*, one of the strongest pool-GWAS associations, by phenotyping and genotyping individual adult female flies. Bristle number increased by 0.58 per major allele carried (β = 0.11 standard deviations, t(68) = 2.88, $P < 0.005$; Fig. 4d; Table S9). This SNP explained 9.5% of the total variance in bristle number (partial adjusted r^2). As expected given our study design, the SNP did not have an effect on ovipositor length (β = 0.02 standard deviations, t(69) = 0.177, $P > 0.05$; Table S10). Out of five variant sites (≥0.05 min. freq.) in the sequenced region, two were in strong LD with the focal SNP (Table S11). Further study will be necessary to identify the causal variant(s) in this region.

Discussion

The plant-penetrating ovipositor of herbivorous insects presents an excellent opportunity to study the evolution and genomic architecture of a complex trait, given its

clear role in egg-laying and the quantitative nature of its traits. We focused on ovipositor evolution in the genus *Scaptomyza*, in which herbivory has evolved relatively recently, ca. 10.4 mya. The wealth of data from the *Drosophila* literature made our analyses possible: genitalic morphological data available from numerous taxa to investigate evolutionary shifts in bristle number across species, and genetic and development knowledge of bristle number in *D. melanogaster* to understand the genetic architecture underlying variation at the population level in *S. flava*.

From a macroevolutionary perspective, we found that ovipositor sensilla number underwent a marked increase that coincided with the evolution of herbivory within *Scaptomyza*, a significantly larger shift than expected from the distribution of background rates of evolution across the phylogeny (Fig. 2a). Surprisingly, we also found that ovipositor sensilla number is evolutionarily malleable, repeatedly increasing and decreasing across the phylogeny, with a five-fold range across *Scaptomyza*. High variability was similarly seen within species, including a 1.5-fold range in *S. flava*. The lack of strong evolutionary constraint over both macro- and microevolutionary timescales, along with heritable standing genetic variation within populations, suggests that ovipositor sensilla number may be highly accessible to adaptive evolution. However, considering there was only one occurrence of the evolution of herbivory among sampled species, it will be necessary to test whether the same patterns exist in other independently evolved herbivorous lineages, such as those that include *D. suzukii*, *Scaptodrosophila notha*, and herbivorous *Lordiphosa*. Further research on herbivorous drosophilids could also test whether heritable variation in ovipositor bristle number could be selected upon for divergent host use (i.e., plants of varying leaf or fruit skin toughness), as seen in fig wasps and pine-specialized sawflies (Bendall et al., 2017; Ghara et al., 2011).

Our phylogenetic analysis (PGLS) suggests that increased sensilla number may have evolved in addition to or partly as a result of ovipositor elongation. A longer ovipositor can accommodate more bristles, and increased ovipositor length has been studied in D. suzukii as a key trait to facilitate cutting into ripe fruit (Green et al., 2019). It will be necessary to examine whether these morphological traits are linked at molecular and developmental levels. Our GWAS in *S. flava* should have, nonetheless, targeted variation in ovipositor bristle number, rather than length, because we used bristle counts adjusted by length, and only bristle number (not length) exhibited narrow sense heritability (Fig. 3c).

Pinpointing genetic changes that gave rise to traits that evolved millions of years ago can be difficult because genetic architectures may differ over short versus long timescales (Stern & Orgogozo, 2008). Still, GWAS can illuminate genes and gene functions that shape standing phenotypic variation and may contribute to evolution over longer timescales. Our GWAS results indicate that broadly conserved neurodevelopmental genes, such as *Gαi* and *slp2*, play a role in ovipositor bristle density (Table 1) (Bhat et al., 2000; Schaefer et al., 2001). Genes encoding transcription repressor proteins were significantly enriched near the strongest GWAS associations, with the majority involved in neural development, neuron differentiation, and are regulated by or regulators of the Notch signaling pathway (GO:0001227, Table 2). For instance,

four genes (h, E(spl)mβ-HLH, dpn, and Hey) repress basic helix-loop-helix (bHLH) proteins, which are important regulators of neurogenesis. These results are consistent with existing knowledge that insect sensilla are developmentally derived from neural precursor cells (sensory organ precursors or SOPs). Bristle patterning begins with expression patterns of proneural genes, like the achaete-scute complex, that generate proneural clusters of cells. Within these clusters, the selection of the SOP is determined by lateral inhibition mediated by Notch signaling, followed by SOP differentiation through asymmetric cell divisions into cells that form the shaft, socket, and sheath, and mechanosensory and chemosensory neurons that innervate the sensilla (Orgogozo et al., 2001). We thus speculate that mutations in or near genes involved in SOP development could cause shifts in SOP patterning, spacing, or density, thus producing more bristles. E(spl)mβ-HLH is a particularly strong candidate (P-value = 6.07E-5), as it is directly regulated by Notch and represses the activity and expression of proneural Achaete and Scute proteins (Couturier et al., 2019).

In other *Drosophila* species, genes involved in neural development also underlie differences in bristle number on the male genitalia, sexcombs of the forelegs (Nagy et al., 2018) and the thorax (Marcellini & Simpson, 2006). From the same overrepresented GO category of transcriptional repressors, the top-scoring SNP (P-value = 1.73E-6) was located near the gene *hairy* (*h*), a direct repressor of achaete (Van Doren et al., 1994). RNAi knockdown of hairy in *Drosophila* has validated its involvement in male genitalic clasper size and bristle number, and association mapping has shown that it falls within a narrow genomic region underpinning divergence in clasper bristle number among sister species of *Drosophila* (Hagen et al., 2021). Its role in bristle and genital development, along with its contribution to intra- and inter-species variation in bristle number, make *hairy* a strong candidate for ovipositor bristle variation. It also presents an opportunity to investigate genetic parallelism for bristle number variation across the body, between sexes and across species.

Studies on the genetic architecture of adaptive traits have largely focused on monogenic, Mendelian traits with large effect loci and lower detection thresholds than genetically complex traits (Lamichhaney et al., 2016; Linnen et al., 2009; Rost et al., 2004). Ovipositor bristle number represents a tractable quantitative trait for genetic dissection because of its meristic nature, high variability, heritability, and clear importance in facilitating entry into a new niche. Despite having a genetic basis similar to many quantitative traits — many, small effect SNPs — we still were able to detect a SNP with moderately large effect (confirmed by individual genotyping). Our results suggest that pool-GWAS can be a viable method for pinpointing genomic regions that underlie quantitative trait variation. Candidate SNPs can then be interrogated through functional experimentation to understand how alternative alleles influence cell division, size expansion, and reorganization during development (Green et al., 2019). Ultimately, this could illuminate how incremental changes could have created this key trait in herbivorous insects.

While Chapter 1 focused on the cutting ovipositor – a morphological trait that provided herbivorous drosophilids, like *Scaptomyza flava*, with access to the soft leaf interior for feeding – Chapter 2 explores adaptations that contend with plant secondary metabolites, another major hurdle in the transition to herbivorous feeding. Faced with a complex sensory environment with numerous potential hosts, herbivorous insects need to evolve chemosensory adaptations that enable attraction and acceptance of their host plants as feeding or egg-laying sites. This often requires shifts in their perception of host plant toxins. These secondary metabolites present a twofold challenge as they also may slow the development of feeding herbivores or kill them outright. Herbivores thus need to evolve mechanisms to tolerate or detoxify these chemical compounds. Chapter 1 explored the genetic basis of the cutting ovipositor by testing associations across the entire genome, but Chapter 2 uses a candidate gene approach, given the purported importance of several large chemosensory and detoxification gene families in driving dietary shifts. Here, I analyze the molecular evolution of these gene families to identify evolutionary patterns that may be driving herbivore-specific adaptations.

Chapter 2. Comparative genomics of chemosensory and detoxification gene family evolution across herbivorous *Drosophila*

Abstract

Herbivorous insects are exceptionally diverse, accounting for a quarter of all known eukaryotic species, but the genetic basis of adaptations that enabled this dietary transition remain poorly understood. Many studies have suggested that expansions and contractions of chemosensory and detoxification gene families – genes that directly mediate interactions with plant chemical defenses – underlie the successful colonization of plants. However, this hypothesis is challenging to test because the origins of herbivory in many lineages are ancient (>150 million mya), obscuring evolutionary patterns. Recently derived (<15 mya) herbivorous lineages hold promise in this regard. Here, we characterized patterns of evolution across *Scaptomyza,* a lineage nested within *Drosophila*, which includes herbivorous mustard specialists (Brassicales, including *Arabidopsis thaliana*), carnation specialists (Caryophyllaceae), and several non-herbivorous species. Comparative genomic analyses revealed that herbivorous *Scaptomyza* have among the smallest chemosensory and detoxification gene repertoires across the species surveyed (five *Scaptomyza* and seven *Drosophila* species), and significantly higher rates of gene turnover (cumulative gains and losses) within these gene families. Among genes most significantly impacted by changes in copy number or selective constraint were those involved in the detection of compounds associated with their ancestral diet (yeast volatiles), or with plant feeding (bitter phytotoxins), as well as a surprising number of changes associated with pheromone, water, salt and cold detection. Although elevated levels of gene duplication were seen across each of the herbivorous species relative to background rates, shared duplications across all herbivores were less dramatic, suggesting a more limited role of gene family expansions in the evolution of herbivory.

Keywords: genomics, trophic shift, gene family evolution, positive selection, plant–herbivore interactions

Introduction

The origin of land plants over 500 million years ago presented a new niche for early insects to colonize (Southwood 1972). The intimate relationship between plants and insects has since generated one of the most ecologically and evolutionarily dominant groups in the earth's history: the herbivorous insects. Herbivorous insects account for over a quarter of all known eukaryotic species and help form the basis of terrestrial food webs (Bernays 1998; Farrell 1998; Strong et al. 1984). It has long been hypothesized that the diversity of herbivorous insects emerged as a result of co-diversification processes with their host plants (Farrell 1998; Mitter et al. 1988; Marvaldi et al. 2002). As described in Ehrlich and Raven's (1964) seminal paper, by evolving adaptations to overcome plant chemical defenses, herbivores can diversify or sub-specialize onto plants bearing the

same chemical defenses (in one or a few plant families), that is, until new plant defenses evolve, allowing plants to escape attack and then also diversify under this release from herbivore pressure, and so on in this 'coevolutionary arms-race' (Ehrlich & Raven 1964). This theory of coevolution, commonly referred to as 'escape and radiate' (Thompson 1988, 1994), has inspired much of the research on insect-plant interactions throughout the last several decades, (Simon et al. 2015; Vertacnik & Linnen 2017; Heidel-Fischer & Vogel 2015). As a result, we have begun to gain insight into the adaptive mechanisms underlying plant colonization, which has specifically focused on genes that mediate interactions with plant secondary compounds: chemosensory and detoxification genes (Ehrlich & Raven 1964; Weimin Li et al. 2003; Futuyma & Agrawal 2009; Edger et al. 2015; Berenbaum 1983). Despite significant progress towards understanding the genetic basis of herbivore-plant interactions, there is still much to learn regarding the evolutionary processes shaping these large, complex and rapidly evolving gene families.

Previous studies have suggested that the source of genetic variation that underlies dietary novelty in herbivorous arthropods arises from extensive gene family expansions and rampant gene loss or pseudogenization. For example, the spider mite (*Tetranychus urticae*) and diamondback moth (*Plutella xylostella*) genomes harbor expanded gene families encoding enzymes involved in detoxification of the plant secondary compounds they encounter (Dermauw et al. 2013; You et al. 2013; Grbić et al. 2011), and genes encoding gustatory receptors (GRs) involved in host finding have undergone extensive lineage-specific duplications in butterflies generally (Briscoe et al. 2013). Despite the identification of these gene-family expansions associated with plant-insect interactions, (Edger et al. 2015; Johnson et al. 2018; McBride 2007), isolating herbivory as the cause of these genome-scale patterns remains controversial (Gardiner et al. 2008). The most diverse living lineages of herbivores, which include butterflies and moths (Lepidoptera), and leaf beetles, weevils and close relatives (Phytophaga), arose in the late Paleozoic (Kawahara et al. 2019) and early Mesozoic (Wiens et al. 2015). For this reason, comparative genomic studies on the evolution of herbivory would benefit from examining herbivores that have more recently diverged from non-herbivore relatives, to better differentiate herbivore-associated changes from those that might have arisen in response to other stochastic processes or agents of selection (i.e. demography or other environmental shifts) (Yassin et al. 2016). While clear patterns have emerged from some older herbivore lineages (You et al. 2013; Suzuki et al. 2018; Karageorgi et al. 2019), it remains challenging to differentiate cause and effect: changes involved in the initial dietary transition from subsequent changes as a consequence of specialization over a hundred million years of selection. Furthermore, we have a limited understanding of whether the majority of these copy number changes functionally impact genes interacting with plant secondary compounds, or whether they share behavioral or physiological functions, because most herbivorous taxa lack the genomic resources and functional tools to test this.

In this study, we address these limitations by using a comparative genomic approach to examine genetic changes associated with the evolution of herbivory within the Drosophilidae. Although most species within Drosophilidae retain the ancestral habit of laying eggs and feeding on decaying plant tissue and associated microbes, herbivory

has evolved several times (Durkin et al. 2021; Okada & Sasakawa 1956; Whiteman et al. 2011; Jones 1998; Maunsell et al. 2017). A clade of herbivorous species are members of the genus *Scaptomyza*, a monophyletic lineage nested within the paraphyletic subgenus *Drosophila,* which also includes Hawaiian *Drosophila*, as well as *D. virilis* and *D. mojavensis* (Fig. 1a). Herbivorous *Scaptomyza,* particularly *S. flava,* have become

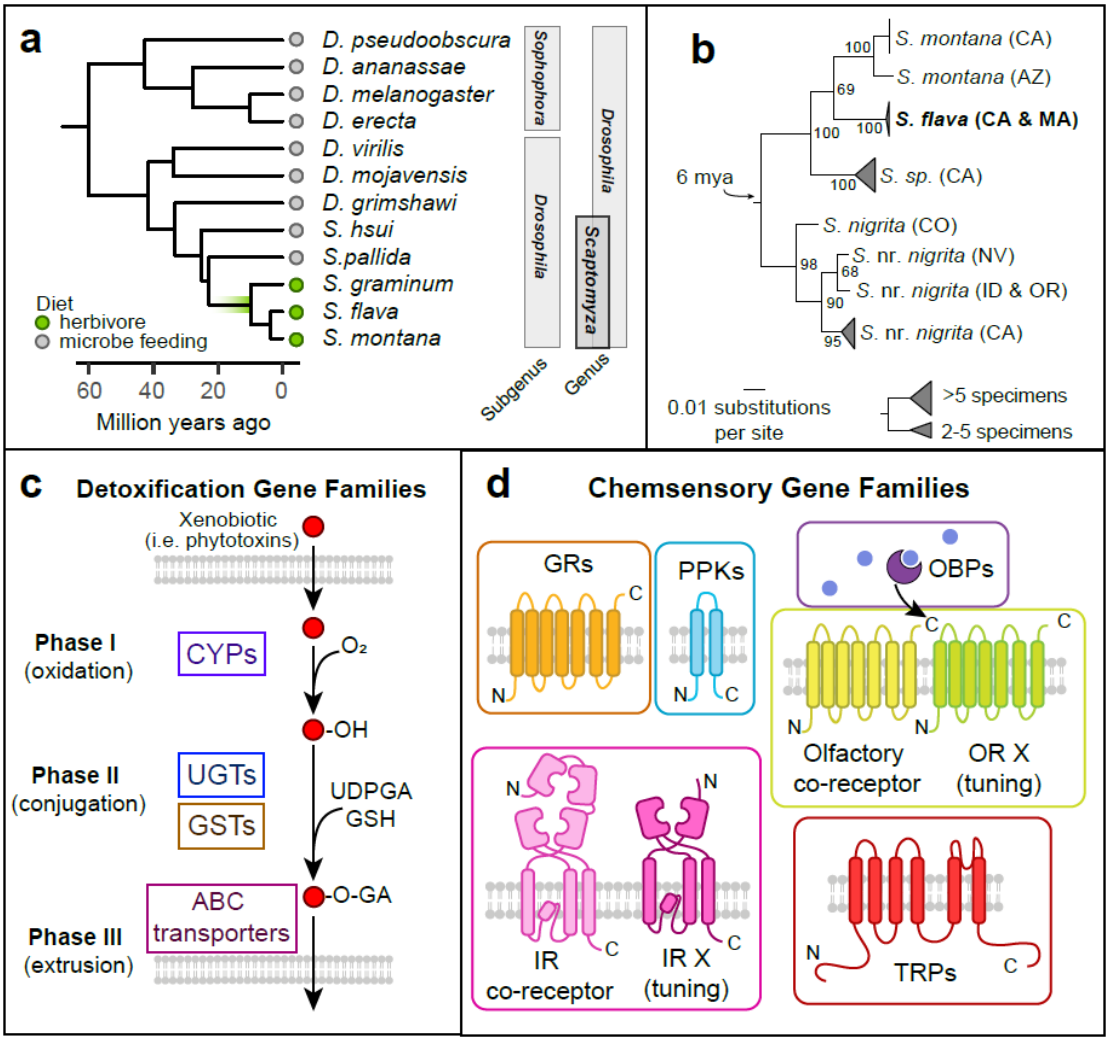


Figure 1. Gene families at the interface of plant-herbivore interactions. (a) Phylogenetic placement of herbivorous *Scaptomyza* within the paraphyletic genus *Drosophila* (Matsunaga et al., 2022). (b) Maximum likelihood nucleotide phylogeny built using *COI* sequences from North American *Scaptomyza* collected feeding on mustard plants (Brassicales spp.). Individuals with < 1% pairwise nucleotide divergence were collapsed into clades. Collection localities indicated by two-letter state abbreviations. For the current study, the genome of an *S. flava* line from MA, USA (bolded) was sequenced and assembled. (c) Detoxification of xenobiotics/plant secondary metabolites (phytotoxins) involves three phases: oxidation/reduction by cytochrome P450s (CYPs), enzymatic conjugation by UDP glucuronosyltransferases (UGTs) or glutathione S-transferases (GSTs), and excretion/transport out of the cells by ABC transporters. (d) Insects detect environmental compounds with six major chemosensory gene families: gustatory receptors (GRs), odorant binding proteins (OBPs), olfactory receptors (ORs), degenerin/epithelial sodium channels (pickpockets PPKs), ionotropic receptors (IRs), and transient receptor potential channels (TRPs).

models to study the evolution of herbivory, not only because they are phylogenetically close to the model *D. melanogaster*, but also because several species are specialists on plants in the Brassicales and can be reared on the genetic model plant *Arabidopsis thaliana* (Whiteman et al. 2011). Genetic dissection of adaptive traits in this lineage has been enabled by the rich knowledge of gene function in *D. melanogaster*, a growing number of high quality genomes across the drosophilid phylogeny (Drosophila 12 Genomes Consortium et al. 2007; Kim et al. 2021), strong phylogenetic frameworks (Finet et al. 2021; O'Grady & DeSalle 2018), and the ability to test hypotheses utilizing genetic tools from both *Drosophila* and *Arabidopsis* (Groen & Whiteman 2016).

While the herbivorous *Scaptomyza* lineage is fairly young (~10 million years) (Lapoint et al. 2013; Peláez, Gloss et al. 2020), DNA barcoding suggests a potential cryptic radiation within this clade (eight species in North America; Fig. 1b), similar in species richness to the *D. melanogaster* subgroup worldwide (nine species), which is ~12 million years old (David et al. 2007). Among these species, we focused on the three herbivorous species *S. flava, S. montana*, and *S. graminum* because there is strong support to indicate that they represent distinct species, based on molecular barcoding and morphological differences. They also represent specialization to varying degrees and on different plant families: *S. graminum* specializes on plants of the Caryophyllaceae family, while both *S. flava and S. montana* are specialists on Brassicales, with *S. montana* being more specialized, exhibited by a strong preference for plants with indole glucosinolates (Gloss et al. 2017). Disentangling genomic patterns associated with specialization and herbivory is challenging as noted above, especially considering the prominent role of specialization in driving herbivore diversification rates. However, by including closely related specialists on different plant families, these comparisons may elucidate patterns associated with herbivory, rather than specialization. Similarly, studies on herbivorous *Scaptomyza* are further aided by the inclusion of closely related non-herbivorous *Scaptomyza* species: here, we include *S. pallida* and *S. hsui*, each from separate subgenera (*Parascaptomyza* and *Hemiscaptomyza*, respectively), which have been collected at the same sites as herbivorous species, feeding on decaying plant material. Including these non-herbivorous close relatives enables us to similarly differentiate herbivore-specific changes from those that arose further back in time along the *Scaptomyza* lineage.

Given the purported importance of gene family expansions and contractions among chemosensory and detoxification gene families in other herbivorous insects (Edger et al. 2015; Johnson et al. 2018; McBride 2007), we hypothesized that herbivorous *Scaptomyza* would have similarly experienced dramatic copy number changes to key genes involved in sensing or detoxifying plant chemical defenses. We suspected that some of these duplicated genes that were retained would have undergone significant protein evolution, as they gained new functions through neofunctionalization or subfunctionalization (Ohno 1970; Lynch & Conery 2000). Conversely, in accordance with the neural limitation theory, we predicted that the loss of chemosensory genes would play an outsized role in the evolution of herbivory, to streamline neural processing in the face of choosing between many host plants (Bernays 2001). To address these hypotheses, we sequenced and assembled the genome of *S. flava* to analyze alongside published genome assemblies of the two herbivorous species (*S. montana* and *S. graminum*), the

two non-herbivores (*S. pallida* and *S. hsui*), and seven non-herbivorous *Drosophila*, ranging in phylogenetic distance from the herbivores (Fig. 1a). This included two non-herbivore specialists outside *Scaptomza* but still relatively closely related (*D. mojavensis* and *D. erecta*) to further polarize patterns of specialization from those of herbivory. Following the curation of the major chemosensory and detoxification gene families (those in Fig. 1c), we then used maximum likelihood methods to estimate rates of gene gain and loss across the tree to determine whether rates significantly differed in herbivore branches from background branches. We next generated codon-based, maximum likelihood models to identify genes with signatures of changing selective constraints along the branch at the base of the herbivore clade. Demographic history was also examined to evaluate whether our results may be explained by demographic events in recent history. Finally, we surveyed each genome assembly for evidence of horizontal gene transfer, which has been implicated in other major transitions to herbivory in invertebrates, from nematodes to insects. Taking into consideration orthologous gene function from *D. melanogaster*, we summarize molecular processes most impacted by the evolution of herbivory. Together, our analyses uncovered dynamic gene family remodeling in a nascent herbivore lineage, illuminated by comparisons with relatives retaining an ancestral, non-herbivorous feeding mode.

Materials and Methods

S. flava genome sequencing, assembly, and annotation.

Full details of these methods can be found in the Supplemental Methods.

PacBio and Dovetail HiC libraries and sequencing. Sequence data for our main *S. flava* assembly (sfla_v2) were generated from a partially inbred laboratory colony. The colony was founded from >150 larvae collected near Dover, NH, USA, and subsequently maintained for several years in the laboratory. 300 male flies were flash frozen and stored at -80°C. SMRTbell libraries (~20kb) for PacBio Sequel were constructed using SMRTbell Template Prep Kit 1.0 (PacBio, Menlo Park, CA, USA) using the manufacturer's recommended protocol. Sequencing was performed on two PacBio Sequel SMRT cells. A Dovetail HiC library was prepared in a similar manner as described previously (Erez Lieberman-Aiden et al., 2009). DNA was sheared to ~350 bp mean fragment size, and sequencing libraries were generated using NEBNext Ultra enzymes and Illumina-compatible adapters. The libraries were sequenced on an Illumina HiSeqX to produce 380 million 2x150 bp paired end reads, approximately 30x sequence coverage.

Illumina library and sequencing. Illumina sequence data was used to polish the PacBio/HiC assembly. Sequences were generated from a laboratory population initially collected in Belmont, MA, USA in 2008 that was inbred through 10 generations of single pair mating on *A. thaliana* Col-0 plants (*S. flava* “OHI-9 line”). Paired-end 180 bp and 300 bp insert libraries and 3 kbp and 5 kbp mate pair libraries from female flies were sequenced with 100 bp read length on an Illumina Hiseq 2000 at the University of Arizona. Reads were quality filtered and Illumina TruSeq3 adapters were removed using

Trimmomatic v0.35 (Bolger et al. 2014) with the following parameters: “LEADING:10 TRAILING:10 SLIDINGWINDOW:4:15 MINLEN:99”.

Genome assembly and scaffolding. The *S. flava* genome (sfla_v2) was assembled using the long-read hybrid assembly pipeline described in Kim et. al (2021), which has been shown to produce highly complete genome assemblies for drosophilid flies resulting in very few coding sequence indels (Kim et al. 2021). We generated an initial draft assembly with Flye 2.9 (Kolmogorov et al. 2019), identified and removed duplicated haplotypes (haplotigs) with purge_dups (Guan et al. 2020). We polished the draft assembly using the PacBio reads with one round of Racon (Vaser et al. 2017), then polished further using Illumina reads with one round of Pilon (Walker et al. 2014), only fixing base-level errors. The fully polished assembly was scanned for contaminant sequences using NCBI BLAST (Johnson et al. 2008) and BlobTools (Laetsch & Blaxter). Repetitive sequences in the assembly were identified with RepeatModeler2 (Flynn et al. 2020). To integrate the long-range assembly information generated with HiC, we scaffolded the assembly with an initial assembly generated by Dovetail (details included in the Supplemental Methods). Specifically, we soft-masked both genomes using the repeat library generated in the previous step, using RepeatMasker (Smit et al. 2013). Then, we created a whole-genome alignment with Progressive Cactus (Armstrong et al. 2020) and used the RagOut reference-based scaffolder (Kolmogorov et al. 2014) to scaffold the new genome. The genome assembly of *S. flava* was scaffolded into 1,252 scaffolds covering 315.4 Mbp (N50 = 32.966 Mb) with a maximum length of 85.98 Mbp and with 460 gaps.

Comparative annotation of *Scaptomyza* genomes. Gene annotations previously created for an Illumina-only *S. flava* assembly (sfla_v1) were transferred to the assemblies of *S. graminum, S. hsui, S. pallida, S. montana*, and to our main *S. flava* assembly (sfla_v2), using whole-genome Progressive Cactus alignments and the Comparative Annotation Toolkit (CAT, (Fiddes et al. 2018)). Details of these methods, including the generation of annotations for the Illumina-only sfla_v1 assembly, can be found in the Supplemental Methods.

Gene model curation & orthology inference.

An iterative curation strategy was used to identify the complement of genes in four published *Drosophila* genomes (*D. melanogaster, D. virilis, D. mojavensis,* and *D. grimshawi) (Drosophila 12 Genomes Consortium et al. 2007)*, four published *Scaptomyza* genomes (*S. pallida, S. hsui, S. graminum,* and *S. montana*) (Kim et al. 2021), and in our *S. flava* assembly (sfla_v2) for six chemosensory gene families (gustatory receptors, GRs; ionotropic receptors, IRs; olfactory receptors, ORs; odorant binding proteins, OBPs; pickpocket proteins, PPKs; transient receptor potential channels, TRPs) and three detoxification gene families (Cytochrome P450s, CYPs; Glutathione *S*-transferases, GSTs; UDP-glycosyltransferases, UGTs). For each species, assembly versions are listed in Table S1, along with which analyses species were included in. First, all protein sequences in each family from *D. melanogaster* were queried against the assembled genomes and annotated proteomes of *D. virilis, D. mojavensis,* and *D. grimshawi*, using BLASTP (e-value cutoff < 1e-3) (Altschul et al. 1997). The resulting collection of genes

was then queried against the automated annotations for *Scaptomyza* species (e-value cutoff < 1e-3). We iteratively ran BLAST searches using the identified genes from each species as queries against their genome assemblies until no new genes were identified. To validate lost genes, we performed an additional TBLASTN search (e-value cutoff < 1e-4) for each putatively lost gene among the *Scaptomyza* species. These searches used the predicted orthologs in *D. grimshawi* as queries, or *D. mojavensis* orthologs if the gene was also lost in *D. grimshawi*. Genes of interest were considered truly absent if TBLASTN searches yielded no hits. Sequences that were identical or nearly identical (>99% identity) but located on separate scaffolds were flagged as putative assembly errors and reduced to a single copy. Gene models significantly deviating in length from their *D. melanogaster* orthologs were manually inspected, and gene models were corrected when necessary, using homology to the annotated genes in other species. Additional validation steps we implemented are described in the Supplementary Methods. Nucleotide sequences were aligned in Geneious v.10.2.6 using MUSCLE (Edgar 2004) and manually inspected.

To verify that these genes belonged to the gene families of interest and that orthology was correctly assigned, maximum likelihood gene trees were constructed from these sequences using RAxML (Stamatakis 2006) with default settings. Orthology was inferred using the phylogenies built for each chemosensory and detoxification gene family. Genes were binned into orthologous groups if they formed a clade with > 70 bootstrap support and their phylogenetic topology suggested they were present as a single gene copy in the ancestor of the genus *Drosophila*. For poorly supported clades (bootstrap support < 70), orthology groups were assigned on the basis of previous orthology assignments (Low et al. 2007; Good et al. 2014; Almeida et al. 2014) and syntenic relationships. After verifying all gene sequences, a final maximum likelihood tree was constructed, using the same methods described above.

Gene family size evolution.

To identify chemosensory or detoxification gene families with rapid expansions and/or contractions among the herbivore lineages, we used the program Computational Analysis of gene Family Evolution (CAFE) v4.2.1 (Han et al. 2013), which models a birth-death process of gene gain and loss across a species tree. In addition to the nine species mentioned above, we included three species of the subgenus *Sophophora* (*D. pseudoobscura*, *D. ananassae*, and *D. erecta*) to improve background rate estimates. We imported gene models for these species from published orthology annotations (Good et al. 2014; Low et al. 2007; Almeida et al. 2014). As input, we provided a matrix of gene counts for each orthologous gene cluster, and we used a published time-calibrated phylogeny of drosophilids (Matsunaga et al. 2022), which included all our species of interest, with the exception of *S. montana*, which was grafted onto this tree using a reported divergence time estimate from *S. flava* (Peláez et al. 2020).

First, we generated models estimating the average turnover rate (λ) (the cumulative rate of gains and losses per gene per unit time). For each gene family, we compared a null model with a single λ rate estimated for all branches to a model with two rates: one rate estimated for a foreground branch, another for the remaining background

branches, where the foreground branch was a branch leading to one of the 12 species (Table S1), or the branch leading to all herbivores. All models were run in triplicate, and the iteration of each model with the highest likelihood was retained. To test whether a two-rate model was significantly better than the single-rate model, we performed likelihood ratio tests (LRT). *P*-values were adjusted at a false discovery rate of 1% using the qvalue package in R (Storey et al 2022). Using the same methods, we also generated models estimating gains and losses (λ, μ) separately, rather than as a single parameter, following the same methodology as described for turnover rate.

To test whether the rates of expansion and contraction of chemosensory and detoxification gene families deviate from those of other orthologous gene clusters genome-wide, we performed the same set of analyses on a generated set of 200 randomly chosen orthology groups. We identified these gene sets using OrthoVenn2 server (Xu et al. 2019; Wang et al. 2015), a web-based orthology assignment tool based on OrthoMCL (Li Li et al. 2003). We uploaded genome-wide proteomes for the five *Scaptomyza* species and utilized protein sequences for the remaining seven species available through the OrthoVenn2 server (which uses the Ensembl database, release January 2019). Default parameters were used with an e-value cutoff of 0.05 and an inflation value of 1.5. We filtered the clusters by those that contained sequences for all 12 species, and randomly chose clusters that summed to ~200 genes per species.

Molecular evolutionary analysis.

To identify genes that experienced substantial changes in selective pressure along the branch leading to the herbivorous drosophilids, we used codon-based models of evolution (*codeml* program) in the Phylogenetic Analysis by Maximum Likelihood (PAML) package (Yang 2007). These models estimate the non-synonymous (dN) to synonymous (dS) substitution rate ratio (ω = dN/dS), wherein protein-coding genes experiencing positive, directional selection may accumulate a significant number of amino acid substitutions resulting in dN/dS>1, those evolving neutrally dN/dS ≈ 1, and those experiencing negative or purifying selection dN/dS<1.

For each group of orthologous genes, nucleotide sequences from the nine taxa (*S. flava, S. montana, S. graminum, S. pallida, S hsui, D. grimshawi, D. mojavensis,* and *D. virilis*, and *D. melanogaster*) were aligned in Geneious v.10.2.6 using the MAFFT v7.450 translation alignment program with default settings (Katoh & Standley 2013; Katoh et al 2002). To generate a species tree for these analyses, we summarized the 100 maximum likelihood-generated gene trees published in (Kim et al. 2021), which included all our species of interest (Kim et al. 2021). To do this, we used ASTRAL v.5.5.9 (Zhang et al. 2018), run with default settings. For paralogous genes, instead of using this species tree, we estimated gene trees using FastTreeMP v2.1 with a GTR+gamma nucleotide substitution model and 100 bootstrap replicates (Price et al. 2010).

We employed branch, branch-site, and clade models within PAML. The branch model allows dN/dS to vary across branches, assuming the same rates across sites (Yang 1998). To assess changes in selection pressures across branches, we compared

a “two-ratio” branch model, where dN/dS is estimated separately for a specified foreground branch and background branches (model = 2, NSites = 0), against a null model (“M0”), which estimates a single dN/dS over all branches (model = 0, NSsites = 0). This test specifically tests whether foreground and background dN/dS rates differ, which could be attributed to positive selection or reduced purifying selection. To specifically test whether the foreground branch has experienced positive selection (dN/dS>1), we compared the two-ratio model to a “constrained two-ratio” model, which includes the same parameters but ω is fixed to 1 (model = 2, NSites = 0, fix_omega = 1). Branch models may be unable to detect positive selection if it acts only at a few sites in the protein because the remaining sites are likely to remain under purifying selection. Therefore, we also used branch-site models, which more realistically model protein evolution, allowing dN/dS to vary across both codons and branches (Yang & Nielsen 2002; Zhang et al. 2005). We compared the “branch-site model A” (model =2, NSsites = 2, fix_omega = 0) against a null model that differs only in that ω is fixed to 1 (model = 2, NSsites = 2, fix_omega = 1). This comparison offers a direct test for positive selection. To identify positively selected amino acids from the branch-site models, we used the Bayes Empirical Bayes (BEB) method for calculating the posterior probability that each site is from a given site class, with site classes 2a and 2b inferred to be under positive selection (Yang et al. 2005).

Finally, we used the “Clade model C” (CmC) to estimate changing selection pressures among gene clades (Weadick & Chang 2012). Based on branch-site models, CmC models allow dN/dS to vary in a proportion of sites across a gene clade, which detects more subtle divergent selection, particularly after duplication events within species (Weadick & Chang 2012). We tested for divergent selection by comparing the CmC model (model = 3, NSsites = 2, fixed = 0) against the null model “M2a_rel” (model = 0, NSsites = 22, fixed = 0), which has the same number of site classes as the CmC model but does not vary across branches. To specifically test for positive selection on the branch leading to the herbivores, we compared the CmC model against a null model in which the ω of the divergent site class is constrained to 1 (model = 3, NSsites = 2, fixed = 1) (Van Nynatten et al. 2015). For clades with foreground divergent sites significantly deviating from the null expectation and multiple paralogs among the herbivore species, we followed up with branch and branch-site tests to test which set of paralogs experienced divergent or specifically positive selection.

For all PAML analyses, model comparisons were made using a likelihood ratio test with a χ^2 distribution and one degree of freedom. *P*-values were adjusted at a false discovery rate of 5% using the qvalue package in R (Storey et a; 2022).

Results

S. flava genome assembly and annotation

To complement existing genome assemblies of non-herbivorous and herbivorous *drosophilids,* we sequenced, assembled, and annotated the genome of the model herbivore, *Scaptomyza flava*, a leaf-mining, mustard specialist. Our main assembly

(sfla_v2) contained 781 scaffolds with an N50 of 31.83 Mb and an assembly size of 331.7 Mb (Table S2). The inflated size of the assembly – relative to the estimate based on flow cytometry (287 Mb) (Whiteman et al. 2011) – is likely due to the partial, not full, inbreeding of the sequenced line, despite taking additional steps during assembly to remove duplicated haplotypes. Inbreeding of *S. flava* is more challenging compared to *Drosophila* that can be reared on media, given that it requires use of living plants for the development of larvae. Using gene models from other *Drosophila* and RNA transcriptome sequences across *S. flava* life stages, we predicted 17,997 genes in the *S. flava* assembly sfla_v1. Carrying these annotations over to our main assembly sfla_v2, we predicted 12,365 genes. We performed the same carry-over procedures on four other *Scaptomyza* species (*S. pallida, S. hsui, S. montana,* and *S. graminum*) to facilitate annotation of their chemosensory and detoxification gene families (Table S2). The completeness of the assemblies was assessed by comparing the assembly against the BUSCO dipteran database. The genome assembly of *S. flava* is near complete, as are the assemblies of the four other *Scaptomyza* species with values comparable to the published assemblies of the seven other *Drosophila* included in subsequent analyses (*D. melanogaster, D. erecta, D. ananassae, D. pseudoobscura, D. mojavensis, D. virilis,* and *D. grimshawi*) (Fig. S1a). Approximately 98.8% of complete BUSCO gene models (98% single-copy and 0.8% duplicated) were identified from the sfla_v2 assembly, 0.4% were found fragmented, and 0.8% were not found. BUSCO scores were also high for the genome-wide automated annotations in all *Scaptomyza* species (Fig. S1b), although slightly lower than those for the full genome assemblies. The completeness of the genome assemblies and automated annotations provided confidence that we would be able to reliably assess the evolution of our gene families of interest.

Contraction and expansion of chemosensory and detoxification gene families

We analyzed the evolution of chemosensory and detoxification gene families to determine if rates of gain, loss, and/or turnover (cumulative gains and losses) were significantly different between non-herbivorous species and herbivorous *Scaptomyza* (both along branches leading to individual herbivorous species and along the branch at the base of all herbivores). We included in these analyses nine non-herbivorous species to compare against three herbivores (all species included in Fig. 1a), all of which have high-quality genome assemblies, and utilize a range of different feeding substrates with varying degrees of specialization. We manually curated six chemosensory gene families (gustatory receptors, GRs; ionotropic receptors, IRs; olfactory receptors, ORs; odorant binding proteins, OBPs; pickpocket proteins, PPKs; transient receptor potential channels, TRPs) and three detoxification gene families (cytochrome P450s, CYPs; glutathione *S*-transferases, GSTs; UDP-glycosyltransferases, UGTs) (RAXML gene trees: Fig. S2-10).

Overall, herbivorous species had smaller repertoires of detoxification and chemosensory genes than non-herbivorous species. (Fig. 2a, Table S3). To explore how gene family expansion and contraction may have influenced these overall family sizes, we used CAFE, a maximum likelihood phylogenetic approach, to generate models in which the rate of gene turnover was estimated: one rate across herbivorous branches and a second across the remaining branches. By comparing these models against a null

model with a single rate for the entire tree, we were able to test whether rates among herbivorous branches were significantly different from the remaining branches. The rate of gene turnover was significantly higher among herbivores than non-herbivores when all chemosensory genes were considered (λ_{herb}=0.006; $\lambda_{non\text{-}herb}$=0.002; q-value<0.0001) and for all detoxification genes (λ_{herb}=0.004; $\lambda_{non\text{-}herb}$=0.002; q-value<0.01) (Fig. 2b (first panel)). The increase in gene turnover rate among chemosensory genes appears most strongly driven by increased turnover rates among GRs, IRs, ORs, and PPKs, and in detoxification genes by increased turnover of GSTs (Fig. 2b (second panel)).

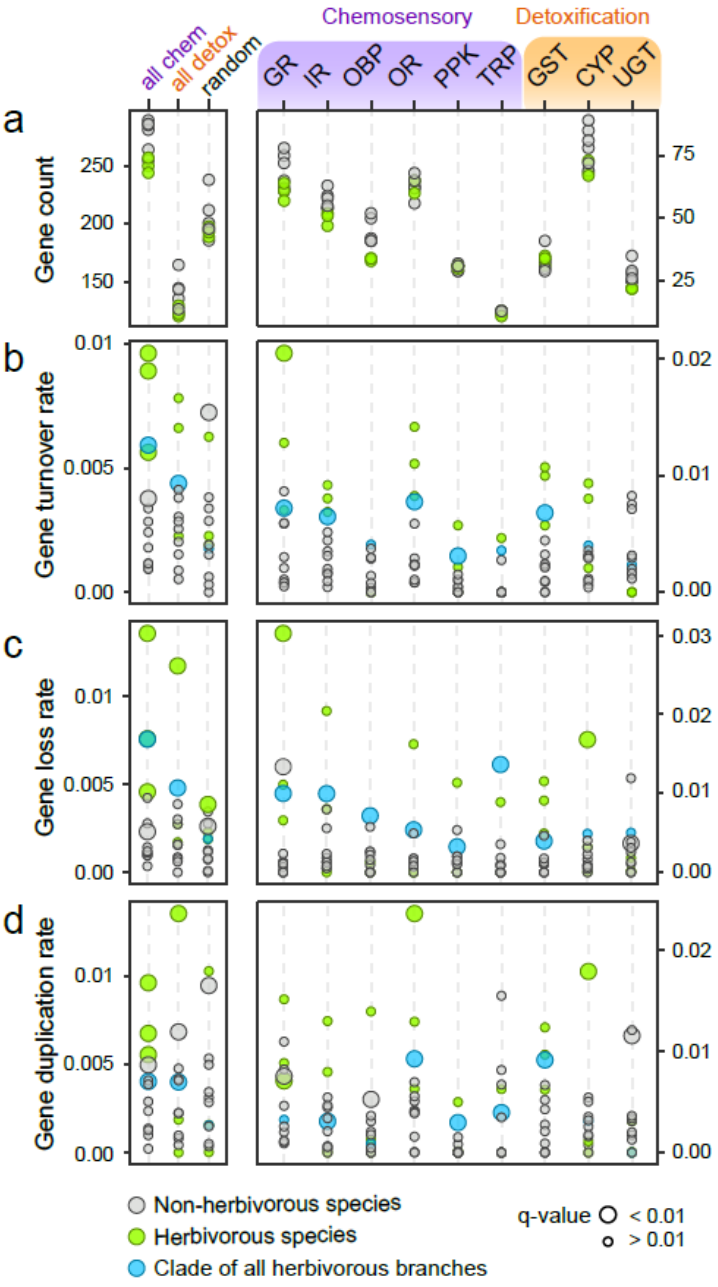


Figure 2. Herbivorous drosophilid lineages exhibit elevated gene turnover within most chemosensory and detoxification gene families. (a) Gene counts per gene family in three herbivorous and nine non-herbivorous *Drosophila*. Evolution of gene family size was estimated by maximum likelihood in CAFE, where the rates of (b) gene turnover (cumulative gains and losses), (c) gene loss, and (d) duplication were estimated for each species. Points represent the rate of a foreground branch from models in which it was allowed to evolve at a rate separate from background branches. Each model was compared to a null model in which a single rate was estimated for the entire phylogeny using a likelihood ratio test at an analysis-wide false discovery rate of 1% (q < 0.01). Foreground branches with significantly higher rates than the rest of the phylogeny are indicated by larger circles. Full details can be found in Tables S3-4.

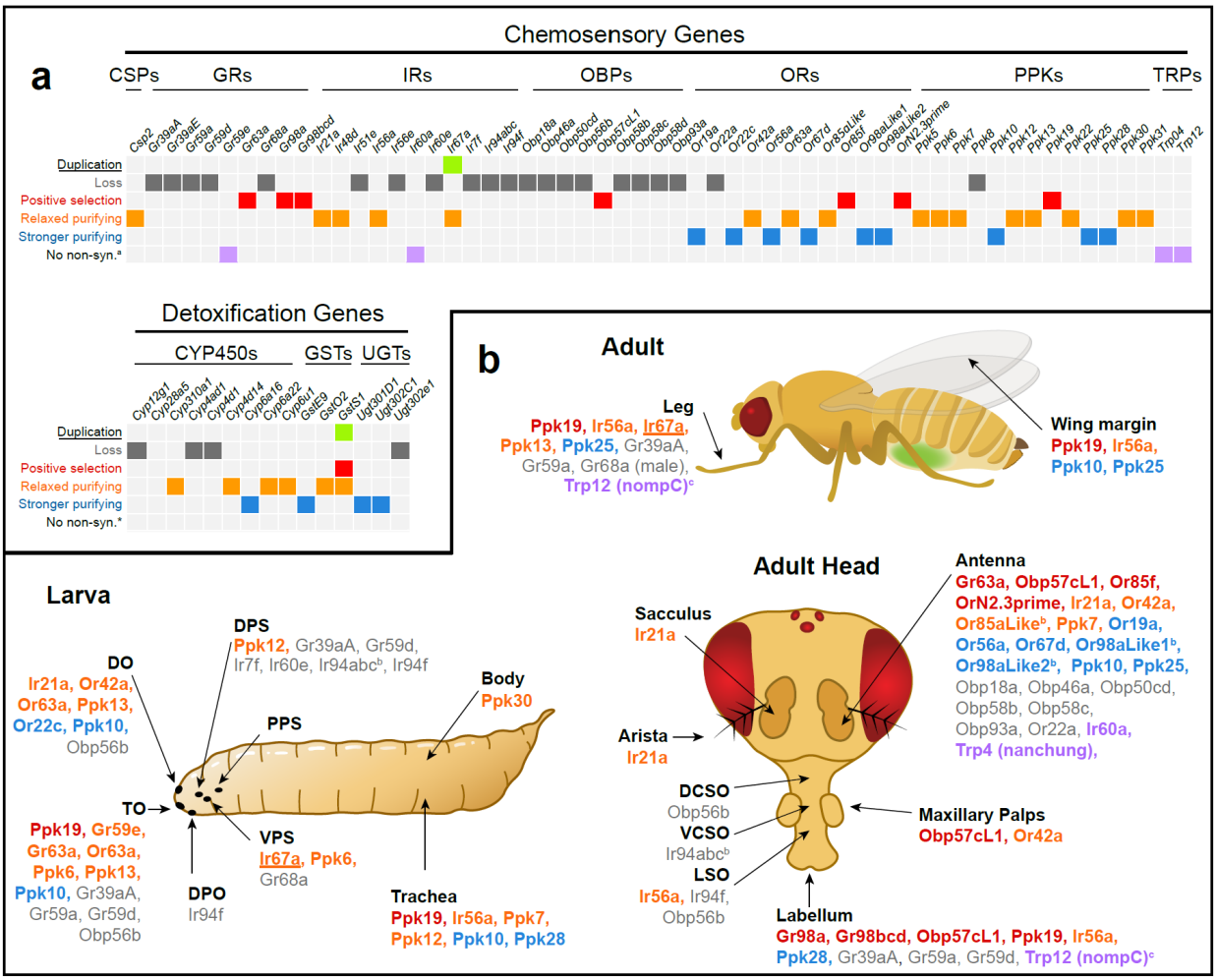
We also estimated rates of gene duplication and loss separately and found that the herbivorous lineages experienced exceptionally high rates of loss among all the chemosensory gene families and GSTs (Fig. 2c), and accelerated rates of gene duplication among the IRs, ORs, PPKs, TRPs, and GSTs (Fig. 2d), which would have been overlooked if we had examined only overall gene counts. Notably, we found evidence for expansions and contractions of several chemosensory genes identified by family-wide and Viterbi *P*-values<0.05 – the former signifies whether an orthology group (e.g. *Gr59d*) has a significantly higher rate of evolution, whereas the latter signifies on which branch the shift

in turnover rate was significant (De Bie et al. 2006). These contractions included OBPs (*Obp18a, Obp58b,* and *Obp58c*) and *Or22a*, an ester-sensitive, yeast volatile receptor, and an expansion of *Ir67a*, which was duplicated in the herbivorous lineage (Table S4). The function of *Ir67a* is not known, but it is expressed in the larval ventral organ and adult legs (Sánchez-Alcañiz et al. 2018), and potentially involved in wing vein development (George et al. 2019). While the number of rapidly expanding orthology groups varies widely across branches of the drosophilid phylogeny, among *Scaptomyza* the ancestral branch leading to the three herbivores has an elevated number of rapidly expanding gene clusters, relative to other branches within the *Scaptomyza* clade as a whole (Fig. S11).

In stark contrast to the CAFE results on chemosensory and detoxification genes, we found evidence that the rate of gene turnover among a random subset of ~200 orthologous clades is significantly lower among herbivores compared to non-herbivores (λ_{herb}=0.002; $\lambda_{non\text{-}herb}$=0.003; q-value<0.02, Fig. 2b). Consistent with this, there were no orthologous groups that were rapidly expanding or contracting (Table S4).

Based on the gene trees we reconstructed for each of the gene families (Fig. S2-S10), we found a surprisingly large number of herbivore-specific losses and a few gains, which are summarized in Fig. 3a (Table S5 for gene counts). The majority of copy number changes are concentrated among the chemosensory gene families (22 out of 27). Herbivores lost several gustatory receptors required for bitter reception or are expressed in bitter gustatory neurons in *D. melanogaster* (paralogs of *Gr39aA*, *Gr59a, Gr59d*) (Weiss et al. 2011; Dweck et al. 2020). Herbivorous lineages also lost *Gr68a*, which is involved in the detection of an anti-aphrodisiac (Bray & Amrein 2003), and *Gr39aE* which has no known ortholog in *D. melanogaster*. The only OR lost was the aforementioned *Or22a*, which has been previously reported (Goldman-Huertas et al. 2015). Among the OBPs, almost all of the herbivore-specific losses (*Obp46a, Obp50cd, Obp58b, Obp58c, Obp58d, Obp93a*) were among the 'plus-C OBPs,' which are characterized as having more than six cysteine residues, and only two losses (*Obp18a* and *Obp56b*) were among the 'classic OBPs' with six cysteines (Fig. S12). Whereas non-herbivorous *Scaptomyza* have 11 Plus-C OBPs; their herbivorous relatives possess only five. Almost all ionotropic receptors that were lost in the herbivores belonged to the 'divergent IR' class (*Ir7f, Ir51e, Ir56e, Ir94abc, Ir94f*), which are expressed in gustatory neurons. The only 'antennal IR' lost across all herbivores was *Ir60a.* Again, the only IR duplication in herbivorous *Scaptomyza* was the divergent IR *Ir67a*. A single loss was found among PPKs (*ppk8*), expressed in the trachea of embryos (Suslak 2018), whose function is otherwise unknown.

There were only five herbivore-specific changes in copy number among the detoxification gene families: the duplication of *GstS1*, which has a role in oxidative stress responses and in flight muscle structure (Singh et al. 2001); the loss of *Ugt302e1*, whose function is currently unknown, and three cytochrome P450s (*Cyp4ad1, Cyp4d1,* and *Cyp12g1*). While no ortholog of *Cyp12g1* is known in *D. melanogaster*, both *Cyp4ad1* and *Cyp4d1* are upregulated in response to the pyrethroid deltamethrin (Liu et al. 2020).



(a) Alternative model was significantly better than null based on LRT, but there were no non-synonymous changes -- dN/dS thus could not be estimated.
(b) Orthologs not present in *D. melanogaster*. Expression data from nearest related gene(s): Or85a, Or98a, Or94a/b/c.
(c) Trp12 (nompC) is expressed in mechanoreceptive bristles, chordotonal organ of haltere, and Johnston organ.

Figure 3. Chemosensory and detoxification genes duplicated, lost, or that have experienced significant changes in selection pressure in the branch at the base of the herbivorous *Scaptomyza* lineage. (a) Gene duplications and losses are based on gene trees generated through RAXML, and genes under divergent or positive selection were identified from PAML selection analyses. Gene counts from each species are shown in Table S5. (b) Localization map of chemosensory gene products in adults and larvae, based on published expression data from *D. melanogaster*. Only shown are those that have been lost in all herbivores (gray), duplicated (underlined), or experienced relaxed purifying selection (orange) or positive selection (magenta). Dorsal organ (DO); terminal organ (TO); dorsal, ventral, and posterior pharyngeal sense organ (DPS, VPS, PPS, respectively); labral sense organ (LSO); ventral cibarial sense organ (VCSO).

Identifying genes under positive selection in the herbivore lineage

We next used a maximum likelihood approach through PAML to identify chemosensory and detoxification genes that may have been subjected to episodic positive selection or significant relaxation from purifying selection in the branch leading to the herbivorous species. We focused on this ancestral herbivore branch, rather than those leading to individual herbivore species, to identify changes that occurred as a result

of herbivory, rather than subsequent specialization on mustard or carnation plants. We used codon-based tests for positive selection, estimating dN/dS values under branch, branch-site, and clade models. We report significant results from branch and branch-site tests below, but for clade models, we only discuss genes that were retained after additional follow-up branch and branch-site tests, as described below. Clade models offer more flexibility/sensitivity in detecting divergent selection among a clade of genes, particularly those with multiple recent paralogs undergoing complex evolution (Weadick & Chang 2012). Thus unsurprisingly, a large number of genes initially showed evidence of divergent selection: 128 genes (90 chemosensory and 38 detoxifying, $P < 0.05$; false discovery rate <5%). In these cases, we additionally tested branch and branch-site models on branches leading to different paralogs to identify which experienced significant changing selection pressure. Tables 1 and 2 summarize all chemosensory and detoxification genes, respectively, wherein dN/dS was significantly different on an herbivore-specific branch compared to background branches. Table 3 summarizes gene function of orthologs in *D. melanogaster.* A total of 38 genes were identified through follow-up branch tests and 12 genes through branch-site tests. Full model details on all 50 genes can be found in Table S6.

Table 1. Summary of selection analyses on chemosensory genes.

	Gene	Model[a]		dN/dS[b]			LRT 2(ΔlnL)	*q-value (FDR 5%)*	Positively selected sites[c]
CSP	*Csp2*	M0 vs 2-ratios	ω0=	0.06	ω1=	0.24	7.14	0.05	
GR	*Gr59e*	M0 vs 2-ratios	ω0=	0.23	ω1=	999	18.26	<0.001	
	Gr63a	Model A vs Model A (ω = 1)	ω0=	0.03 (81%)	ω1=	1 (7%)	16.52	<0.01	V13 (0.62), A17 (0.69), N66 (0.69), V115 (0.7), H134 (0.69), A278 (0.69), T412 (0.69)
			ω2a=	1 (11%)	ω2b=	1 (1%)			
	Anc*(SmonGr98a1, SflaGr98a1, SgraGr98a2)*	M0 vs 2-ratios	ω0=	0.29	ω1=	0.87	7.74	0.04	
		Model A vs Model (ω = 1)	ω0=	0.22 (71%)	ω1=	1 (26%)	11.4	0.03	L265 (0.69), V275 (0.52), V282 (0.52), D320 (0.93), Q325 (0.95)
			ω2a=	32.44 (2%)	ω2b=	32.44 (1%)			
	Gr98bcd	Model A vs Model (ω = 1)	ω0=	0.16 (73%)	ω1=	1 (27%)	10.2	0.05	V49 (0.84), F154 (0.5), L377 (0.9)
			ω2a=	179.55 (0.004%)	ω2b=	179.55 (0.001%)			
IR	*Ir21a*	M0 vs 2-ratios	ω0=	0.11	ω1=	0.26	9.2	0.02	
	Ir48d	M0 vs 2-ratios	ω0=	0.18	ω1=	0.38	7.96	0.04	
	Anc(*SflaIr56a, SmonIr56a*)	M0 vs 2-ratios	ω0=	0.33	ω1=	0.98	18.8	<0.001	
	Anc(*SgraIr56a1, SgraIr56a2, SgraIr56a4, SgraIr56a5*)	M0 vs 2-ratios	ω0=	0.33	ω1=	3.66	7.98	0.03	
		2-ratios vs 2 ratios (ω = 1)	ω0=	0.34	ω1=	1	1.56	0.31	
	Ir60a	Model A vs Model A (ω = 1)	ω0=	0.06 (83%)	ω1=	1 (16%)	10	0.05	
			ω2a=	999 (0.005%)	ω2b=	999 (0.001%)			
	Ir67a	M0 vs 2-ratios	ω0=	0.3	ω1=	0.6	10.6	0.01	

OBP	Obp57cL1	Model A vs Model A (ω = 1)	ω0=	0.21 (59%)	ω1=	1 (36%)	15.68	<0.01	K5 (0.62), **L6 (0.99)**, Q51 (0.84), L95 (0.83), G100 (0.62), P140 (0.67)
			ω2a=	148.51 (3%)	ω2b =	148.51 (2%)			
OR	Or19a	M0 vs 2-ratios	ω0=	0.3	ω1=	0.09	13.7	<0.01	
	Or22c	M0 vs 2-ratios	ω0=	0.18	ω1=	0.05	7.24	0.05	
	Anc(*SflaOr42a1, SmonOr42a1, SgraOr42a1*)	M0 vs 2-ratios	ω0=	0.13	ω1=	0.45	11.46	<0.01	
	Or56a	M0 vs 2-ratios	ω0=	0.11	ω1=	0.03	8.98	0.02	
	Or63a	M0 vs 2-ratios	ω0=	0.2	ω1=	0.76	29.5	<0.0001	
	Or67d	M0 vs 2-ratios	ω0=	0.18	ω1=	0.03	15.18	<0.01	
	Or85aLike	M0 vs 2-ratios	ω0=	0.13	ω1=	0.29	6.98	0.05	
	Or85f	Model A vs Model A (ω = 1)	ω0=	0.13 (88%)	ω1=	1 (11%)	12.54	0.03	**A99 (0.99)**, S270 (0.94)
			ω2a=	46.61 (1%)	ω2b =	46.61 (0.001%)			
	Or98aLike1	M0 vs 2-ratios	ω0=	0.19	ω1=	0.05	7.58	0.04	
	Or98aLike2	M0 vs 2-ratios	ω0=	0.2	ω1=	0	35.92	<0.0001	
	OrN2.3prime	Model A vs Model A (ω = 1)	ω0=	0.16 (73%)	ω1=	1 (20%)	15.6	<0.01	V64 (0.51), Y77 (0.94), I97 (0.69), S114 (0.7), I170 (0.83), V211 (0.82), E222 (0.51), I230 (0.95), N253 (0.72), L310 (0.5), **S341 (0.98)**, A356 (0.92), **V388 (0.97)**
			ω2a=	15.3 (5%)	ω2b =	15.3 (1%)			
PPK	*Ppk10*	M0 vs 2-ratios	ω0=	0.1	ω1=	0.01	12.12	<0.01	
	Ppk12	M0 vs 2-ratios	ω0=	0.18	ω1=	0.38	10.07	0.02	
	Ppk13	M0 vs 2-ratios	ω0=	0.05	ω1=	0.19	16.12	<0.01	
	Ppk19	Model A vs Model A (ω = 1)	ω0=	0.11 (75%)	ω1=	1 (23%)	13.98	0.01	H42 (0.75), **P113 (0.96)**, Y203 (0.92), M264 (0.8), T271 (0.91), V394 (0.9), T439 (0.84), T468 (0.9)
			ω2a=	15.75 (2%)	ω2b =	15.75 (1%)			
	Ppk22	M0 vs 2-ratios	ω0=	0.16	ω1=	0.37	7.7	0.04	
	Ppk25	M0 vs 2-ratios	ω0=	0.21	ω1=	0.04	16.82	<0.01	
	Ppk28	M0 vs 2-ratios	ω0=	0.1	ω1=	0.03	11.72	<0.01	
	Ppk30	M0 vs 2-ratios	ω0=	0.2	ω1=	0.44	12.2	<0.01	
	Ppk31	M0 vs 2-ratios	ω0=	0.12	ω1=	0.39	17.23	<0.01	
	Ppk5	M0 vs 2-ratios	ω0=	0.11	ω1=	0.36	15.16	<0.01	
	Ppk6	M0 vs 2-ratios	ω0=	0.09	ω1=	0.25	10.35	0.01	
	Ppk7	M0 vs 2-ratios	ω0=	0.15	ω1=	0.27	7.75	0.04	
TRP	*Trp04*	M0 vs 2-ratios	ω0=	0.03	ω1=	0.09	10.66	0.01	
		Model A vs Model A (ω = 1)	ω0=	0.01 (94%)	ω1=	1 (6%)	12.2	0.03	
			ω2a=	999 (0.004%)	ω2b =	999 (0.0002%)			

Trp12	Model A vs Model A (ω = 1)	ω0=	0.01 (97%)	ω1=	1 (3%)	11.8	0.03	
		ω2a=	999 (0.001%)	ω2b =	999 (0.00003 %)			

[a] Null model against Model A is Model A with fixed divergent sites (ω = 1).

[b] Omega values are reported only for the alternative model. Detailed results can be found in Table S6.

[c] Based on Bayes Empirical Bayes. Posterior probabilities are shown in parentheses with those P > 0.95 shown in bold.

Table 2. Summary of selection analyses on detoxification genes

	Gene	Model[a]	dN/dS[b]				LRT 2(ΔlnL)	*q*-value *(FDR 5%)*	Positively selected sites[c]
GST	*GstE9*	M0 vs 2-ratios	ω0=	0.15	ω1=	0.01	9.35	0.02	
	GstO2	M0 vs 2-ratios	ω0=	0.12	ω1=	0	10.3	0.01	
	Anc (*SflaGstS1b, SmonGstS1b, SgraGstS1b*)	M0 vs 2-ratios	ω0=	0.09	ω1=	3.27	50.43	<0.0001	
		Model A vs Model A (ω = 1)	ω0=	0.08 (84%)	ω1=	1 (2%)	14.04	0.01	**V110 (0.99), I114 (0.99), F119 (0.99), C127 (0.97), L149 (1), S155 (1), E159 (0.99), D160 (0.99), L167 (0.99), T169 (0.99), N171 (0.96), D190 (0.99), L193 (0.99), I208 (0.99), M212 (0.94), Y214 (0.99), V216 (0.99), V234 (0.9), V252 (0.99)**
			ω2a=	38.47 (13%)	ω2b=	38.47 (0.004%)			
CYP	*Cyp28a5*	M0 vs 2-ratios	ω0=	0.13	ω1=	0.05	9.95	0.02	
	Cyp310a1	M0 vs 2-ratios	ω0=	0.19	ω1=	0.57	18.43	<0.001	
	Cyp4d14	Model A vs Model A (ω = 1)	ω0=	0.06 (84%)	ω1=	1 (15%)	11.44	0.03	A153 (0.51), -272 (0.84), Y414 (0.53), **T470 (0.99)**
			ω2a=	809.22 (1%)	ω2b=	809.22 (0.001%)			
	Cyp6a16	M0 vs 2-ratios	ω0=	0.21	ω1=	0.1	7.57	0.04	
	Cyp6a22	M0 vs 2-ratios	ω0=	0.07	ω1=	0.26	21.17	<0.001	
	Cyp6u1	M0 vs 2-ratios	ω0=	0.14	ω1=	0.61	20.06	<0.001	
UGT	*Ugt301D1*	M0 vs 2-ratios	ω0=	0.09	ω1=	0.03	7.22	0.05	
	Ugt302C1	M0 vs 2-ratios	ω0=	0.11	ω1=	0.04	7.42	0.05	

[a] Null model against Model A is Model A with fixed divergent sites (ω = 1).

[b] Omega values are reported only for the alternative model. Detailed results can be found in Table S6.

[c] Based on Bayes Empirical Bayes. Posterior probabilities are shown in parentheses with those *P* > 0.95 shown in bold.

Among the 50 divergently selected genes, nine were identified as being under positive selection (dN/dS >1) in the herbivore lineage: *Gr63a, Gr98a, Gr98bcd, GstS1, Obp57cL1, Or85f, OrN2.3prime, Cyp4d14*, and *Ppk19* (Tables 1-2). The proportion of positively selected sites among these genes was generally low, with a few exceptions. One of these exceptions was *GstS1b* that had 13% positively selected sites (dN/dS =38.47) with almost all of these sites having exceptionally strong support (i.e., high posterior probabilities, *P*>0.95). As noted earlier, *GstS1* was duplicated in the ancestor of the three herbivores included here. Despite expression of both *GstS1a* and *GstS1b* in the *S. flava* larval gut (Gloss et al. 2019), we only found that *GstS1b* was rapidly evolving in the branch at the base of herbivores. However, in response to the mustard-specific toxin isothiocyanates, neither paralog was up-regulated in larvae, suggesting it might detoxify other plant compounds (Gloss et al. 2019). *GstS1* in *D. melanogaster* has been found to

be involved in oxidative stress response, conjugation of the lipid peroxidation product 4-HNE, and modulating methylmercury toxicity (Whitworth et al. 2005, Singh et al. 2001; Vorojeikina et al. 2017). Notably, isothiocyanates do induce the formation of peroxidized lipids (Gago-Dominguez et al. 2007). *GstS1* is expressed during development and in adults in the central nervous system (Whitworth et al. 2005). However, it is most highly expressed in flight muscle in *D. melanogaster*, which is unusual for a GST, and it has been hypothesized that it plays a structural role (Clayton et al, 1998).

Gr63a also had a relatively high number of positively selected sites in the herbivorous clade (12%). It is an unusual GR in that it is expressed in adult olfactory neurons and is a receptor for carbon dioxide (Kwon et al. 2007). While the yeast-feeding *D. melanogaster* displays context-dependent attraction/aversion to CO_2, the rapid accumulation of amino acid changing substitutions in *Gr63a* among herbivores suggests that feeding on living leaves may provide a starkly different interaction with environmental CO_2 levels compared to feeding on fermenting fruit.

Several of the other herbivore-specific positively selected genes have gustatory roles. *Gr98a* and *Gr98bcd* are both involved in the detection of noxious, 'bitter' compounds – *Gr98a* in detecting histamine which is high in fermented foods (Aryal & Lee 2022) and *Gr98b* in detecting L-canavanine, a plant-derived toxic amino acid (Shim et al. 2015). (Note that three genes, *Gr98b, Gr98c,* and *Gr98d*, are present in *D. melanogaster*, whereas only a single copy *Gr98bcd*, sister to the three *D. melanogaster* orthologs, is present in all species in the subgenus *Drosophila*). *Ppk19* plays a key role in detecting high salt concentrations in both larvae and adults (Liu et al 2003; Alves et al. 2014). And *Obp57cL*, although it is not present in *D. melanogaster*, is closely related to *Obp57c*, which is the only OBP expressed in all olfactory sensilla to our knowledge (Galindo & Smith 2001) and is also enriched in gustatory neurons in the labellum (Jeong et al. 2013). However, expression data from the herbivore *S. flava* and non-herbivore *S. pallida* suggest that Obp57cL1 is only expressed in gustatory organs (labellum and tarsi) (Suzuki et al. unpublished). The functions of *Obp57c* and *Obp57cL1* remain unknown.

Among genes that did not show a strong indication of positive selection but were found to be evolving at significantly different rates between herbivores and background branches, many (29/41) had higher foreground dN/dS values (Tables 1-2), indicating relaxation from selective constraints among the herbivores, whereas the remaining experienced stronger purifying selection.

Many olfactory receptors in *D. melanogaster* have been de-orphanized, providing a wealth of information about the breadth and sensitivity of individual ORs (e.g., Hallem et al. 2004). Though the properties of *D. melanogaster*'s ORs cannot be used to precisely determine those of other species, it offers a starting point to generate hypotheses about how ORs are evolving in closely related species. For ORs that experienced divergent selection in the herbivores and that have been de-orphanized, we utilized the DoOR database (Münch & Galizia 2016) to investigate patterns in the types of ligands that activate these receptors. We found that ORs experiencing relaxed purifying selection (*Or42a, Or85aLike*), stronger purifying selection (*Or98aLike1*, *Or98aLike2*), and also

gene loss (*Or22a*) detect esters and alcohols (Fig. S13), which are produced in high abundance by yeasts and fruit, and are attractive to microbe-feeders (Becher et al. 2012), but not herbivores (Goldman-Huertas et al. 2015; Matsunaga, Reisenman et al. 2021).

Interestingly, *Ppk28* shows stronger negative selection along the branch basal to all herbivores (ω_{bkgrd} = 0.1, ω_{frgrd} = 0.03, q<0.01). This gene is involved in sensing water or low osmolarity in gustatory neurons (Cameron et al. 2010). Conservation of this gene may be critical for herbivorous *Scaptomyza* species that are more typically found in habitats near flowing water. Moreover, larvae are essentially in a liquid environment while creating mines in the leaf mesophyll. Similarly, *Ppk25* showed stronger purifying selection in the basal herbivorous branch (*Ppk25*: ω_{bkgrd} = 0.1, ω_{frgrd} = 0.03, q<0.01). Expressed in a wide range of chemosensory tissues (larval chemosensory organs, adult wings, legs, coeloconic antennal sensilla) (Menuz et al. 2014; Liu et al 2003; He et al 2019; Lin et al. 2005; Fig. 3b), *Ppk25* is notably involved in the detection of the female pheromone 7,11-heptacosadiene (Liu et al. 2018). Additionally, it has also been implicated in noxious cold detection, along with *Ppk12* and *Ir21a*, which we found both experienced relaxed purifying selection in the ancestral herbivore lineage (Table 1) (Williamson 2015; Ni et al. 2016). How these molecular changes (or lack thereof) affect water sensing, pheromone detection, or cold tolerance in herbivores remains to be investigated.

Detoxification genes that experienced relaxed purifying selection in the herbivore lineage were *Cyp310a, Cyp6a22,* and *Cyp6u1*, and those that experienced. stronger purifying selection were *GstE9, GstO2, Cyp28a5, Cyp6a16, Ugt301d1,* and *Ugt302c1* (Table 2). The majority of these have been implicated in detoxification, although some have putative functions in development, hormone metabolism, cold tolerance, and olfaction (Table 3).

Table 3. Gene functions of candidate chemosensory and detoxification genes duplicated, lost, or having experienced positive selection in the ancestral lineage of herbivorous *Scaptomyza*.

Gene	Result	Gene Function in *D. melanogaster*	References
Bitter Detection			
Gr39aA	loss	Involved in the detection of many bitter compounds	(H. K. M. Dweck & Carlson, 2020; Sung et al., 2017)
Gr59a	loss	Putative bitter reception: expressed in labellar S-a and S-b type bitter gustatory neurons	(Weiss et al., 2011)
Gr59d	loss	Bitter detection in larvae. Expressed in labellar S-a and I-a type bitter gustatory neurons	(H. Kim et al., 2016; Weiss et al., 2011)
Gr98a	dN/dS>1	Detects the toxic amino acid histamine, found in high amounts in fermented foods	(Aryal & Lee, 2022)
Gr98bcd	dN/dS>1	Gr98b is required for detection of the toxic plant-derived amino acid L-canavanine	(Shim et al., 2015)
Yeast/fruit volatile detection			
Obp18a	loss	Involved in detecting diverse yeast/fruit volatiles (propanol, benzaldehyde, citral, 2-heptanone, isoamylacetate, methyl salicylate, phenyl-ethanol, d-carvone). Down-regulated in *D. sechellia*	(Dworkin & Jones, 2009; Swarup et al., 2011)
Obp58b	loss	Involved in detecting diverse yeast/fruit volatiles (1-hexanol, 2-ethylpyrazine, d-carvone, isoamylacetate, methyl salicylate)	(Swarup et al., 2011)
Obp58c	loss	Involved in detecting diverse yeast/fruit volatiles (1-hexanol, geraniol, 2-heptanone)	(Swarup et al., 2011)
Obp93a	loss	Involved in detecting diverse yeast/fruit volatiles (1-hexanol, geraniol, ethylpyrazine, benzaldehyde)	(Swarup et al., 2011)
Or22a	loss	Detects a range of esters and alcohols produced during fermentation by yeast	(Stensmyr et al., 2003)
Pheromone detection			
Gr39aA	loss	Mutant males show reduced courtship	(Watanabe et al., 2011)

Gr68a	loss	Detection of female sex pheromone	(Bray & Amrein, 2003)
Other chemosensory detection			
Gr63a	dN/dS>1	Carbon dioxide detection	(Kwon et al., 2007)
Ppk19	dN/dS>1	Detection of high salt concentrations in both larvae and adults	(Alves et al., 2014; L. Liu et al., 2003)
Or85f	dN/dS>1	Detects parasitoid wasp (*Leptopilina*) volatiles. Best ligand is the ketone acetophenone.	(Ebrahim et al., 2015)
Detoxification of toxins			
GstS1	duplication & dN/dS>1	Detoxification of lipid peroxidation product, 4-HN, oxidation product of adrenalin, adrenochrome, and methylmercury. Expressed in CNS and during development, but highest in flight muscle.	(Clayton et al., 1998; Saisawang et al., 2012; Singh et al., 2001; Vorojeikina et al., 2017)
Cyp4ad1	loss	Up-regulated in response to insecticide pyrethroid deltamethrin. Down-regulated by ecdysteroid agonists. Expressed in larval gonads.	(Davies et al., 2006; Y. Liu et al., 2020)
Cyp4d1	loss	Up-regulated in response to insecticide pyrethroid deltamethrin. Down-regulated by ecdysteroid agonists. Expressed in midgut and fat body. The only P450 that exhibits alternate splicing.	(Chung et al., 2009; Davies et al., 2006; Y. Liu et al., 2020)
Cyp4d14	dN/dS>1	Up-regulated in response to phenobarbital, caffeine, mycotoxins. Expressed in larval midgut. Also lost in *D. sechellia.*	(Chung et al., 2009; Rane et al., 2019; Sun et al., 2006; Trienens et al., 2017; Willoughby et al., 2006)
Other functions			
Ir67a	duplication	Wing vein development	(George et al., 2019)
Ppk8	loss	Involved in regulating neuronal excitability through interactions with the gene seizure. Expressed in embryonic trachea and adult fat body.	(Suslak, 2015; K. Xu et al., 2011; Yehuda, 2012)

Unknown Function in *Drosophila*									
Gr39aE	loss	*Ir60e*	loss	*Ir94f*	loss	*Obp56b*	loss	*OrN2.3prime*	dN/dS>1
Ir56e	loss	*Ir7f*	loss	*Obp46a*	loss	*Obp57cL1*	dN/dS>1	*Cyp12g1*	loss
Ir51e	loss	*Ir94abc*	loss	*Obp50cd*	loss	*Obp58d*	loss	*Ugt302e1*	loss

Effects of demographic history

Demographic processes, such as population bottlenecks, can weaken the efficacy of natural selection, leading to an accelerated fixation of slightly deleterious gene gains and losses (Gardiner et al. 2008). To investigate the possibility that our results on gene family evolution were confounded by demographic events, we examined three lines of evidence from our analyses. First, as discussed above, the randomly selected set of orthologous clusters (~200 genes) exhibited gain, loss, and turnover rates that were not significantly higher than those observed in other *Drosophila* (Fig. 2). These results suggest that although we cannot eliminate or control for their effects in our analyses, demographic processes were not the underlying cause of elevated rates of turnover among chemosensory and detoxification genes because these processes would have generated similar patterns genome-wide. Second, we inferred the level of nucleotide diversity (π) in a population of one of the herbivore species, *S. flava*, using pooled whole genome sequencing (methods described in the Supplemental Methods). Autosomal nucleotide diversity, which is proportional to the coalescent effective population size, was similar between a single population of *S. flava* (π = 0.0056) and North American populations of *D. melanogaster* (the DGRP; π = 0.0060 (Mackay et al. 2012)). The relationship between physical distance and linkage disequilibrium was also similar between the two species and decayed quickly (Peláez et al. 2020), consistent with sharing similarly large coalescent effective population sizes (Sved 1971).

Horizontal gene transfer

The acquisition of novel genes via horizontal gene transfer (HGT) can also shape genome content and gene family evolution. The genomes of diverse arthropod herbivores and nematode plant parasites encode horizontally transferred genes, principally from microbes, that facilitate acquisition and digestion of plant tissues and resistance to plant toxins (Fraenkel 1959; Ehrlich & Raven 1964; Swain 1977; Berenbaum & Zangerl 2008). We recently used a genome-wide sequence similarity screen to search for such genes in the *S. flava* genome assembly (Verster et al. 2019), and we expanded upon that search here with a more liberal approach (see Supplemental Methods). We identified no new evidence for HGT beyond the single, previously-described HGT event (Verster et al. 2019): the toxin-encoding *cytolethal distending toxin subunit B* (*cdtB*) gene suspected to defend flies against parasitoids and unlikely to aid in digestion or detoxification.

Discussion

In this study, we sought to determine the extent to which the evolution of herbivory affected patterns of molecular evolution across chemosensory and detoxification gene families in the Drosophilidae. We focused in particular on a clade within the genus of *Scaptomyza* – nested within the paraphyletic *Drosophila* – that evolved to feed as leafminers of living plants ~15 million years ago. A great deal is now known regarding how these gene families, especially chemosensory ones, evolve across *Drosophila* in the context of diet shifts and diet specialization. However, we have far less insight into how these gene families evolve in response to a niche shift to a truly herbivorous life history (i.e. larval development is completed by feeding on living plants).

Here, we identified, through an iterative curation approach, including manual inspection, the full complement of chemosensory and detoxification gene families for five *Scaptomyza* species – three herbivorous species within the subgenus *Scaptomyza* (*S. flava*, *S. montana,* and *S. graminum*) and two non-herbivorous species representing two other subgenera (*S. pallida* in *Parascaptomyza* and *S. hsui* in *Hemiscaptomyza*) – and more distantly related *Drosophila* species. By including species across a range of diets and phylogenetic distances, we were able to disentangle *Scaptomyza*-wide effects on genome evolution from herbivore-specific or host-plant specific specialization changes among Brassicales-feeding *Scaptomyza* (*S. flava* and *S. montana*) and Caryophyllaceae-feeding *Scaptomyza* (*S. graminum*).

We recognize, however, that ideally a study of this type should include multiple instances (convergence) of the evolution of herbivory to more clearly delineate patterns of adaptive evolution from neutral changes that drift to fixation. Additional herbivorous drosophilid species that could be included in future analyses include the leaf-mining species of the *Lordiphosa* genus (Atallah et al. 2012), for which there are already several genome assemblies now available (Kim et al. 2021), and bracken fern-feeding *Scaptodrosophila* species (Thomson et al. 1982). Nonetheless, there is currently a growing wealth of genetic data and analyses on these gene families from many herbivorous and specialist drosophilids and other insects, which we can draw on to infer generalizations about how these gene families are reshaped by the evolution of herbivory.

We discuss here the potential implications of specific genetic changes, assuming functional homology with those orthologs in *D. melanogaster*.

Concerted losses of and positive selection on genes involved in olfaction

Gene family sizes among herbivorous species were similar to non-herbivorous species (Fig 2a, Table S3), suggesting that there is overall conservation of function among these genes and similar mechanisms of chemosensory detection and detoxification of xenobiotics. However, there were significant reductions in gene family size among all *Scaptomyza* herbivores, and our phylogenetic analyses revealed that rates of gene duplication but more prominently gene loss were elevated among these gene families in the basal herbivore branch (Fig. 2). These patterns among chemosensory genes are consistent with the neural limitation hypothesis, which suggests that the loss of chemosensory genes, by limiting sensory inputs, facilitates rapid and accurate decision-making in herbivores that are faced with many host plant options (Bernays 2001). Gene losses likely occurred as these genes to detect diverse compounds were no longer needed, resulting in relaxed purifying selection and the accumulation of deleterious, non-synonymous changes. To decipher whether these losses are associated with a niche shift, an important question is whether genes that share related, ecologically important functions are lost concertedly during these transitions (Robertson 2019; McBride 2007).

Goldman-Huertas et al. (2015) showed that the evolution of herbivory in *S. flava* was associated with OR gene losses that reduced attraction towards their ancestral diet of yeast-feeding: through a behavioral loss of attraction to yeast volatiles, putatively ancestral feeding attractants, and with the step-wise loss of olfactory receptors (*Or22a*, *Or42b*, and *Or85d*) tuned to detect them across the evolution of herbivory in *Scaptomyza (Goldman-Huertas et al. 2015)*. We broadened the scope of these analyses to now include full annotations from four other *Scaptomyza* and our results are consistent with this finding (Fig. S5). However, we also find evidence through tests for selection that there was relaxed selective constraint on additional ORs putatively involved in detecting yeast volatiles (*Or42a* and *Or85aLike,* respectively*;* Table 2, Fig. S13). Furthermore, in line with past studies on *Or22a* in various species, the loss of *Or22a*, which was the only OR loss shared exclusively by all herbivores in this study, strongly indicates an outsized role for *Or22* in association with diet shifts. *Or22a* and *Or22b* (sometimes found as a single copy in some lineages of *Drosophila*) have been subjected to repeated, independent bouts of natural selection, resulting in: increased expression and sensitivity to esters specific to noni fruit (*Morinda citrifolia*) in *D. sechellia (Dekker et al. 2006)*; increased sensitivity to *Pandanus* fruit volatiles in *D. erecta (Linz et al. 2013)*; attraction to marula fruit (*Sclerocarya birrea*) in some populations of *D. melanogaster*; and finally, loss of sensitivity towards fermentation odors and increased sensitivity to certain leaf volatiles in the ripe fruit specialist *D. suzukii (Keesey et al. 2015)*. Collectively, this indicates the importance of *Or22* to drive attraction to fermenting plant tissues in drosophilids, which is no longer a niche used by ovipositing females of herbivorous species, like herbivorous *Scaptomyza* and *D. suzukii.*

The most striking herbivore-specific loss of genes occurred in a gene family predominantly expressed in olfactory neurons, the odorant binding proteins. OBPs have been classified into three groups: classic OBPs that have six cysteine residues, minus-C with less than six, and plus-C with more than six (Zhou et al. 2006). The majority of OBPs lost in herbivores (six out of eight) were among the Plus-C class (Table S5, Figure S11). Though the functional significance of bearing additional cysteine residues is unclear, we speculate that the Plus-C OBPs may be more vulnerable to damage by plant toxins, like the Brassicales-specific isothiocyanates, which are highly electrophilic and attack the nucleophilic sulfhydryl moieties of the cysteine residues (Kawakishi & Kaneko 1987). Some of the OBPs lost in herbivores (*Obp58b, Obp58c,* and *Or85a*) are expressed only in the antennae and/or head of *D. melanogaster*, suggesting an olfactory role, but other lost OBPs have been found additionally in the legs or body of adults (*Obp46a, Obp50c, Obp93a*), suggesting non-olfactory functions (Zhou et al. 2004). For instance, members of the *Obp50a-d* cluster are expressed in metabolic tissues, the male testis and accessory gland, where they function as carriers for lipophilic compounds (Johnstun et al. 2021). Overall, the losses of Plus-C OBPs and yeast-sensitive ORs among herbivorous *Scaptomyza* suggests that concerted loss of genes with shared functions (i.e. olfaction) may play a central role in dietary niche shifts, but also reveals that herbivorous feeding may exert selective pressure on proteins with similar structures that may not be directly involved in the host plant interactions.

In a similar vein, the single OR that we identified under positive selection, and for which there exists functional data in *D. melanogaster* was *Or85f* (Table 1), which detects parasitoid wasp odors in *D. melanogaster* (Ebrahim et al. 2015). How these amino acid changes have affected the function of this receptor would be an interesting pursuit, given that parasitoid wasp pressure is exceptionally high in some populations of the herbivore *S. flava* (80% parasitism, Martin 2012), and the parasitoid community that attacks leaf-mining flies are distinct from those that attack free-living drosophilids. Furthermore, *S. flava* has also horizontally acquired the gene *cdtB* (Verster et al. 2019), which has been shown in other insect species to confer parasitoid resistance (Oliver et al. 2010), which further suggests the selective pressure from parasitoid attack.

Previous divergence-based selection analyses that included only *S. flava* and the 12 original *Drosophila* species with genome annotations found evidence of positive selection on *Or63a, Or67b* paralogs, *Or88a*, and *Or98a* (Goldman-Huertas et al. 2015; Masunaga et al. 2021), whereas here we only found evidence for positive selection on the basal herbivore branch for *Or85f* and *OrN2.3prime*. This highlights the advantage of including additional herbivorous species and non-herbivorous species that are more closely related (i.e. *S. pallida* at 23 million years diverged versus *D. grimshawi* at 33 million years (Matsunaga et al. 2021)). In particular, there is now strong evidence that the triplication and strong positive pressure on *Or67b* paralogs is specifically related to *S. flava*'s specialization on plants of the order Brassicales, as these ORs are specifically tuned to the volatile isothiocyanates (Matsunaga et al. 2021). Similarly, we find that previous selection analyses that focused on *S. flava* without the inclusion of close *Scaptomyza* species identified the rapid evolution and repeated duplication of *GstE5-8* (5 copies) as being a significant event in for the evolution of herbivory in this lineage, but our

analyses reveal that it is restricted to just the mustard specialists, *S. flava* and *S. montana*. This is consistent with the hypothesis that there are evolutionary changes associated with the transition to herbivory that are distinct from those evolutionary changes associated with specialization on particular host plant lineages and their particular chemical defenses.

Evolutionary patterns across gustatory receptors involved in detecting bitter compounds

Numerous studies have now shown that expansions and contractions of bitter gustatory receptors are typical in herbivorous species, as the diversity of these receptors mirrors the incredible diversity of toxic or noxious plant chemical compounds. It has been hypothesized that expansions of bitter GRs enable generalist species to detect a wide variety of plant-derived bitter compounds, while specialist herbivores (as well as specialist non-herbivores) tend to lose these bitter GRs, in line with ecological specialization and the neural limitation hypothesis, as their exposure to diverse plant compounds narrows. This pattern has been seen in butterflies, moths, aphids, flies, beetles, sawflies, mites and likely many other arthropod herbivore lineages (Smadja et al. 2009; Crava et al. 2020; Xu et al. 2016; Suzuki et al. 2018; Vertacnik et al. 2021). While it is impossible to know whether the *Scaptomyza* herbivorous ancestors were specialists or generalists, our results suggest that the ancestral herbivore lineage experienced a marked change in bitter detection: the majority of GR losses were involved bitter reception (paralogs of *Gr39aA*, *Gr59a, Gr59d*), the only changes in selective constraint on GRs involved positive selection on two bitter GRs (*Gr98a* and *Gr98bcd*), and there was additionally relaxed constraint on an ionotropic receptor also involved in bitter detection, *Ir56a*. These genetic changes likely reduced the ability of these flies to detect bitter compounds, which would otherwise limit their intake of their host plants. It should be noted that this reduction in bitter sensitivity has also been found in *D. suzukii* (Dweck et al. 2021), which has evolved herbivory (feeding on living, ripe fruit) on a similar timescale as herbivorous *Scaptomyza (Suvorov et al. 2022)*. However, although *D. suzukii* is herbivorous, it is actually a generalist or polyphagous on many plant families (Poyet et al. 2015), suggesting that the loss of bitter reception is a trait involved in the transition to herbivory, independent of specialization.

With the exception of *Gr39aA*, these candidate bitter GRs do not encode the “commonly expressed receptors” – a set of bitter GRs expressed in all bitter gustatory neurons (Dweck & Carlson 2020; Weiss et al. 2011). This suggests that, despite a diet based on bitter, toxic compounds, they are still able to detect bitter compounds generally, which has also been found in the specialist feeders on toxic and bitter hosts *D. sechellia* (noni fruit specialist), *D. erecta* (*Pandanus* fruit specialist), and *D. suzukii* (Dweck et al. 2021; Dweck & Carlson 2020). This ability is likely important for these flies that have evolved to detoxify host plant toxins but are not sophisticated enough to avert all damage. We have observed that in *S. flava*, feeding on plants bearing aliphatic and indolic glucosinolates causes them to develop more slowly (Gloss et al. 2019). Being able to differentiate varying levels of bitter compounds may enable them to differentiate toxins

levels between individual leaves or host plants, old from young, or healthy from damaged plants.

The losses of *Gr39aA, Gr59a*, and *Gr59d* are particularly interesting candidates to focus on for future study for several reasons, particularly in relation to the evolution of specialization. All three GRs have been implicated in dietary shifts in other species. *Gr59a* and *Gr59d* both underwent expansions in *D. suzukii* (Hickner et al. 2016), while both of these genes were lost in *D. sechellia* and *D. erecta* (McBride & Arguello 2007). As mentioned previously, *D. suzukii* has lost some bitter sensitivity, but this was achieved through transcriptional down-regulation of bitter GRs (Dweck et al. 2021). The pattern of expansion of *Gr59a* and *Gr59d* in a generalist (*D. suzukii*) and contraction in specialists (*D. sechellia*, and *D. erecta*) seems to strongly suggest involvement in mediating host breadth, rather than an evolution to herbivory per se. Furthermore, signals of selection have also been detected in *Gr59a* in *D. yakuba mayottensis,* which has convergently evolved to specialize on the noni fruit along with *D. sechellia* (Ferreira et al. 2020). Similarly, losses of *Gr39aA* have occurred across four additional independently evolved specialist lineages: *D. sechellia* and *D. erecta* (McBride & Arguello 2007), *D. sproati* (specialist on Araliaceae), leaf-mining *Lordiphosa*, and cactus specialists of the repleta group (Table S9). These results seem to highlight how changes to these bitter GRs is not necessarily a trait associated with the evolution of herbivory but is more strongly determined by specialization. What is also striking about these three GRs is that each of these genes exhibit numerous tandem duplications across the phylogeny of drosophilids. For instance, *S. pallida* and *S. hsui* bear nine and seven copies of *Gr59d*, respectively, while all herbivores have only three or four, and other *Drosophila* species have between one and seven. We also identified a remarkably high rate of turnovers for *Gr59d* and *Gr39aA* (Table S4). This seems to suggest that there may be transposable elements in the vicinity driving these duplications, which can then be selected upon during dietary shifts. *Gr39aA* is one of the few GRs that is alternatively spliced, is expressed in all bitter gustatory neurons, and is involved in the detection of many bitter compounds tested to date in *D. melanogaster* (Dweck & Carlson 2020). However, *D. melanogaster* only bears a single copy of the A exon, so whether these additional *Gr39aA* copies are expressed in species bearing multiple copies remains to be studied, but the difference in exon copies might suggest a dosage effect, where more copies may enable stronger bitter detection. Altogether our results on bitter GR evolution indicate some potential reduction in bitter detection in the ancestral herbivorous *Scaptomyza*. How exactly the combinatorial nature of GRs factors into these evolutionary changes will also be interesting to study.

Diverse functions of chemosensory genes associated with herbivory

In this study, we hypothesized that the majority of genetic changes associated with the evolution of herbivory would involve genes that interact with host plant compounds. It was thus surprising to find significant changes in genes involved in detecting various other compounds: carbon dioxide (positive selection on *Gr63a*), high salinity (positive selection on *Ppk19*), and water/low osmolarity (stronger purifying selection on *Ppk28*), and cool temperatures (relaxed purifying selection on *Ir21a* and *Ppk12*, strong purifying selection on *Ppk25*), and pheromones (loss of *Gr68a,* strong purifying selection on *Ppk25*). This

suggests that while the toxins in their diet play a significant role in driving the evolution of these gene families, there are also additional challenges presented by the leaf-mining lifestyle. The larvae of herbivorous species navigate through the mesophyll of the leaf – a fluid-filled cavity, where salt levels and temperature gradients may be more stable, and risk of carbon dioxide poisoning may be lower in leaves than in decaying organic material.

It has been suggested that the impacts of herbivory on mating behavior may be significant and varied: through the influences of host plant chemistry on insect pheromone production (Geiselhardt et al. 2012), on regulating pheromone release or initiation of reproductive behaviors (Raina et al. 1992), and/or by volatile organic compounds operating synergistically with insect pheromones to attract mates (Yang et al. 2004). Studies on the chemical ecology and reproductive system of herbivorous *Scaptomyza* are needed to identify any of these types of changes, but our results suggested some surprising implications. We found evidence of stronger negative selection on *Ppk25,* which detects the female pheromone 7,11-heptacosadiene, one of the two primary aphrodisiac pheromones that coat the *D. melanogaster* female cuticle (Ferveur et al. 1997). Meanwhile the Gr68a receptor*,* which is expressed in the male foretarsi of *D. melanogaster* and detects a male anti-aphrodisiac (Bray and Amrein, 2003; Shankar et al., 2015), was lost in herbivorous *Scaptomyza.* Together, these results could indicate a slight shift in mating strategy, where female pheromones are still required for successful courtship, but male-male interactions may be altered. Goldman-Huertas (2018) speculated that the loss of *Gr68a* may suggest that males employ physical, rather than chemical, protection of females from other mates, noting that *S. flava* males cling to the female abdomen to ward off other males. Investigations into the mating systems of herbivorous *Scaptomyza* are rife with opportunity, particularly given the pronounced shift from feeding and congregating on ephemeral decaying plants to more persistent living plants, which may drive different patterns of intraspecific communication (Goldman-Huertas 2018). Furthermore, herbivorous *Scaptomyza* may offer genetic insight into the effects of host plant use on generating prezygotic barriers to gene flow. For instance, cuticular hydrocarbon composition of the mustard leaf beetle *Phaedon cochleariae* (Chrysomelidae), which also feeds on various plants of the Brassicaceae family, is influenced by host plant species, such that males prefer females that have fed on the same host plant species (Geiselhardt et al. 2012). Similarly, cuticular hydrocarbon profiles and mating success in *D. mojavensis* have also been found to be influenced by the species of cactus they have been reared on (Etges et al. 2009). Collection data on herbivorous *Scaptomyza* in the wild suggests that while several species occur in sympatry, they also vary in degrees of specialization and host plant ranges – from monophagy in *S. nigrita* (note: polyphagy is seen in *S. nr. nigrita*) to requiring specific host plant compounds for feeding in *S. montana* (indole glucosinolates (Gloss et al. 2017) to the expansion onto non-Brassicaceae families in populations of *S. flava* outside of North America (Table S8).

Duplication and positive selection on a detoxification gene.

Numerous studies have indicated that gene duplications, followed by functional divergence can spur biological novelties – new traits or adaptations to new niches

(reviewed in (Carscadden et al.) 2022). Examples abound across the diversity of life and across complex traits, from trichromatic vision in old world monkeys to snake venom phospholipase genes (Dulai et al. 1999; Lynch 2007). Contrary to our initial hypothesis that we might find numerous instances of gene duplications across chemosensory ande detoxification gene families, we instead found many detected within individual herbivore species or among just the mustard specialists, but only two genes that duplicated along the branch at the base of the three herbivores. This suggests that gene duplications may spur further specialization but may not play a prominent role in initial transitions to herbivory. We cannot speculate on the role of the *Ir67a* duplication because the function is unknown in other species. However, the role of the *GstS1* duplication may still have been a critical step in the early transitioning herbivore, considering the strong positive selection we detected in the *GstS1b* paralog (Table 3), which shares 73-93% sequence similarity with *GstS1a* (in *S. flava* the 5' end is truncated by 126 bp). This GST is involved in detoxifying oxidation products, lipid peroxidation products, and organometallic compounds (Saisawang et al. 2012; Whitworth et al. 2005; Singh et al. 2001; Vorojeikina et al. 2017). Both *GstS1* copies are expressed in *S. flava* larvae, but there is not strong up-regulation of either copy in response to isothiocyanates, the defense compound in mustard plants, reinforcing that this duplication is not specific to mustard feeding. In *D. melanogaster*, *GstS1* has high expression in flight muscle and in the adult central nervous system, with previous studies suggesting both structural and detoxifying functions (Clayton et al, 1998; Whitworth et al. 2005). Functional genetic testing in herbivorous drosophilids should reveal the fate and possible functional divergence of these two paralogs: whether expression patterns differ across tissues, whether the loss of either or both paralog affects flight muscle structure, and the specificity of each paralog towards diet-derived xenobiotics.

In Chapter 3, I investigate changes in bitter taste perception in the herbivorous drosophilid *Scaptomyza flava*. The results of Chapter 2 highlighted significant gene losses of and relaxed purifying selection on gustatory receptor (GR) genes whose orthologs in *Drosophila melanogaster* have been reported to be involved in detecting bitter compounds that elicit aversive reactions. These results suggested that herbivorous *Scaptomyza* may have overcome their ancestral disgust of plant secondary metabolites by losing these GRs or accumulating mutations to reduce their functionality. In this chapter, I tested this hypothesis first by evaluating the bitter feeding preferences of *S. flava*, relative to non-herbivorous species. In particular, I focused on their evolved responses towards compounds that are specific to the plant family they specialize on. I coupled these data with recordings from labellar taste sensilla to test whether changes in bitter preferences could be directly correlated with changes in bitter sensitivity within particular sensilla types. Considering the GR candidates identified from Chapter 2 in combination with this behavioral and electrophysiological data, these results suggest a mechanism through which specific bitter GR losses could have caused reduced bitter sensitivity in one class of sensilla, and ultimately reduced behavioral aversion against their host's chemical defenses.

Chapter 3. Taste evolution in an herbivorous drosophilid

Abstract

Plant secondary metabolites pose a challenge for generalist herbivorous insects because they are not only potentially toxic, they also may trigger aversion or repellence. On the contrary, some highly specialized herbivorous insects evolved to use these same compounds as 'token stimuli' for unambiguous determination of their host plants. Two open questions that emerge from these observations is how herbivores overcome this aversion to host plant toxins and whether evolutionarily young herbivore lineages evolve increased attraction to these same compounds. In this study, we investigated these questions by focusing on the evolution of bitter taste preferences in the herbivorous drosophilid *Scaptomyza flava*. We examined the behavioral and neural responses of *S. flava* towards plant bitter compounds in a comparative approach across a phylogenetic gradient of non-herbivorous close relatives (*S. pallida*, *S. hsui*, and *D. melanogaster*). We observed that female *S. flava* experienced a broad change in bitter detection, rather than changes toward the mustard-specific defense compounds, glucosinolates. In a dye-based consumption assay, *S. flava* exhibited shifts in aversion toward the non-mustard bitter compounds caffeine and lobeline, and reduced aversion towards glucosinolates, whereas non-herbivorous *Scaptomyza* showed strong aversion to all tested bitter compounds. We then examined whether these changes in bitter preferences could be explained by changes in sensitivity in the peripheral nervous system and compared electrophysiological responses from the labellar sensilla of *S. flava*, *S. pallida*, and *D. melanogaster.* Using scanning electron microscopy, we created a map of labellar sensilla in *S. pallida* and *S. flava* and assigned each sensilla to a functional sensilla class based on their morphology and initial response profiles to bitter and sweet compounds. Despite a high degree of conservation in the morphology and spatial placement of sensilla between *S. pallida* and *S. flava*, electrophysiology revealed that *S. flava* had reduced sensitivity to glucosinolates, although this varied across glucosinolates. This reduction was specifically found in only I type sensilla. Finally, we speculate on the potential role that genetic changes in gustatory receptors between *S. pallida* and *S. flava* we previously discovered may play in driving these patterns. Specifically, we hypothesize that the evolution of bitter receptors expressed in I type sensilla may have driven the reduced sensitivity observed in *S. flava*, and ultimately, its reduced bitter aversion. While it still remains unclear the extent to which young herbivores evolve gustatory attraction towards plant defense compounds, the *S. flava* system showcases the importance of reduced aversion to bitter defense compounds, and how this may be achieved.

Introduction

The evolution of new dietary shifts critically relies on an animal's ability to sense particular chemicals and textures in their food and decide whether to consume it. This is true from octopi that search the seafloor, tasting with their tentacles (Wells, 1963) to newborn human and non-human primates, that wince at the taste of bitter quinine (Steiner et al., 2001). In terrestrial animals, these sensations often involve not only long-range

attraction mediated by vision and olfaction, but also short-range tactile gustation. Taste, which has been less well-studied in terms of niche shifts, operates as a final checkpoint for the acceptance of the substrate as an acceptable food source for themselves and/or their offspring (Scott, 2018). In recent decades, there has been a wealth of literature suggesting that the evolution of attraction to novel food sources can be attributed in large part to the dynamic evolution of the large families of chemosensory proteins expressed in chemosensory organs (McBride, 2007; Edger et al., 2015). However, what remains unknown is the progression of steps through which feeding preferences evolve, the mechanisms and sensory cues involved, and how genetic changes translate to real differences in feeding choices across species with different diets.

In no other group has chemosensory evolution been studied more than among herbivorous insects, which are among the most diverse organisms on Earth, accounting for a quarter of all known eukaryotic species (Bernays, 1998; Wiens et al., 2015). At the interface of herbivorous insects and their host plants are plant chemical defenses, which have evolved and diversified in a chemical arms race to repel and poison potential herbivores (Ehrlich and Raven, 1964). Plant secondary metabolites pose a challenge for insects to evolve herbivory because they are not only potentially toxic to non-specialist insects, they also may trigger an innate aversion or repellence response. Insects possess gustatory bristles or sensilla across their bodies, on the proboscis, legs, and wings. These sensilla house gustatory neurons whose cell membranes are embedded with gustatory receptors (GRs) that detect chemical cues from their food and environment (Nayak and Singh, 1985; Clyne et al., 2000). Studies on gustatory evolution in generalist herbivores – insects feeding on a broad range of plant families – have illuminated a strong correlation between host breadth and GR diversity: specifically, dramatic expansions of GR repertoires have been implicated in driving the detection of a large diversity of host plant species, given the differential ligand specificity of different GRs (Wanner and Robertson, 2008; Smadja et al., 2009; Briscoe et al., 2013; Xu et al., 2016; Suzuki et al., 2018). However, the majority of herbivores, over 80%, are oligophagous and feed on one or a restricted set of plant families (Schoonhoven and Jermy, 1998).

In contrast to generalist herbivores, much of the literature on taste preferences of specialists has focused on their attraction to compounds derived from their host plant as stimulants or host identifiers. Because of their specialization on one or a few plant families, these ‘token stimuli’ provide unambiguous identification of their hosts (Verschaffelt, 1910; Lipke and Fraenkel, 1956). For instance, the arctiid moth *Utetheisa ornatrix* is not only attracted to but sequesters high concentrations of pyrrolizidine alkaloids from host plants (Hoina et al., 2013). Similarly, Brassicaceae-specific glucosinolates strongly stimulate feeding and oviposition in *Pieris* butterfly species (Chew and Renwick, 1995). In several specialist herbivores, individual gustatory receptors that detect these plant compounds have even been identified: Chinese citrus fly *Bactrocera minax* uses *BminGr59f* to detect compounds from unripe citrus fruits (Zhang et al., 2022); *Pieris rapae* uses *PrapGr28* to detect the glucosinolate sinigrin from their host plants (Yang et al., 2021); and *Papilio xuthus*, specialist on Rutaceae, uses *PxutGr1* to detect synephrine, an oviposition stimulant (Ozaki et al., 2011). However, most of the herbivores that have been studied evolved herbivory over 100 million years ago (Gloss et al., 2019),

and do not provide any insight into gustatory evolution during the early stages of herbivory and host plant specialization. Examples from *Drosophila* include more recently evolved specialists, like *D. sechellia* (monophagous noni specialist) and *D. suzukii* (polyphagous ripe fruit specialist), which show that in response to feeding on bitter hosts, they have experienced an overall evolutionary loss of aversion to bitter compounds, likely due to dramatic gene losses and/or down-regulation of GRs tuned to bitter compounds (McBride et al., 2007; Dweck and Carlson, 2020; Dweck et al., 2021). In addition, at least in *D. sechellia,* it has evolved increased sensitivity and attraction to specific host plant compounds (Dekker et al., 2006). With few examples of recently evolved herbivore lineages, it remains controversial whether host plant specific compounds evolve as feeding stimulants early on in the transition to herbivory, and whether specialists evolve greater or lower sensitivity to bitter compounds (Bernays and Chapman, 1994; Bernays et al., 2000; Schoonhoven and Van Loon, 2002).

To address this gap in our knowledge, we used the model herbivore *Scaptomyza flava*, a member of an herbivorous clade within the genus *Scaptomyza*. *Scaptomyza* is sister to the Hawaiian *Drosophila* clade and phylogenetically nested within the paraphyletic *Drosophila* genus (Lapoint et al., 2013). Most herbivores from this clade are specialists on mustard plants in the Brassicaceae family and relatives in the Brassicales. *S. flava* offers an ideal system to investigate bitter taste evolution because this species can be readily caught from the wild and reared in the lab on the model plant *Arabidopsis thaliana (Whiteman et al., 2011)*. Close non-herbivore relatives, like *S. pallida* and *S. hsui*, each from distinct but closely related subgenera *Parascaptomyza* and *Hemiscaptomyza*, can also be easily reared in the lab. With genome assemblies now available for these *Scaptomyza* species and many other drosophilid species (Chapter 2; (Kim et al., 2021)), alongside decades of accumulated knowledge on the molecular and neural basis of *D. melanogaster*'s gustatory system, *S. flava* is a prime system to dissect bitter taste evolution.

The primary chemical defenses in mustard plants are derived from glucosinolates, sulfur-rich anionic thioglucosidases that are hydrolyzed by the plant enzyme myrosinase (a beta glucosidase) upon tissue damage (Halkier and Gershenzon, 2006). The hydrolysis of glucosinolates by myrosinases – the so-called "mustard oil bomb" – generates several different breakdown products (e.g., isothiocyanates, thiocyanates, and nitriles). The most well-known are isothiocyanates, which are toxic to feeding insects because of their strong electrophilic reactivity, binding to DNA, RNA, and proteins and cleaving disulfide bonds (Kawakishi et al., 1983). Despite the toxicity of their breakdown products, glucosinolates are used by more than 25 herbivorous mustard specialists (e.g., the cabbage white *Pieris rapae* and cabbage root fly *Delia radicum*) as token stimuli (Fahey et al., 2001; Hopkins et al., 2008). These insects use glucosinolates, which are largely non-toxic, as cues to stimulate feeding and egg-laying. However, responses to individual glucosinolates vary depending on the glucosinolate, glucosinolate class, and insect species (Bidart-Bouzat and Kliebenstein, 2008).

Glucosinolates are categorized into one of three major structural groups based on the amino acids from which their side chain is derived: aliphatic glucosinolates mainly

from methionine, indole glucosinolates from tryptophan, and aromatic glucosinolates from phenylalanine or tyrosine. Not all mustard specialists exhibit the same pattern of responses to these different classes, but there are some general trends. Specialists tend to be indifferent towards indolic glucosinolates because the mustard oils they form are unstable, breaking down quickly other derivative, less toxic products like indole-3 carbinol (Wittstock et al. 2003). In contrast, many specialists have retained sensitivity towards aliphatics because they produce toxic, stable mustard oils, and have been shown to reduce herbivore performance (Kliebenstein et al. 2004). Because *S. flava* has evolved a gained attraction towards the breakdown products of glucosinolates (Matsunaga et al., 2022), we hypothesized that in addition to lost aversion to glucosinolates, they might have also gained attraction towards these toxin precursors for increased efficiency in finding appropriate hosts.

To gain insight into changes in the gustatory system of relatively recently evolved specialist herbivores, we examined the behavioral and neural responses of *S. flava* towards plant bitter compounds, contrasting its responses against those of non-herbivorous close relatives (*S. pallida*, *S. hsui*, and *D. melanogaster*). We utilized a consumption-based feeding assay to identify changing patterns of general bitter detection and changes towards mustard-specific compounds. We then used scanning electron microscopy and single sensillum electrophysiology to investigate whether there were alterations at the level of labellar sensilla sensitivity that could explain the behavioral differences. Altogether, we identified a mechanism by which glucosinolate sensitivity was reduced in *S. flava*, while retaining some general bitter sensitivity, and then speculate on how the genomic changes among candidate gustatory receptors might have driven these changes.

Materials and Methods

Fly stocks and maintenance.

Species with different feeding ecologies required different rearing media. *D. melanogaster* (wild-type Canton-S) were maintained on standard Bloomington fly media, provided by the UC Berkeley fly food facility. A colony of *S. flava* (originally collected near Dover, NH, USA) was maintained on *Arabidopsis thaliana* (Col-0 accession) and 10% honey-water solution. Colonies of *S. pallida* and *S. hsui* (bred from isofemale lines, collected from Berkeley, CA, USA) were reared on Bloomington media topped with organic chopped spinach. All fly species were maintained at 22ºC ± 2ºC, 12L/12D photoperiod, and 70% relative humidity. Only three species – *S. flava*, *S. pallida*, and *D. melanogaster* – were used for electrophysiological recordings for practical reasons. For these experiments, *S. pallida* and *D. melanogaster* were reared on Nutri-Fly German food formulation (Genessee, Cat. No. 66-115), while *S. flava* was reared on *A. thaliana* (Col-0).

Feeding assays with individual bitter tastants.

To test the feeding preferences of each species, we used a two-choice feeding assay. Newly emerged adult flies were provided honey water *ad libidum*. Roughly twenty

mated female flies (5-8 days old) were transferred to a vial with moistened filter paper and starved at room temperature for 24 hours. Flies were briefly knocked out with CO_2 and transferred to the experimental vials. These vials were capped with 3D-printed lids through which four capillaries were inserted (Fig. 1b). The capillaries had alternating solutions of 5mM sucrose alone and 5mM sucrose with a bitter tastant (0.1, 1, or 10mM) (Fig. 1b). Red (0.2 mg/ml sulforhodamine B) and blue (0.08 mg/ml erioglaucine) dyes were added to the two solutions, and were randomly assigned to either solution across different experiments to minimize the effect of dye preferences (Mack and Zhang, 2021). To characterize how herbivorous species may differ from their non-herbivorous relatives in their breadth and sensitivity towards bitter tastants, the test compounds included both those that have been found to elicit aversive responses in *D. melanogaster*, as well as compounds found in mustard plants, specifically glucosinolates, the precursor molecules to the toxic isothiocyanates (listed in Table 1). After feeding in the dark for nine hours, flies were anesthetized with CO_2 and abdomens were scored as red, blue, purple, or no color. In pilot feeding assays, nine hours was determined as the optimal run time to allow for a majority of *S. flava* and *S. hsui* to feed, without significant acclimation towards the bitter compounds. Five replicates (vials) were performed for each species per tastant at a given concentration, unless specified otherwise. Replicates were included if there were ≥15 flies in the vial at the conclusion of the experiment and ≥30% flies fed.

For each tastant, experiments were performed across at least three separate days. A preference index (PI) was calculated for each vial of flies with PI = (number of flies with red abdomens + ½ the number of flies with purple abdomens)/(total number of flies with any color in their abdomens), if the red dye was added to the bitter solution (if blue, the numerator included the number of flies with blue abdomens instead of red). Preference data were analyzed for differences between species using the Kruskal-Wallis test in R v4.2.1. If significant differences were found ($P < 0.05$), post-hoc Dunn's tests were performed to determine which species had significantly different preferences.

Table 1. Chemical compounds used in feeding experiments and electrophysiology

Class	Chemical	Source	Catalog Number
Dye	Brilliant blue FCF	Sigma-Aldrich	80717
Dye	Sulforhodamine B	Sigma-Aldrich	230162
Glucosinolate	(-)-Sinigrin hydrate	Sigma-Aldrich	85440
Glucosinolate	Glucobrassicin	Sigma-Aldrich	PHL80593
Glucosinolate	Gluconasturtiin	Sigma-Aldrich	PHL89689
Glucosinolate	Glucoraphanin	Extrasynthese	2509 S
Glucosinolate	Glucotropaeolin	Extrasynthese	2510 S
Glucosinolate	Neoglucobrassicin	Extrasynthese	2519 S
Non-mustard bitter	Caffeine	Sigma-Aldrich	C0750
Non-mustard bitter	(−)-Lobeline hydrochloride	Sigma-Aldrich	141879
Solvent	Tricholine citrate	Sigma-Aldrich	T0252

Scanning electron microscopy (SEM).

To identify sensillum types based on morphology among the labellar sensilla in *S. pallida* and *S. flava*, SEM was performed on adult female flies of each species. Fly heads were separated from their bodies to improve the penetration of fixatives into the heads.

Heads were fixed in 2% glutaraldehyde in 0.1M sodium cacodylate, stained with osmium tetroxide, and dried through an ethanol series (35-100%). The ethanol was evaporated with a critical point dryer (Tousimis AutoSamdri 815). Specimens were mounted onto stubs, coated with gold-palladium using a Tousimis sputter coater, and imaged with a Hitachi SM-5000 scanning electron microscope (200 – 1500x). All SEM work was carried out at the UC Berkeley Electron Microscope Laboratory.

Electrophysiology.

We performed electrophysiological recordings on labellar sensilla of *S. pallida, S. flava,* and *D. melanogaster* to determine the extent to which: (1) bitter feeding preferences involve the gustatory system and (2) differences in bitter preferences across species are explained by differences in sensitivity of neuronal responses. Newly eclosed flies were transferred to German food formulation for several days, allowed to mate, and then starved on water for 24 hours prior to recordings. Females (5-10 days old) were immobilized and prepared as previously described (French et al., 2015). Using the tip recording method (Hodgson and Roeder, 1956), the recording electrode simultaneously delivers a tastant and transmits the electrical pulse to a TasteProbe amplifier (Marion-Poll and van der Pers, 1996), and analyzed in the spike sorting software dbWave. The recording began immediately on contact with the sensilla for a duration of two seconds. Tricholine citrate (TCC, 30mM) was added to all solutions to silence the water cell (Wieczorek and Wolff, 1989). For each tastant, 5-15 recordings were made per sensilla. Freshly made solutions were used for only one week. Spikes were manually sorted and based on amplitude and shape of spikes. Stimulations to the same sensilla were separated at least by ten minutes.

We first performed recordings from all labellar sensilla using 10 mM sucrose, 1 mM lobeline, and 10 mM caffeine to determine whether *S. flava* and *S. pallida* had similar sensillar classes as *D. melanogaster*. In *D. melanogaster*, these three compounds can be used to differentiate L, I-a, I-b, S-a, S-b, and S-c classes of labellar sensilla (Weiss et al., 2011). Because behavioral responses to these compounds were similar between *Scaptomyza* species and *D. melanogaster*, this suggested that sensillar sensitivity to these compounds might be conserved. A hierarchical cluster analysis using Ward's classification method in the PAST program was then used to identify distinct sensillar classes based on their responses to these three compounds (Hammer et al., 2001). Based on clustering, we chose representative sensilla from each class. To further test the differences in sensitivity of these sensilla, we generated a dose response curve for lobeline (0.1, 1, 10 mM) and sucrose and caffeine (1, 10, 100 mM).

Recordings were then performed using glucosinolates, the mustard-specific toxin precursor molecules (listed in Table 1). Gluconasturtiin, neoglucobrassicin, and sinigrin were tested at 0.1, 1, 10, and 30mM concentrations, and glucoraphanin, glucobrassicin, and glucoraphanin were tested at 1 and 10mM concentrations.

Results

Mustard specialist shows less aversion to some bitter compounds compared to microbe-feeders

To examine how bitter taste perception may have evolved in the herbivorous *S. flava*, we compared its feeding preferences on bitter compounds against those of microbe-feeding drosophilids across a phylogenetic gradient (*D. melanogaster*, *S. pallida*, and *S. shui*). We tested both glucosinolates – the mustard-derived precursors to the toxic breakdown products – and non-mustard bitter compounds to evaluate whether changes in bitter perception were mustard-specific. In a dye-based consumption assay, flies were presented with two sucrose solutions, dyed red and blue, one of which was randomly chosen to also have the bitter tastant added. Appetitive, neutral, or aversive responses were indicated by abdomen color after nine hours of feeding, with a preference index (PI) of zero indicating absolute aversion/repulsion, 0.5 no preference, and one indicating attraction. All species, including the herbivore *S. flava*, were averse to the non-mustard specific compounds caffeine and lobeline (Fig. 1c). However, at a lower concentration of 1mM caffeine, *S. flava* experienced reduced aversion compared to the non-herbivorous *Scaptomyza* (*S. flava*, PI=0.4; *S. pallida,* PI=0.14; *S. hsui*, PI=0.19). In contrast, while most other microbe-feeders showed less aversion to lobeline at 1mM compared to at 10 mM, *S. flava* showed almost complete aversion even at 1mM (PI=0.01), showing stronger aversion than *S. pallida* (PI=0.22), although the same level as *S. hsui* (PI=0.05). These results indicate that while *S. flava* has evolved reduced aversion to some bitter compounds (caffeine), they may have evolved stronger aversion towards others (lobeline).

The responses towards glucosinolates varied depending on the species and glucosinolate class (Fig. 1d). Most glucosinolates elicited aversive responses in all species. The strongest aversive responses were from aromatic and indolic glucosinolates, with *S. pallida* showing almost complete aversion towards some compounds (e.g., 10 mM glucotropaeolin, PI=0.03; 10 mM neoglucobrassicin, PI=0.02). However, *S. flava* found aromatic glucosinolates (glucotropaeolin and gluconasturtiin) less aversive as compared to non-herbivorous *Scaptomyza*, showing no preference at a lower concentration of 1 mM (glucotropaeolin, PI=0.43; gluconasturtiin, PI=0.66). In response to indolic glucosinolates (neoglucobrassicin and glucobrassicin), *S. flava* showed almost complete indifference to these indole glucosinolates even at a higher concentration of 10 mM (neoglucobrassicin, PI=0.4; glucobrassicin, PI=0.48), relative to non-herbivorous *Scaptomyza* (i.e., *S. pallida*: neoglucobrassicin, PI=0.02; glucobrassicin, PI=0.08).

All species were generally more indifferent towards aliphatic glucosinolates (sinigrin and glucoraphanin) than those towards aromatic and indolic glucosinolates. Furthermore, we did not find statistically significant reductions in aversion towards glucoraphanin in *S. flava* at either of the concentrations tested (1 and 10mM), and only reduced aversion towards sinigrin at 10mM. Interestingly, *D. melanogaster* did not show any aversion or preference towards the aliphatic glucosinolates at all (PI ≈ 0.5 across all tested concentrations of both aliphatic compounds), but in general, *D. melanogaster* also showed lower aversion towards all the other glucosinolates, relative to the non-herbivorous *Scaptomyza.* Overall, these behavioral experiments involving glucosinolates

suggest that *S. flava* evolved reduced aversion to glucosinolates, but preferences varied across different glucosinolates in herbivores and non-herbivores alike.

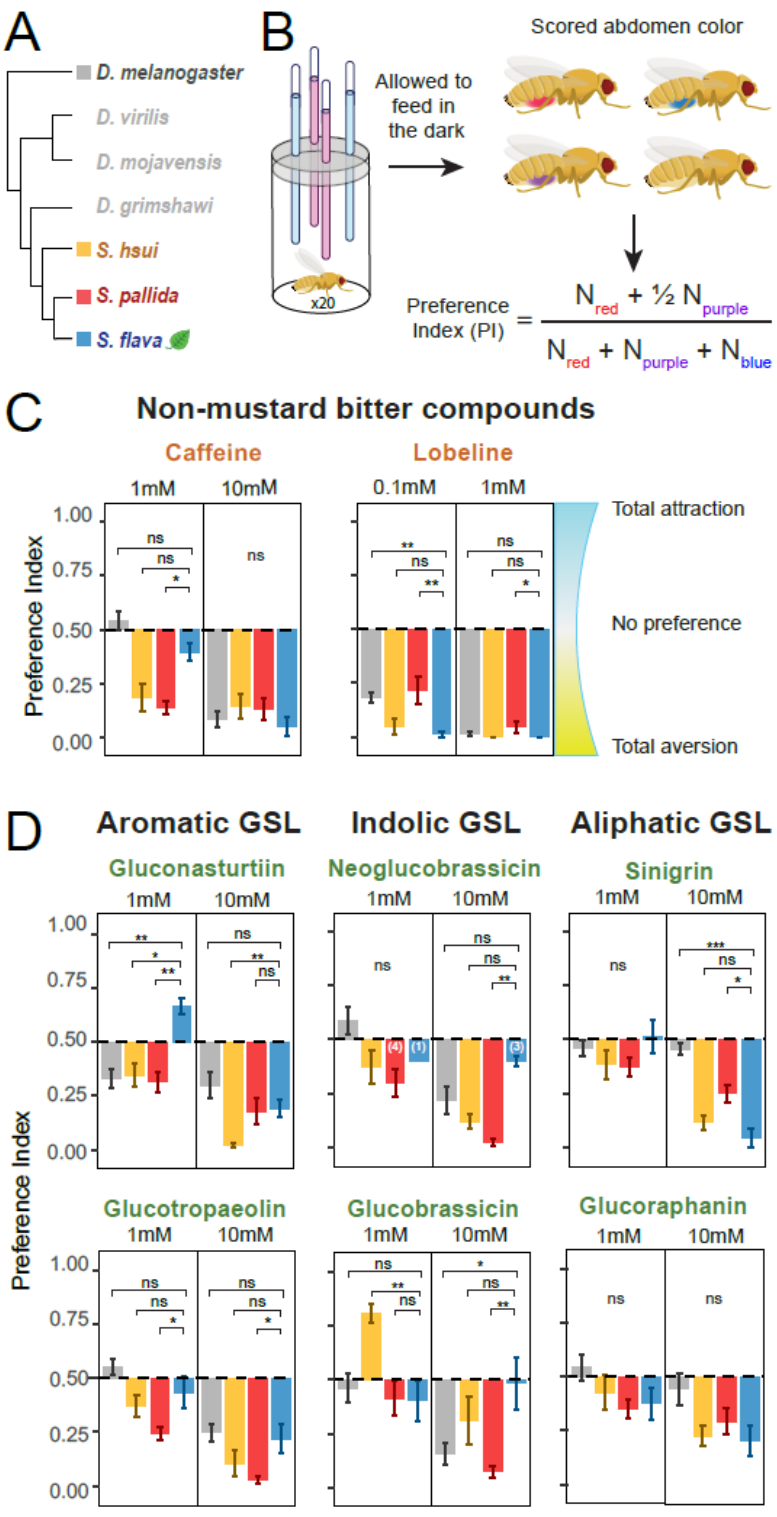


Figure 1. *Scaptomyza flava* shows reduced bitter aversion to some plant-derived compounds. (a) Phylogenetic placement of the herbivore *S. flava* among microbe-feeding drosophilids. (b) Schematic of the two-choice feeding assay. Feeding preferences towards (c) non-mustard bitter compounds (caffeine and lobeline) and (d) mustard-specific glucosinolates, the precursor molecules to toxic isothiocyanates. Glucosinolates are grouped by class based on the amino acids they are derived from: aromatic (phenylalanine, tyrosine), indolic (tryptophan), and aliphatic (methionine). For panels b and c, bar graphs represent the average, and error bars represent the standard error of the mean. PI = 0 indicates no preference towards the bitter tastant, PI < 0 indicates aversion, and PI > 0 indicates attraction. Preference differences among species were tested with m Kruskal-Wallis test, followed by Post-hoc Dunn's test (*p<0.05, **p<0.01, ***p<0.001, ns = not significant). For panels c and d, n > 5 vials of 15-20 flies, unless specified in parentheses.

Conservation of sensilla number and location within *Scaptomyza*

Using scanning electron microscopy, we found that the number and location of sensilla were largely conserved between *D. melanogaster* and *Scaptomyza* species, and even more similar between *S. flava* and *S. pallida* (Fig. 2). All three species shared the

same number and location of long L type sensilla. There were, however, typically two fewer intermediate I type sensilla in both *S. flava* and *S. pallida*, compared to *D. melanogaster*, specifically along the dorsal region of the labellum. Similarly, there was a slight difference in S type sensilla between the two genera, with 12 S type sensilla found in the two *Scaptomyza* species, and typically 10 in *D. melanogaster*. The number of I and S type sensilla varies across individuals of *D. melanogaster* as well (Hiroi et al., 2002), and we observed during electrophysiological recordings that both I and S type sensilla numbers were variable across all 3 species, with either one more or fewer present across some individuals. It is unclear whether this smaller number of I type sensilla and higher number of S type in *Scaptomyza* impacts their taste preferences. However, these comparisons clearly show that between *Scaptomyza* species, the number and location of sensilla types are conserved, which suggests individual sensilla between *S. flava* and *S. pallida* are homologous, and their electrophysiology sensitivities can be compared.

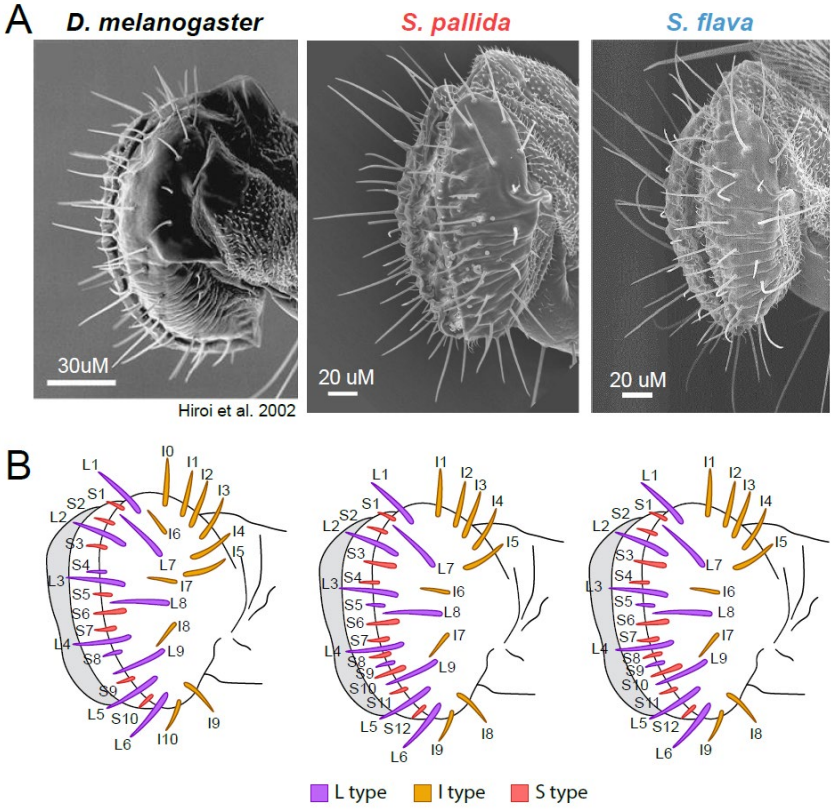


Figure 2. Labellar sensilla types are conserved across *D. melanogaster, S. flava,* and *S. pallida*, with a few changes consistent in both non-herbivorous and herbivorous *Scaptomyza*. (a) Representative images of scanning electron microscopy (SEM), showing lateral view of the labellum. Sensilla of *Scaptomyza* species tend to be longer among all sensilla classes than those of *D. melanogaster*. *D. melanogaster* SEM is reprinted from Hiro et al. 2002. (b) Illustration of sensilla types and their location based on sensilla lengths from SEM.

Identification of sensilla types based on electrophysiology response profiles

To identify herbivore-specific changes in bitter sensitivity in the peripheral nervous system, we first sought to identify whether *Scaptomyza* species have different functional sensilla classes than *D. melanogaster* (Weiss et al. 2011). In *D. melanogaster* and other *Drosophila* species, sensilla classes can be differentiated based on their electrophysiological responses to caffeine, lobeline, and sucrose (Weiss et al. 2011; Dweck et al. 2021; Dweck et al. 2020). Responses to these compounds may have evolved differently in *Scaptomyza* species, but because *Scaptomyza* also show aversion to

caffeine and lobeline and attraction to sucrose, similar to *D. melanogaster* (Figure 1c), we reasoned that there may be conservation of functional sensilla classes. Following electrophysiological recordings, we used a hierarchical clustering analysis to assign each of the ~30 sensilla of each species to a sensilla class.

L type sensilla across all species responded in a similar manner as those in *D. melanogaster*: unresponsive to bitter compounds (caffeine and lobeline) but strongly stimulated by sugars (sucrose) (Fig. 3a). However, the spike responses of L type sensilla were slightly lower in *S. flava* and *S. pallida* compared to *D. melanogaster.* This could be attributed to the fact that *D. melanogaster* feeds mainly on rotting fruits, which are higher in sugars than decaying or living vegetative tissues, which *S. pallida* and *S. flava* feed on, respectively. The I and S type sensilla of *S. flava* and *S. pallida* both responded to bitter compounds and sugars, as they do in *D. melanogaster*. I type sensilla in *D. melanogaster* can be separated into two classes: I-a type, which can be stimulated by lobeline, but not caffeine, and I-b type, which can be stimulated by caffeine but not lobeline (Weiss et al., 2011). In *Scaptomyza*, we see this distinction between the two types disappear, although there is some intersensilla variation, particularly in *S. pallida* (Fig. 3a). The lack of two I types is supported by a hierarchical cluster analysis, which shows almost all I types clustering together in both *S. flava* and *S. pallida* (Fig. S1). It is possible that more bitter compounds could help differentiate I subtypes, but two I subtypes have only been found in *D. melanogaster* and *D. simulans* (Dweck et al. 2020) and no other *Drosophila* (Dweck et al. 2020, Dweck et al. 2021). This pattern suggests that two I types may be derived in the lineage leading to *D. melanogaster* and *D. simulans*. Most notably, while responses to sucrose and lobeline are similar between *S. flava* and *S. pallida* among I types, there is a reduced spike response to caffeine in *S. flava* compared to *S. pallida* (Fig. 3a and 3c, GLM, $t = -2.88$, $df = 47$, $P = 0.006$), which mirrors their feeding preferences (Fig. 1c).

Among S type sensilla, 3 subtypes were present in *S. flava*, S-a, S-b, and S-c (Fig. 3a), which are also present in *D. melanogaster*. S-c sensilla have a similar profile as L sensilla as they also do not have a bitter gustatory neuron, and thus only respond to sugars. In *S. flava*, the S-c sensilla corresponded to sensilla S5 and S9 (S4 and S8 in *D. melanogaster*). The same sensilla (S5 and S9) appear to be S-c sensilla in *S. pallida*, and other sensilla may functionally be S-c sensilla as well (S4, S7, S8, S10, S11, S12). In *D. melanogaster*, S-a and S-b sensilla can be distinguished by their sensitivity towards caffeine, where the three S-b sensilla are more sensitive to caffeine than the five S-a (Weiss et al. 2011). In *S. flava*, the three S-b sensilla correspond to S4, S7, and S11, that are differentiated from the seven remaining S-a sensilla, not by their sensitivity to caffeine, but by their sensitivity to lobeline, as S-b sensilla exhibited higher spike rates after stimulation with lobeline than S-a (Fig. 3a). Interestingly, the sensitivities of both S-a and S-b sensilla towards lobeline in *S. flava* are higher than any sensilla in any of the species we surveyed. This is in congruence with the increased behavioral aversion we found in *S. flava* compared to most of the non-herbivores (Fig. 1c). In *S. pallida*, based on responses from only caffeine and lobeline, we are not able to differentiate S-a and S-b subtypes: all short sensilla that respond to bitter compounds (S1, S2, S3, S6) respond with similar intensity. The delineation of short sensilla types in *S. pallida* is still tentative, and more compounds are needed to differentiate if any S-b type sensilla exist. Given the

conservation of sensilla morphology and spatial locations between *S. flava* and *S. pallida* (Fig. 2), it is possible that the S-b sensilla in *S. pallida* could be the same as those identified in *S. flava,* in which case *S. pallida* S-b sensilla have lost their sensitivity towards bitter compounds.

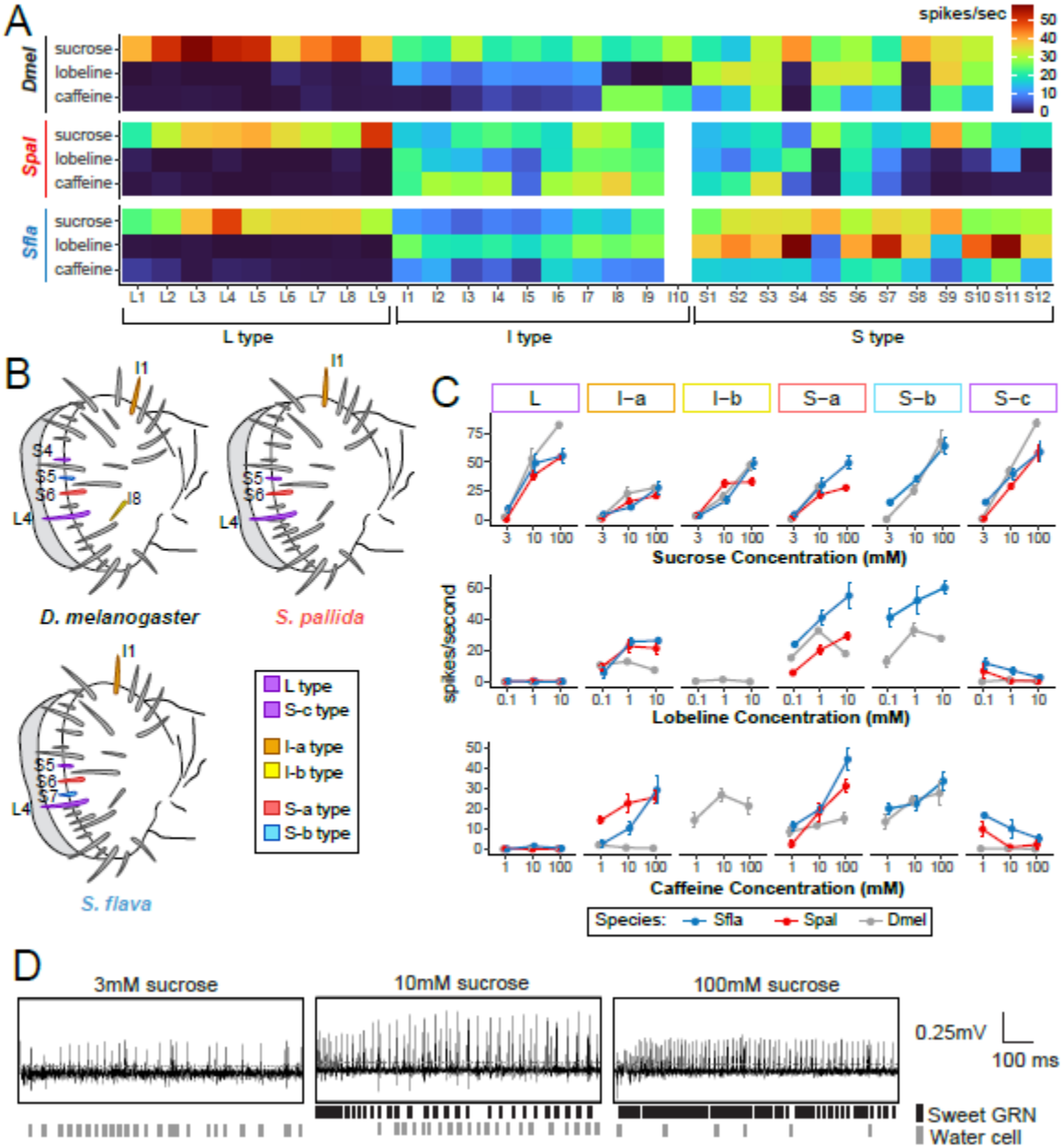


Figure 3. Identification of labellar sensilla classes based on response profiles (a) Heatmap of spike responses to caffeine (10mM), lobeline (1mM), sucrose (10mM), illustrating sensilla types by their different responses to these compounds, across all ~30 sensilla in 3 species. (b) Representatives of each sensilla class chosen based on hierarchical clustering (Figure S1), which were recorded from for (c) dose response curves for caffeine, lobeline, and sucrose (and glucosinolates, Figure 4). Error bars represent the standard error of the mean. (d) Representative spike recordings from *S. flava* S5 sensilla (S-c type) at 3 concentrations of sucrose show an inverse relationship between the rate of sweet GRN firing (black bars) and water cell firing (gray bars).

We chose representative sensilla from each class within each species (Fig. 3b) for the remaining recordings. To confirm the sensitivity of each sensilla type towards caffeine, lobeline and sucrose, we generated dose response curves for each representative sensillum. It was not possible to reliably record from *D. melanogaster* S sensilla with 100mM because at high concentrations the spike trains became too noisy for spike identification (Hiroi et al., 2002), so five replicates for these sensilla could not always be achieved. In most cases, spike rates increased with increasing concentrations of the tastant, the only exception being S-c sensilla in *S. flava* and *S. pallida* (Fig. 3c). Indeed, we noticed during the initial mapping experiment that there was a small response towards bitter compounds in *S. flava* S-c sensilla (Fig. 3a). The negative dose dependent responses, however, suggest that these spikes correspond to the water cells that show higher firing rates in response to lower solute concentrations (Meunier et al., 2003). In *D. melanogaster*, the inclusion of tricholine citrate (>30mM) in the stimulating solution is effective at abolishing water cell firing (Wieczorek and Wolff, 1989). Tricholine citrate appears to be less effective in this regard within the *Scaptomyza* species. In preliminary recordings, higher concentrations of tricholine citrate were used (up to 300 mM), but there was no reduction in spike rate. Across all sensilla types, spike responses from the water cell are only presented in our results when no other neurons within the sensilla were activated. If sweet or bitter sensitive neurons fired, the spikes from the water cell could be differentiated from these neurons based on the amplitude difference in spikes (Fig. 3d). These spikes from water cells were not typically seen in I type sensilla, consistent with the lack of water cells in these sensilla (Hiroi et al., 2004).

Lastly, although we only found one I type in *Scaptomyza* and no S-b type in *S. pallida*, we recorded from sensilla that could represent these sensilla types based on orthologous morphology and location (Fig. S2a). The results confirm our initial findings that these sensilla types are no longer present in these species (Fig. S2b).

Reduced sensitivity towards mustard-derived glucosinolates

To understand how herbivore responses may have evolved towards host-specific compounds, we performed electrophysiology on the representative sensilla seen in Fig. 3b. We tested responses towards three glucosinolates, one from each glucosinolate class: gluconasturtiin (aromatic), neoglucobrassicin (indole), and sinigrin (aliphatic). The results were consistent with our findings on their behavioral responses to these compounds. *S. pallida* showed strong behavioral aversion to all three glucosinolates. This species was also strongly sensitive to the glucosinolates gluconasturtiin and neoglucobrassicin, which were both detected by I type and S-a type sensilla (Fig. 4a). In contrast, *S. flava*, which had reduced behavioral aversion to these glucosinolates compared to the non-herbivores, also showed reduced sensitivity to these compounds among I type sensilla (Fig. 4a; neoglucobrassicin: GLM, $t = -2.04$, $df = 68$, $P = 0.05$; gluconasturtiin: not significant but trending lower). Interestingly, although S-a sensilla detected gluconasturtiin and neoglucobrassicin, there was no reduced sensitivity compared to responses in *S. pallida*, and S-b sensilla showed no response at all to gluconasturtiin and neoglucobrassicin. This lack of response in S-b sensilla in *S. flava* is surprising, given their increased sensitivity towards lobeline (Fig. 3a and 3c). These

results collectively suggest that *S. flava*'s loss of aversion to some glucosinolates was achieved through reduced bitter sensitivity in I type sensilla.

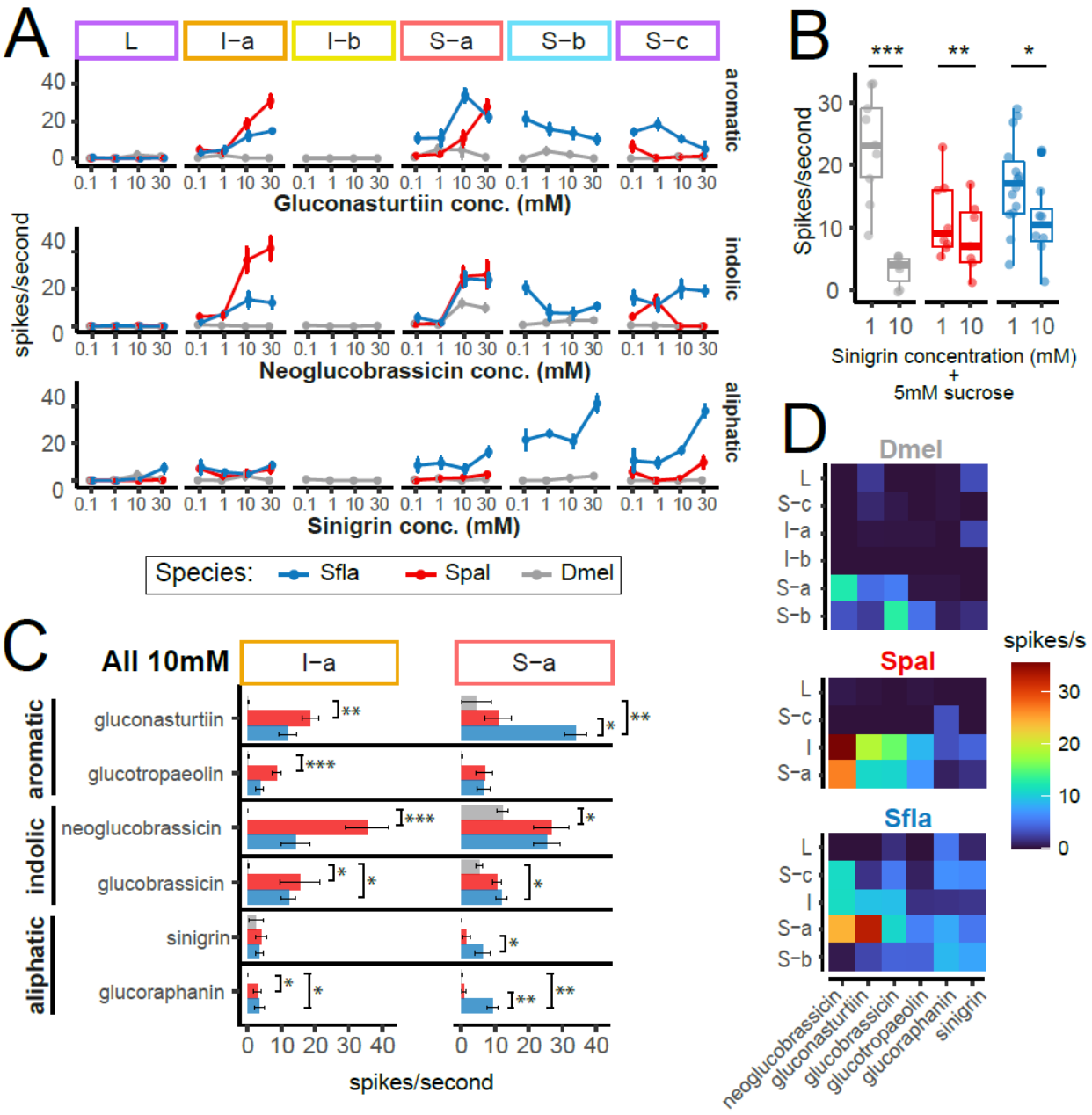


Figure 4. Glucosinolate sensitivity across labellar sensilla class. (a) Responses of representative sensilla to three representative glucosinolates: gluconasturtiin (aromatic), neoglucobrassicin (indolic), and sinigrin (aliphatic). (b) Spike responses to a mixed solution of sucrose (5 mM) and sinigrin at two concentrations (1 and 10 mM), showing reduced sweet neuron firing in the presence of sinigrin in all three species. (c) Responses to six glucosinolates (10 mM), including 3 additional glucosinolates, one additional from each glucosinolates class. Error bars represent the standard error of the mean. (d) Heatplot of spike responses towards all glucosinolates (10mM).

D. melanogaster showed the lowest levels of aversion in feeding experiments, and their electrophysiological responses towards all three glucosinolates were also extremely low (Fig. 4a). Considering *D. melanogaster* did show some glucosinolate aversion, the lack of spike responses from labellar sensilla could indicate that tarsal sensilla play a role in mediating aversion to these compounds. Tarsal responses to sinigrin in *D. melanogaster* have been found previously, although weakly as well (Ling et al., 2014).

In contrast to aromatic and indolic glucosinolates, the responses to sinigrin, an aliphatic glucosinolate, were significantly different across all sensilla in all species. Both *S. pallida* and *D. melanogaster* showed little sensitivity towards sinigrin across all sensilla. Strikingly, *S. flava* exhibited a positive correlation between sinigrin concentration and spike rate, even in the S-c type sensilla that we would not expect to respond to a bitter compound (Fig. 4a). While it is possible that a bitter gustatory neuron could be responding to sinigrin in the labellar sensilla, it may be that a salt-responsive gustatory neuron was stimulated by the salt in the sinigrin solution (Fujishiro et al., 1984). This still begs the question as to how species like *S. pallida*, *D. melanogaster*, and even still *S. flava* detect glucosinolates like sinigrin. As mentioned earlier, in *D. melanogaster*, singrin is detected by tarsal sensilla (Ling et al., 2014), and it could be possible that some glucosinolates, like sinigrin, are only detected by tarsi. An alternative or additional hypothesis is that sinigrin aversion may be caused by direct or indirect effects of sinigrin on sweet neuron spiking. We addressed this with recordings involving solutions filled with both sucrose and sinigrin and tested high (10mM) and low (1mM) sinigrin concentrations. There were significant differences between the spike rate between high and low sinigrin concentrations in all three species (Fig. 4b; *D. melanogaster*: $t(19) = 4.15$, $P < 0.001$; *S. pallida*: $t(10) = 4.47$, $P = 0.001$; *S. flava*: $t(15) = 2.11$, $P = 0.05$), which suggests that sinigrin (and potentially other glucosinolates) may be inhibiting sweet neuron firing as well. While this suggests one mechanism in which aliphatic glucosinolates elicit aversion, it remains unclear whether they are directly detected by gustatory receptors.

We performed electrophysiological recordings on three additional glucosinolates, one from each glucosinolate class – glucotropaeolin (aromatic), glucobrassicin (indolic), and glucoraphanin (aliphatic) – to test whether I and S-a responses were consistent within glucosinolate classes. Spike rates varied among the aromatic and indolic glucosinolates (all tested at 10mM), but there was a consistent trend, although not statistically significant, that *S. flava* had lower spike rates than *S. pallida* among I sensilla (Fig. 4c). Similarly, responses to the aliphatic glucosinolate glucoraphanin were low, as with sinigrin.

Discussion

Understanding how herbivore specialists evolve feeding preferences for their new host plants, particularly towards host plant defense compounds, is a key component to understanding their adaptation to this new niche and their exceptional diversity. Here, we investigated how the taste preferences of a mustard specialist, *S. flava*, have evolved towards glucosinolates, compounds specific to the order Brassicales, upon which this species feeds. We found that *S. flava* is less averse to glucosinolates, and this appears

to be attributable to the reduced sensitivity of I type sensilla towards glucosinolates and another bitter compound (caffeine).

While most generalist herbivores and non-herbivores find glucosinolates aversive, many mustard specialists have evolved attraction towards these compounds. We confirmed that non-herbivorous drosophilids (*S. pallida*, *S. hsui*, and *D. melanogaster*) are averse to glucosinolates, with *S. pallida* and *S. shui* showing especially strong sensitivity to glucosinolates. These findings are consistent with the fact that these two non-herbivores likely encounter these compounds while feeding on decaying plant material. However, we also found that the young mustard specialist *S. flava* was still averse to higher concentrations of glucosinolates, and this aversion, in *Scaptomyza* herbivores and non-herbivores alike, is mediated by specific sensilla classes that house bitter gustatory neurons, specifically I type and S-a type sensilla. Maintaining the capability of sensing glucosinolates is not surprising given that *S. flava* is still an evolutionarily young herbivore lineage and is still directly influenced by the toxic effects of these host plant defenses, as well as other toxins it may encounter in the environment. Specifically, *S. flava* development is slowed and weight gain is reduced by the presence of glucosinolates (Gloss et al., 2019). Moreover, *S. flava* is unique among known specialists of Brassicales plants, like cabbage white butterflies and diamondback moths, in not having a physiologic mechanism that prevents the formation of toxic isothiocyanates and other hydrolysis products in its body. Drosophilids, including *S. flava* and *D. melanogaster,* use the ancient mercapturic acid pathway to detoxify glucosinolates just like humans (Gloss et al. 2019). At lower glucosinolate concentrations, *S. flava* has lost or reduced its aversion to most of the glucosinolates tested, which enables these herbivores to feed on plants harboring lower levels of these compounds. Estimates of glucosinolate concentrations from leaves have been reported in the range of 0-10mM in some plant species (Merritt, 1996). But glucosinolate concentrations can vary depending on the age of the plant, the species, the tissue type, and environmental conditions (Feeny and Rosenberry, 1982; Brown et al., 2003; Wentzell and Kliebenstein, 2008). The reduced glucosinolate sensitivity in *S. flava* would thus allow them to sense and avoid higher, toxic concentrations, while readily feeding on lower, less toxic concentrations.

A lingering question is whether there is gustatory attraction towards any glucosinolates – potentially in a context-dependent manner (e.g., positional attraction or oviposition stimulation) in *S. flava*. We know from *D. melanogaster* that the same compounds that can be aversive while feeding can also be stimulatory for egg-laying through differential expression in different gustatory tissues (Joseph and Heberlein, 2012). One *Scaptomyza* herbivore, *S. nigrita*, which feeds exclusively on the mustard *Cardamine cordifolia*, shows a preference for laying eggs on the lower rosette leaves of their host, which are lower in glucosinolate content (Humphrey et al., 2016), suggesting ovipositional stimulation by glucosinolates may not have evolved in young mustard specialists. We do know that volatile isothiocyanates activate olfactory receptors in *S. flava* that are specifically tuned to these mustard oils – mediated by recently triplicated copies of the odorant receptor gene *Or67b* (Matsunaga et al. 2021). Further study will be needed to test whether olfactory cues alone stimulate feeding and egg-laying in herbivorous *Scaptomyza*, or whether additional cues are necessary to elicit these behaviors.

We suspected that there would be differences in the responses of mustard feeders towards different classes of glucosinolates (aliphatic, aromatic, and indolic). While testing with more compounds and at more concentrations is needed to conclusively determine whether class differences exist, we found that responses varied across different glucosinolates, with some trends across different classes. In particular the strongest reduction in aversion was in response to indole glucosinolates (Fig. 1d) (although there was not a dramatic drop in sensilla sensitivity relative to other glucosinolates (Fig. 4a and 4c)). We might expect that mustard specialists may evolve greater losses of aversion to these compounds because the isothiocyanates formed by indole glucosinolates are unstable and break down rapidly (Chevolleau et al., 1997; Agerbirk et al., 1998), leaving potentially less toxic end products like indole-3-carbinol. It is thus possible that indole glucosinolates may be a less reliable indicator of toxicity for these herbivores. Over 120 glucosinolates have been found naturally occurring in mustard plants, and this chemical diversity has evolved to evade herbivore counter-adaptations (Fahey et al., 2001). It is thus unsurprising that different glucosinolates elicit different responses across the flies we tested.

The diversity of plant chemical defenses is matched by the diversity of gustatory receptors that have evolved in insects to detect bitter compounds: out of the roughly 68 GR proteins in *D. melanogaster*, seven of them function to detect sugars (Dahanukar et al., 2007; Slone et al., 2007; Fujii et al., 2015), while almost half of the remaining GRs are involved in bitter detection or are at least expressed in bitter gustatory neurons (Moon et al., 2006; Lee et al., 2009, 2012; Weiss et al., 2011; Sung et al., 2017). While we still do not know which gustatory receptors are expressed in *S. flava* and in which gustatory neurons and sensilla, in Chapter 2, I discussed some of the GR changes that were specific to *Scaptomyza* herbivores. Copy number changes only affected bitter GR genes – most notably losses of *Gr28bA, Gr59e, Gr59f*, and paralogs of *Gr39aA, Gr59ab, Gr59cd* – and a duplication of *Obp49a*. *D. melanogaster* orthologs of *Gr59e* and *Gr59f* are specific to larvae (Kwon et al., 2011), and these losses could be involved in larval changes in feeding preferences. Meanwhile, orthologs of *Gr59ab* and *Gr59cd* are restricted to short S-a and S-b type sensilla and thus their losses may not be strongly responsible for reductions in glucosinolate aversion. *Gr28bA* and *Gr39aA*, however, are strong contenders for causing this gustatory shift, as they are expressed in both I and S type sensilla (Fig. 5). In *D. melanogaster*, *Gr28bA* has been implicated in the detection of saponin, triterpene glycosides found in a variety of flowering plant families (Sang et al., 2019). *Gr39aA* is a splice variant of *Gr39a* that is critically important for the detection of a wide variety of bitter compounds (Weiss et al., 2011; Dweck and Carlson, 2020). But while only a single copy is found in *D. melanogaster,* this splice variant uniquely has undergone numerous duplications in many other species (Gardiner et al., 2008), including *S. pallida* and *S. hsui* that both have nine copies each. Meanwhile, *S. graminum*, which is not a mustard specialist but rather a specialist on Caryophyllaceae (Hackman, 1959) has eight copies. *S. flava* has six, and *S. montana,* an even more specialized mustard feeder than *S. flava*, has only four. The number of *Gr39aA* copies could have a dosage effect on bitter reception, which is further suggested by losses of *Gr39aA* in several other specialist lineages, including in noni and Pandanus fruit specialists (*D. sechellia* and *D.*

erecta), cactus rot specialists (*D. navajoa* and *D. mojavensis*), *Lordiphosa* spp. leafminers (*L. magnipectinata, L. clarofinis, L. collinella*), and Hawaiian bark/stem specialists on Araliaceae (*D. sproati*) (Chapter 2). Interestingly, the loss of *Gr39a* in *D. melanogaster* caused a reduction in caffeine sensitivity in I-b type sensilla, while also causing an increase in lobeline sensitivity in I-a type sensilla (Dweck et al. 2020). The implications of this are muddled considering *Scaptomyza* only has a single I type sensilla, but it does suggest a way in which the loss of a GR could cause the simultaneous reduction in sensitivity towards some compounds, while increasing sensitivity to others. Increased sensitivity due to GR loss has been shown for several other GRs and compounds as well (Dweck and Carlson, 2020). It is worthwhile to note that like *S. flava*, the noni specialist *D. sechellia* has also lost caffeine sensitivity. The researchers who discovered this in *D. sechellia* similarly speculate that it may prevent high-level activation of aversive taste circuits (Dweck et al., 2021).

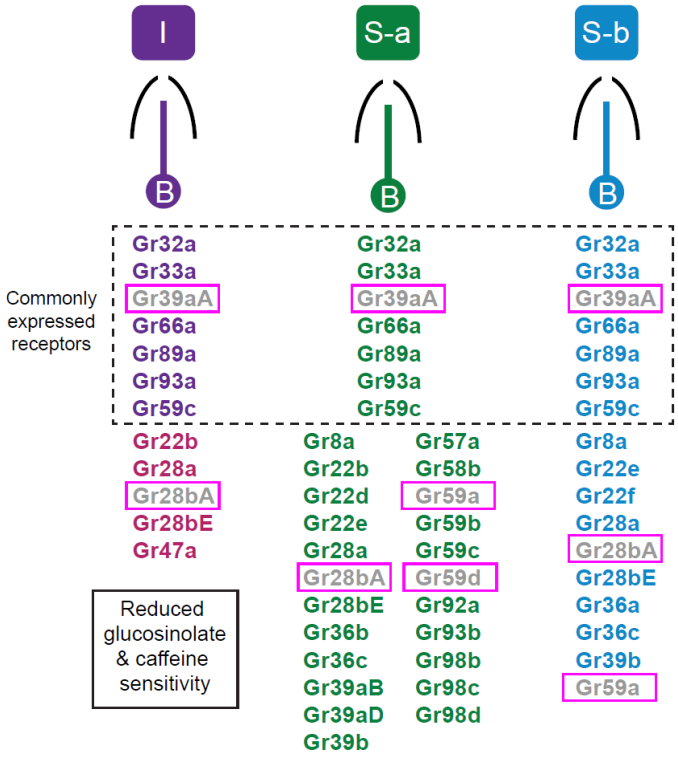


Figure 5. Candidate changes in gustatory receptors in the herbivore *Scaptomyza flava*. Schematic of bitter gustatory neurons within I, S-a and S-b type sensilla. Shown are those GRs expressed in homologous sensilla in *D. melanogaster*. GRs that are found in all bitter gustatory neurons in *D. melanogaster* are delineated as those in the dashed box. GRs that have been lost or pseudogenized in *S. flava*, relative to the *S. pallida*, are highlighted with pink boxes.

In summary, our study supports an evolutionary model wherein reduced bitter sensitivity at the level of sensory neurons has enabled an herbivorous drosophild to feed on living vegetative plants tissues rife with bitter and toxic plant defense chemicals. The redundancy of the gustatory taste system has allowed these flies to maintain some bitter sensitivity, critical for insects still inhabiting complex environments in which they encounter diverse toxins. With the identification of candidate GRs that have been lost in *S. flava*, particularly *Gr39aA*, it will be interesting to functionally test the role of these genetic changes in driving feeding preferences.

Concluding Remarks

Through my dissertation work, I sought to understand the evolution of herbivorous insects and specific adaptations that enabled them to colonize their new host plants. Using the *Scaptomyza* herbivore system, I investigated plant cutting ovipositors in Chapter 1, shifts in chemosensation and detoxification in Chapter 2, and in Chapter 3, changes in bitter feeding preferences. Altogether, these results suggest that insects can evolve new traits to facilitate herbivorous feeding through regulatory changes affecting conserved developmental genes (Chapter 1), or through dynamic remodeling of whole gene families (Chapter 2). They can evolve incrementally as in the case of many small-effect mutations giving rising to a highly dentate ovipositor, but also in the case of more and more bitter gustatory receptors being purged from the genome to reduce bitter sensitivity. In summary, this work highlights the incredible insight we can gain into the evolution of herbivory, by leveraging developing resources for the herbivore lineage in *Scaptomyza* with the wealth of genetic knowledge on *Drosophila melanogaster*.

References

Agerbirk, N., Olsen, C. E., and Sørensen, H. (1998). Initial and Final Products, Nitriles, and Ascorbigens Produced in Myrosinase-Catalyzed Hydrolysis of Indole Glucosinolates. *J. Agric. Food Chem.* 46, 1563–1571.

Almeida, F. C., Sánchez-Gracia, A., Campos, J. L., & Rozas, J. (2014). Family Size Evolution in Drosophila Chemosensory Gene Families: A Comparative Analysis with a Critical Appraisal of Methods. In *Genome Biology and Evolution* (Vol. 6, Issue 7, pp. 1669–1682). https://doi.org/10.1093/gbe/evu130

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, *25*(17), 3389–3402.

Alves, G., Sallé, J., Chaudy, S., Dupas, S., & Manière, G. (2014). High-NaCl perception in Drosophila melanogaster. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *34*(33), 10884–10891.

Armstrong, J., Hickey, G., Diekhans, M., Fiddes, I. T., Novak, A. M., Deran, A., Fang, Q., Xie, D., Feng, S., Stiller, J., Genereux, D., Johnson, J., Marinescu, V. D., Alföldi, J., Harris, R. S., Lindblad-Toh, K., Haussler, D., Karlsson, E., Jarvis, E. D., … Paten, B. (2020). Progressive Cactus is a multiple-genome aligner for the thousand-genome era. *Nature*, *587*(7833), 246–251.

Aryal, B., & Lee, Y. (2022). Histamine avoidance through three gustatory receptors in Drosophila melanogaster. *Insect Biochemistry and Molecular Biology*, *144*, 103760.

Atallah, J., Teixeira, L., Salazar, R., Zaragoza, G., & Kopp, A. (2014). The making of a pest: the evolution of a fruit-penetrating ovipositor in Drosophila suzukii and related species. *Proceedings. Biological Sciences / The Royal Society*, *281*(1781), 20132840.

Atallah, J., Watabe, H., & Kopp, A. (2012). Many ways to make a novel structure: a new mode of sex comb development in Drosophilidae. *Evolution & Development*, *14*(6), 476–483.

Ayres, D. L., Darling, A., Zwickl, D. J., Beerli, P., Holder, M. T., Lewis, P. O., Huelsenbeck, J. P., Ronquist, F., Swofford, D. L., Cummings, M. P., Rambaut, A., & Suchard, M. A. (2012). BEAGLE: an application programming interface and high-performance computing library for statistical phylogenetics. *Systematic Biology*, *61*(1), 170–173.

Balakrishnan, S. S., Basu, U., & Raghu, P. (2015). Phosphoinositide signalling in Drosophila. *Biochimica et Biophysica Acta*, *1851*(6), 770–784.

Bastide, H., Betancourt, A., Nolte, V., Tobler, R., Stöbe, P., Futschik, A., & Schlötterer, C. (2013). A genome-wide, fine-scale map of natural pigmentation variation in Drosophila melanogaster. *PLoS Genetics*, *9*(6), e1003534.

Bendall, E. E., Vertacnik, K. L., & Linnen, C. R. (2017). Oviposition traits generate extrinsic postzygotic isolation between two pine sawfly species. *BMC Evolutionary Biology*, *17*(1), 26.

Berenbaum, M. (1983). Coumarins and Caterpillars: A Case for Coevolution. *Evolution; International Journal of Organic Evolution*, *37*(1), 163–179.

Berenbaum, M. R., & Zangerl, A. R. (2008). Facing the future of plant-insect interaction research: le retour à la “raison d’être.” *Plant Physiology*, *146*(3), 804.

Bernays, E. A. (1998). Evolution of Feeding Behavior in Insect Herbivores. *Bioscience* 48, 35–44.
Bernays, E. A. (2001). Neural limitations in phytophagous insects: implications for diet breadth and evolution of host affiliation. *Annual Review of Entomology*, *46*, 703–727.
Bernays, E. A., and Chapman, R. E. eds. (1994). "Behavior: The Process of Host-Plant Selection," in *Host-Plant Selection by Phytophagous Insects* (Boston, MA: Springer US), 95–165.
Bernays, E. A., Oppenheim, S., Chapman, R. F., Kwon, H., and Gould, F. (2000). Taste sensitivity of insect herbivores to deterrents is greater in specialists than in generalists: a behavioral test of the hypothesis with two closely related caterpillars. *J. Chem. Ecol.* 26, 547–563.
Bhat, K. M., van Beers, E. H., & Bhat, P. (2000). Sloppy paired acts as the downstream target of wingless in the Drosophila CNS and interaction between sloppy paired and gooseberry inhibits sloppy paired during neurogenesis. *Development, 127*(3), 655–665.
Bidart-Bouzat, M. G., and Kliebenstein, D. J. (2008). Differential levels of insect herbivory in the field associated with genotypic variation in glucosinolates in Arabidopsis thaliana. *J. Chem. Ecol.* 34, 1026–1037.
Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120.
Bolger, Anthony M., Marc Lohse, and Bjoern Usadel. 2014. "Trimmomatic: A Flexible Trimmer for Illumina Sequence Data." *Bioinformatics* 30 (15): 2114–20.
Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.-H., Xie, D., Suchard, M. A., Rambaut, A., & Drummond, A. J. (2014). BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Computational Biology*, *10*(4), e1003537.
Bourgeois, Y., & Boissinot, S. (2019). On the Population Dynamics of Junk: A Review on the Population Genomics of Transposable Elements. *Genes*, *10*(6), 419.
Bray, S., & Amrein, H. (2003). A putative Drosophila pheromone receptor expressed in male-specific taste neurons is required for efficient courtship. *Neuron*, *39*(6), 1019–1029.
Briscoe, A. D., Macias-Muñoz, A., Kozak, K. M., Walters, J. R., Yuan, F., Jamie, G. A., Martin, S. H., Dasmahapatra, K. K., Ferguson, L. C., Mallet, J., Jacquin-Joly, E., & Jiggins, C. D. (2013). Female behaviour drives expression and evolution of gustatory receptors in butterflies. *PLoS Genetics*, *9*(7), e1003620.
Brown, P. D., Tokuhisa, J. G., Reichelt, M., and Gershenzon, J. (2003). Variation of glucosinolate accumulation among different organs and developmental stages of Arabidopsis thaliana. *Phytochemistry* 62, 471–481.
Bushnell, B. 2014. "BBTools Software Package."
Butler, M. A., & King, A. A. (2004). Phylogenetic Comparative Analysis: A Modeling Approach for Adaptive Evolution. *The American Naturalist*, *164*(6), 683–695.
Calla, B., Noble, K., Johnson, R. M., Walden, K. K. O., Schuler, M. A., Robertson, H. M., & Berenbaum, M. R. (2017). Cytochrome P450 diversification and hostplant utilization patterns in specialist and generalist moths: Birth, death and adaptation. *Molecular Ecology*, *26*(21), 6021–6035.

Carlson, M., Falcon, S., Pages, H., & Li, N. (2017). GO. db: A set of annotation maps describing the entire Gene Ontology. *R Package Version*, *3*(0), 10–18129.
Carscadden, K., Batstone, R., & Hauser, F. (n.d.). *No title*. Retrieved August 22, 2022, from https://ecoevorxiv.org/yujcb
Chevolleau, S., Gasc, N., Rollin, P., and Tulliez, J. (1997). Enzymatic, Chemical, and Thermal Breakdown of 3H-Labeled Glucobrassicin, the Parent Indole Glucosinolate. *J. Agric. Food Chem.* 45, 4290–4296.
Chew, F. S., and Renwick, J. A. A. (1995). "Host Plant Choice in Pieris Butterflies," in *Chemical Ecology of Insects 2*, eds. R. T. Cardé and W. J. Bell (Boston, MA: Springer US), 214–238.
Chung, H., Sztal, T., Pasricha, S., Sridhar, M., Batterham, P., & Daborn, P. J. (2009). Characterization of Drosophila melanogaster cytochrome P450 genes. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(14), 5731–5736.
Clayton, J. D., Cripps, R. M., Sparrow, J. C., & Bullard, B. (1998). Interaction of troponin-H and glutathione S-transferase-2 in the indirect flight muscles of Drosophila melanogaster. *Journal of Muscle Research and Cell Motility*, *19*(2), 117–127.
Clyne, P. J., Warr, C. G., and Carlson, J. R. (2000). Candidate taste receptors in Drosophila. *Science* 287, 1830–1834.
Connor, E. F., & Taverner, M. P. (1997). The Evolution and Adaptive Significance of the Leaf-Mining Habit. *Oikos*, *79*(1), 6–25.
Couturier, L., Mazouni, K., Corson, F., & Schweisguth, F. (2019). Regulation of Notch output dynamics via specific E (spl)-HLH factors during bristle patterning in Drosophila. *Nature Communications*, *10*(1), 1–13.
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.
Crava, C. M., Zanini, D., Amati, S., Sollai, G., Crnjar, R., Paoli, M., Rossi-Stacconi, M. V., Rota-Stabelli, O., Tait, G., Haase, A., Romani, R., & Anfora, G. (2020). Structural and transcriptional evidence of mechanotransduction in the Drosophila suzukii ovipositor. *Journal of Insect Physiology*, *125*, 104088.
Dahanukar, A., Lei, Y.-T., Kwon, J. Y., and Carlson, J. R. (2007). Two Gr genes underlie sugar reception in Drosophila. *Neuron* 56, 503–516.
David, J. R., Lemeunier, F., Tsacas, L., & Yassin, A. (2007). The historical discovery of the nine species in the Drosophila melanogaster species subgroup. *Genetics*, *177*(4), 1969–1973.
Davies, L., Williams, D. R., Aguiar-Santana, I. A., Pedersen, J., Turner, P. C., & Rees, H. H. (2006). Expression and down-regulation of cytochrome P450 genes of the CYP4 family by ecdysteroid agonists in Spodoptera littoralis and Drosophila melanogaster. *Insect Biochemistry and Molecular Biology*, *36*(10), 801–807.
De Bie, T., Cristianini, N., Demuth, J. P., & Hahn, M. W. (2006). CAFE: a computational tool for the study of gene family evolution. *Bioinformatics, 22*(10), 1269–1271.

Dekker, T., Ibba, I., Siju, K. P., Stensmyr, M. C., & Hansson, B. S. (2006). Olfactory shifts parallel superspecialism for toxic fruit in Drosophila melanogaster sibling, D. sechellia. *Current Biology: CB*, *16*(1), 101–109.

Dermauw, W., Osborne, E. J., Clark, R. M., Grbić, M., Tirry, L., & Van Leeuwen, T. (2013). A burst of ABC genes in the genome of the polyphagous spider mite Tetranychus urticae. *BMC Genomics*, *14*, 317.

Devlin, B., & Roeder, K. (1999). Genomic control for association studies. *Biometrics*, *55*(4), 997–1004.

Dmitriev, D. A., & Rakitov, R. A. (2008). Decoding of superimposed traces produced by direct sequencing of heterozygous indels. *PLoS Computational Biology*, *4*(7), e1000113.

Dobin, Alexander, Carrie A. Davis, Felix Schlesinger, Jorg Drenkow, Chris Zaleski, Sonali Jha, Philippe Batut, Mark Chaisson, and Thomas R. Gingeras. 2013. "STAR: Ultrafast Universal RNA-Seq Aligner." *Bioinformatics* 29 (1): 15–21.

Drosophila 12 Genomes Consortium, 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., … MacCallum, I. (2007). Evolution of genes and genomes on the Drosophila phylogeny. *Nature*, *450*(7167), 203–218.

Dulai, K. S., von Dornum, M., Mollon, J. D., & Hunt, D. M. (1999). The evolution of trichromatic color vision by opsin gene duplication in New World and Old World primates. *Genome Research*, *9*(7), 629–638.

Durkin, S. M., Chakraborty, M., Abrieux, A., Lewald, K. M., Gadau, A., Svetec, N., Peng, J., Kopyto, M., Langer, C. B., Chiu, J. C., Emerson, J. J., & Zhao, L. (2021). Behavioral and Genomic Sensory Adaptations Underlying the Pest Activity of Drosophila suzukii. *Molecular Biology and Evolution*, *38*(6), 2532–2546.

Dweck, H. K. M., and Carlson, J. R. (2020). Molecular Logic and Evolution of Bitter Taste in Drosophila. *Curr. Biol.* 30, 17–30.e3.

Dweck, H. K., Talross, G. J., Wang, W., and Carlson, J. R. (2021). Evolutionary shifts in taste coding in the fruit pest Drosophila suzukii. *Elife* 10. doi: 10.7554/eLife.64317.

Dworkin, I., & Jones, C. D. (2009). Genetic changes accompanying the evolution of host specialization in Drosophila sechellia. *Genetics*, *181*(2), 721–736.

Ebrahim, S. A. M., Dweck, H. K. M., Stökl, J., Hofferberth, J. E., Trona, F., Weniger, K., Rybak, J., Seki, Y., Stensmyr, M. C., Sachse, S., Hansson, B. S., & Knaden, M. (2015). Drosophila Avoids Parasitoids by Sensing Their Semiochemicals via a Dedicated Olfactory Circuit. *PLoS Biology*, *13*(12), e1002318.

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, *32*(5), 1792–1797.

Edger, P. P., Heidel-Fischer, H. M., Bekaert, M., Rota, J., Glöckner, G., Platts, A. E., Heckel, D. G., Der, J. P., Wafula, E. K., Tang, M., Hofberger, J. A., Smithson, A., Hall, J. C., Blanchette, M., Bureau, T. E., Wright, S. I., dePamphilis, C. W., Eric Schranz, M., Barker, M. S., … Wheat, C. W. (2015). The butterfly plant arms-race escalated by gene and genome duplications. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(27), 8362–8366.

Ehrlich, P. R., and Raven, P. H. (1964). Butterflies and Plants: A Study in Coevolution. *Evolution* 18, 586–608.

Eiseman, C., Charney, N., & Carlson, J. (2010). *Tracks & Sign of Insects & Other Invertebrates: A Guide to North American Species*. Stackpole Books.

Etges, W. J., de Oliveira, C. C., Ritchie, M. G., & Noor, M. A. F. (2009). Genetics of incipient speciation in Drosophila mojavensis: II. Host plants and mating status influence cuticular hydrocarbon QTL expression and G x E interactions. *Evolution; International Journal of Organic Evolution*, *63*(7), 1712–1730.

Fahey, J. W., Zalcmann, A. T., and Talalay, P. (2001). The chemical diversity and distribution of glucosinolates and isothiocyanates among plants. *Phytochemistry* 56, 5–51.

Farrell, B. D. (1998). " Inordinate fondness" explained: Why are there so many beetles? *Science*. https://science.sciencemag.org/content/281/5376/555.short

Feder, A. F., Petrov, D. A., & Bergland, A. O. (2012). LDx: estimation of linkage disequilibrium from high-throughput pooled resequencing data. *PloS One*, *7*(11), e48588.

Feeny, P., and Rosenberry, L. (1982). Seasonal variation in the glucosinolate content of North American Brassica nigra and Dentaria species. *Biochem. Syst. Ecol.* 10, 23–32.

Ferveur, J. F., Savarit, F., O'Kane, C. J., Sureau, G., Greenspan, R. J., & Jallon, J. M. (1997). Genetic feminization of pheromones and its behavioral consequences in Drosophila males. *Science*, *276*(5318), 1555–1558.

Fiddes, I. T., Armstrong, J., Diekhans, M., Nachtweide, S., Kronenberg, Z. N., Underwood, J. G., Gordon, D., Earl, D., Keane, T., Eichler, E. E., Haussler, D., Stanke, M., & Paten, B. (2018). Comparative Annotation Toolkit (CAT)—simultaneous clade and personal genome annotation. *Genome Research*, *28*(7), 1029–1038.

Finet, C., Kassner, V. A., Carvalho, A. B., Chung, H., Day, J. P., Day, S., Delaney, E. K., De Ré, F. C., Dufour, H. D., Dupim, E., Izumitani, H. F., Gautério, T. B., Justen, J., Katoh, T., Kopp, A., Koshikawa, S., Longdon, B., Loreto, E. L., Nunes, M. D. S., … Marlétaz, F. (2021). DrosoPhyla: Resources for Drosophilid Phylogeny and Systematics. *Genome Biology and Evolution*, *13*(8). https://doi.org/10.1093/gbe/evab179

Flynn, J. M., Hubley, R., Goubert, C., Rosen, J., Clark, A. G., Feschotte, C., & Smit, A. F. (2020). RepeatModeler2 for automated genomic discovery of transposable element families. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(17), 9451–9457.

Fraenkel, G. S. (1959). The raison d'etre of secondary plant substances. *Science*. https://www.jstor.org/stable/1756998

French, A., Ali Agha, M., Mitra, A., Yanagawa, A., Sellier, M.-J., and Marion-Poll, F. (2015). Drosophila Bitter Taste(s). *Front. Integr. Neurosci.* 9, 58.

Fujii, S., Yavuz, A., Slone, J., Jagge, C., Song, X., and Amrein, H. (2015). Drosophila sugar receptors in sweet taste perception, olfaction, and internal nutrient sensing. *Curr. Biol.* 25, 621–627.

Fujishiro, N., Kijima, H., and Morita, H. (1984). Impulse frequency and action potential amplitude in labellar chemosensory neurones of Drosophila melanogaster. *J. Insect Physiol.* 30, 317–325.

Futuyma, D. J., & Agrawal, A. A. (2009). Macroevolution and the biological diversity of plants and herbivores. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(43), 18054–18061.

Gago-Dominguez, M., Jiang, X., & Castelao, J. E. (2007). Lipid peroxidation, oxidative stress genes and dietary factors in breast cancer protection: a hypothesis. *Breast Cancer Research: BCR*, *9*(1), 201.

Gardiner, A., Barker, D., Butlin, R. K., Jordan, W. C., & Ritchie, M. G. (2008). Drosophila chemoreceptor gene evolution: selection, specialization and genome size. *Molecular Ecology*, *17*(7), 1648–1657.

Gardiner, A., Barker, D., Butlin, R. K., Jordan, W. C., and Ritchie, M. G. (2008). Evolution of a complex locus: exon gain, loss and divergence at the Gr39a locus in Drosophila. *PLoS One* 3, e1513.

Geiselhardt, S., Otte, T., & Hilker, M. (2012). Looking for a similar partner: host plants shape mating preferences of herbivorous insects by altering their contact pheromones. *Ecology Letters*, *15*(9), 971–977.

Gentleman, R. C., Carey, V. J., Bates, D. M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J., Hornik, K., Hothorn, T., Huber, W., Iacus, S., Irizarry, R., Leisch, F., Li, C., Maechler, M., Rossini, A. J., … Zhang, J. (2004). Bioconductor: open software development for computational biology and bioinformatics. *Genome Biology*, *5*(10), R80.

George, L. F., Pradhan, S. J., Mitchell, D., Josey, M., Casey, J., Belus, M. T., Fedder, K. N., Dahal, G. R., & Bates, E. A. (2019). Ion Channel Contributions to Wing Development in Drosophila melanogaster. *G3* , *9*(4), 999–1008.

Ghara, M., Kundanati, L., & Borges, R. M. (2011). Nature's Swiss Army knives: ovipositor structure mirrors ecology in a multitrophic fig wasp community. *PloS One*, *6*(8), e23642.

Gloss, A. D., Brachi, B., Feldmann, M. J., Groen, S. C., Bartoli, C., Gouzy, J., LaPlante, E. R., Meyer, C. G., Pyon, H. S., Rogan, S. C., Roux, F., Bergelson, J., & Whiteman, N. K. (2017). *Genetic variants affecting plant size and chemical defenses jointly shape herbivory inArabidopsis*. https://doi.org/10.1101/156299

Gloss, A. D., Nelson Dittrich, A. C., Lapoint, R. T., Goldman-Huertas, B., Verster, K. I., Pelaez, J. L., Nelson, A. D. L., Aguilar, J., Armstrong, E., Charboneau, J. L. M., Groen, S. C., Hembry, D. H., Ochoa, C. J., O'Connor, T. K., Prost, S., Suzuki, H. C., Zaaijer, S., Nabity, P. D., & Whiteman, N. K. (2019). Evolution of herbivory remodels a Drosophila genome. In *bioRxiv* (p. 767160). https://doi.org/10.1101/767160

Gloss, A. D., Vassão, D. G., Hailey, A. L., Nelson Dittrich, A. C., Schramm, K., Reichelt, M., Rast, T. J., Weichsel, A., Cravens, M. G., Gershenzon, J., Montfort, W. R., & Whiteman, N. K. (2014). Evolution in an ancient detoxification pathway is coupled with a transition to herbivory in the drosophilidae. *Molecular Biology and Evolution*, *31*(9), 2441–2456.

Gloss, Andrew D., Amélie Vergnol, Timothy C. Morton, Peter J. Laurin, Fabrice Roux, and Joy Bergelson. 2022. "Genome-Wide Association Mapping within a Local Arabidopsis Thaliana Population More Fully Reveals the Genetic Architecture for Defensive Metabolite Diversity." *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 377 (1855): 20200512.

Gnerre, Sante, Iain Maccallum, Dariusz Przybylski, Filipe J. Ribeiro, Joshua N. Burton, Bruce J. Walker, Ted Sharpe, et al. 2011. “High-Quality Draft Assemblies of Mammalian Genomes from Massively Parallel Sequence Data.” *Proceedings of the National Academy of Sciences of the United States of America* 108 (4): 1513–18.

Goldman-Huertas, B., Mitchell, R. F., Lapoint, R. T., Faucher, C. P., Hildebrand, J. G., & Whiteman, N. K. (2015). Evolution of herbivory in Drosophilidae linked to loss of behaviors, antennal responses, odorant receptors, and ancestral diet. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(10), 3026–3031.

Good, R. T., Gramzow, L., Battlay, P., Sztal, T., Batterham, P., & Robin, C. (2014). The molecular evolution of cytochrome P450 genes within and between drosophila species. *Genome Biology and Evolution*, *6*(5), 1118–1134.

Grbić, M., Van Leeuwen, T., Clark, R. M., Rombauts, S., Rouzé, P., Grbić, V., Osborne, E. J., Dermauw, W., Ngoc, P. C. T., Ortego, F., Hernández-Crespo, P., Diaz, I., Martinez, M., Navajas, M., Sucena, É., Magalhães, S., Nagy, L., Pace, R. M., Djuranović, S., … Van de Peer, Y. (2011). The genome of Tetranychus urticae reveals herbivorous pest adaptations. *Nature*, *479*(7374), 487–492.

Green, J. E., Cavey, M., Médina Caturegli, E., Aigouy, B., Gompel, N., & Prud’homme, B. (2019). Evolution of Ovipositor Length in Drosophila suzukii Is Driven by Enhanced Cell Size Expansion and Anisotropic Tissue Reorganization. *Current Biology: CB*, *29*(12), 2075–2082.e6.

Groen, S. C., & Whiteman, N. K. (2016). Using Drosophila to study the evolution of herbivory and diet specialization. *Current Opinion in Insect Science*, *14*, 66–72.

Guan, D., McCarthy, S. A., Wood, J., Howe, K., Wang, Y., & Durbin, R. (2020). Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics, 36*(9), 2896–2898.

Hackman, W. (1959). On the genus Scaptomyza Hardy (Dipt., Drosophilidae). *With descriptions of new species from various parts of*. Available at: https://core.ac.uk/download/pdf/14926346.pdf.

Hagen, J. F. D., Mendes, C. C., Booth, S. R., Figueras Jimenez, J., Tanaka, K. M., Franke, F. A., Baudouin-Gonzalez, L., Ridgway, A. M., Arif, S., Nunes, M. D. S., & McGregor, A. P. (2021). Unraveling the Genetic Basis for the Rapid Diversification of Male Genitalia between Drosophila Species. *Molecular Biology and Evolution*, *38*(2), 437–448.

Halkier, B. A., and Gershenzon, J. (2006). Biology and biochemistry of glucosinolates. *Annu. Rev. Plant Biol.* 57, 303–333.

Hammer, Ø., Harper, D. A. T., Ryan, P. D., and Others (2001). PAST: Paleontological statistics software package for education and data analysis. *Palaeontol. Electronica* 4, 9.

Han, M. V., Thomas, G. W. C., Lugo-Martinez, J., & Hahn, M. W. (2013). Estimating gene gain and loss rates in the presence of error in genome assembly and annotation using CAFE 3. *Molecular Biology and Evolution*, *30*(8), 1987–1997.

Harmon, L. J., Weir, J. T., Brock, C. D., Glor, R. E., & Challenger, W. (2008). GEIGER: investigating evolutionary radiations. *Bioinformatics, 24*(1), 129–131.

Hassan, B. A., Prokopenko, S. N., Breuer, S., Zhang, B., Paululat, A., & Bellen, H. J. (1998). skittles, a Drosophila phosphatidylinositol 4-phosphate 5-kinase, is required

for cell viability, germline development and bristle morphology, but not for neurotransmitter release. *Genetics*, *150*(4), 1527–1537.
Heidel-Fischer, H. M., & Vogel, H. (2015). Molecular mechanisms of insect adaptation to plant secondary compounds. *Current Opinion in Insect Science*, *8*, 8–14.
Hiroi, M., Marion-Poll, F., and Tanimura, T. (2002). Differentiated response to sugars among labellar chemosensilla in Drosophila. *Zoolog. Sci.* 19, 1009–1018.
Hiroi, M., Meunier, N., Marion-Poll, F., and Tanimura, T. (2004). Two antagonistic gustatory receptor neurons responding to sweet-salty and bitter taste in Drosophila. *J. Neurobiol.* 61, 333–342.
Hodgson, E. S., and Roeder, K. D. (1956). Electrophysiological studies of arthropod chemoreception. I. General properties of the labellar chemoreceptors of Diptera. *J. Cell. Comp. Physiol.* 48, 51–75.
Hoina, A., Martins, C. H. Z., Trigo, J. R., and Cogni, R. (2013). Preference for high concentrations of plant pyrrolizidine alkaloids in the specialist arctiid moth Utetheisa ornatrix depends on previous experience. *Arthropod Plant Interact.* 7, 169–175.
Holt, Carson, and Mark Yandell. 2011. "MAKER2: An Annotation Pipeline and Genome-Database Management Tool for Second-Generation Genome Projects." *BMC Bioinformatics* 12 (December): 491.
Hopkins, R. J., van Dam, N. M., and van Loon, J. J. A. (2008). Role of Glucosinolates in Insect-Plant Relationships and Multitrophic Interactions. doi: 10.1146/annurev.ento.54.110807.090623.
Humphrey, P. T., Gloss, A. D., Alexandre, N. M., Villalobos, M. M., Fremgen, M. R., Groen, S. C., et al. (2016). Aversion and attraction to harmful plant secondary compounds jointly shape the foraging ecology of a specialist herbivore. *Ecol. Evol.* 6, 3256–3268.
Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S., & Madden, T. L. (2008). NCBI BLAST: a better web interface. *Nucleic Acids Research*, *36*(Web Server issue), W5–W9.
Johnson, R. M., Harpur, B. A., Dogantzis, K. A., Zayed, A., & Berenbaum, M. R. (2018). Genomic footprint of evolution of eusociality in bees: floral food use and CYPome "blooms." *Insectes Sociaux*, *65*(3), 445–454.
Jones, C. D. (1998). The genetic basis of Drosophila sechellia's resistance to a host plant toxin. *Genetics*, *149*(4), 1899–1908.
Joseph, R. M., and Heberlein, U. (2012). Tissue-specific activation of a single gustatory receptor produces opposing behavioral responses in Drosophila. *Genetics* 192, 521–532.
Karageorgi, M., Groen, S. C., Sumbul, F., Pelaez, J. N., Verster, K. I., Aguilar, J. M., Hastings, A. P., Bernstein, S. L., Matsunaga, T., Astourian, M., Guerra, G., Rico, F., Dobler, S., Agrawal, A. A., & Whiteman, N. K. (2019). Genome editing retraces the evolution of toxin resistance in the monarch butterfly. *Nature*, *574*(7778), 409–412.
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.
Kawahara, A. Y., Plotkin, D., Espeland, M., Meusemann, K., Toussaint, E. F. A., Donath, A., Gimnich, F., Frandsen, P. B., Zwick, A., Dos Reis, M., Barber, J. R., Peters, R. S., Liu, S., Zhou, X., Mayer, C., Podsiadlowski, L., Storer, C., Yack, J.

E., Misof, B., & Breinholt, J. W. (2019). Phylogenomics reveals the evolutionary timing and pattern of butterflies and moths. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.1907847116

Kawakishi, S., & Kaneko, T. (1987). Interaction of proteins with allyl isothiocyanate. *Journal of Agricultural and Food Chemistry*, *35*(1), 85–88.

Kawakishi, S., Goto, T., and Namiki, M. (1983). Oxidative scission of the disulfide bond of cystine and polypeptides by the action of allyl isothiocyanate. *Agric. Biol. Chem.* Available at: https://www.jstage.jst.go.jp/article/bbb1961/47/9/47_9_2071/_article/-char/ja/.

Keesey, I. W., Knaden, M., & Hansson, B. S. (2015). Olfactory specialization in Drosophila suzukii supports an ecological shift in host preference from rotten to fresh fruit. *Journal of Chemical Ecology*, *41*(2), 121–128.

Kembel, S. W., Cowan, P. D., Helmus, M. R., Cornwell, W. K., Morlon, H., Ackerly, D. D., Blomberg, S. P., & Webb, C. O. (2010). Picante: R tools for integrating phylogenies and ecology. *Bioinformatics, 26*(11), 1463–1464.

Kim, B. Y., Wang, J., Miller, D. E., Barmina, O., Delaney, E. K., Thompson, A., Comeault, A. A., Peede, D., D'Agostino, E. R., Pelaez, J., Aguilar, J. M., Haji, D., Matsunaga, T., Armstrong, 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*. https://doi.org/10.7554/eLife.66405

Kim, H., Choi, M. S., Kang, K., & Kwon, J. Y. (2016). Behavioral Analysis of Bitter Taste Perception in Drosophila Larvae. *Chemical Senses*, *41*(1), 85–94.

Kofler, R., & Schlötterer, C. (2012). Gowinda: unbiased analysis of gene set enrichment for genome-wide association studies. *Bioinformatics* , *28*(15), 2084–2085.

Kofler, R., Pandey, R. V., & Schlötterer, C. (2011). PoPoolation2: identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). *Bioinformatics* , *27*(24), 3435–3436.

Kofler, Robert, Pablo Orozco-terWengel, Nicola De Maio, Ram Vinay Pandey, Viola Nolte, Andreas Futschik, Carolin Kosiol, and Christian Schlötterer. 2011. "PoPoolation: A Toolbox for Population Genetic Analysis of next Generation Sequencing Data from Pooled Individuals." *PloS One* 6 (1): e15925.

Kolmogorov, M., Raney, B., Paten, B., & Pham, S. (2014). Ragout—a reference-assisted assembly tool for bacterial genomes. *Bioinformatics, 30*(12), i302–i309.

Kolmogorov, M., Yuan, J., Lin, Y., & Pevzner, P. A. (2019). Assembly of long, error-prone reads using repeat graphs. *Nature Biotechnology*, *37*(5), 540–546.

Kvon, E. Z., Kazmar, T., Stampfel, G., Yáñez-Cuna, J. O., Pagani, M., Schernhuber, K., Dickson, B. J., & Stark, A. (2014). Genome-scale functional characterization of Drosophila developmental enhancers in vivo. *Nature*, *512*(7512), 91–95.

Kwon, J. Y., Dahanukar, A., Weiss, L. A., & Carlson, J. R. (2007). The molecular basis of CO2 reception in Drosophila. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(9), 3574–3578.

Laetsch, & Blaxter. (n.d.). BlobTools: Interrogation of genome assemblies [version 1; peer review: 2 approved with reservations]. F1000Res. 2017; 6: 1287. *Publisher Full Text*.

Lamichhaney, S., Han, F., Berglund, J., Wang, C., Almén, M. S., Webster, M. T., Grant, B. R., Grant, P. R., & Andersson, L. (2016). A beak size locus in Darwin's finches facilitated character displacement during a drought. *Science*, *352*(6284), 470–474.
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.
Lee, Y., Kang, M. J., Shim, J., Cheong, C. U., Moon, S. J., and Montell, C. (2012). Gustatory receptors required for avoiding the insecticide L-canavanine. *J. Neurosci.* 32, 1429–1435.
Lee, Y., Moon, S. J., and Montell, C. (2009). Multiple gustatory receptors required for the caffeine response in Drosophila. *Proc. Natl. Acad. Sci. U. S. A.* 106, 4495–4500.
Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* , *25*(14), 1754–1760.
Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* , *25*(16), 2078–2079.
Li, L., Stoeckert, C. J., Jr, & Roos, D. S. (2003). OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Research*, *13*(9), 2178–2189.
Li, W., Schuler, M. A., & Berenbaum, M. R. (2003). Diversification of furanocoumarin-metabolizing cytochrome P450 monooxygenases in two papilionids: Specificity and substrate encounter rate. *Proceedings of the National Academy of Sciences of the United States of America*, *100 Suppl 2*, 14593–14598.
Lieberman-Aiden, Erez, Nynke L. van Berkum, Louise Williams, Maxim Imakaev, Tobias Ragoczy, Agnes Telling, Ido Amit, et al. 2009. "Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome." *Science* 326 (5950): 289–93.
Ling, F., Dahanukar, A., Weiss, L. A., Kwon, J. Y., and Carlson, J. R. (2014). The molecular and cellular basis of taste coding in the legs of Drosophila. *J. Neurosci.* 34, 7148–7164.
Linnen, C. R., Kingsley, E. P., Jensen, J. D., & Hoekstra, H. E. (2009). On the origin and spread of an adaptive allele in deer mice. *Science*, *325*(5944), 1095–1098.
Linz, J., Baschwitz, A., Strutz, A., Dweck, H. K. M., Sachse, S., Hansson, B. S., & Stensmyr, M. C. (2013). Host plant-driven sensory specialization in Drosophila erecta. *Proceedings. Biological Sciences / The Royal Society*, *280*(1760), 20130626.
Lipke, H., and Fraenkel, G. (1956). Insect nutrition. *Annu. Rev. Entomol.* 1, 17–44.
Liu, L., Leonard, A. S., Motto, D. G., Feller, M. A., Price, M. P., Johnson, W. A., & Welsh, M. J. (2003). Contribution of Drosophila DEG/ENaC genes to salt taste. *Neuron*, *39*(1), 133–146.
Liu, Y., Ge, M., Zhang, T., Chen, L., Xing, Y., Liu, L., Li, F., & Cheng, L. (2020). Exploring the correlation between deltamethrin stress and Keap1-Nrf2-ARE pathway from Drosophila melanogaster RNASeq data. *Genomics*, *112*(2), 1300–1308.

Low, W. Y., Ng, H. L., Morton, C. J., Parker, M. W., Batterham, P., & Robin, C. (2007). Molecular evolution of glutathione S-transferases in the genus Drosophila. *Genetics*, *177*(3), 1363–1375.

Luo, Ruibang, Binghang Liu, Yinlong Xie, Zhenyu Li, Weihua Huang, Jianying Yuan, Guangzhu He, et al. 2012. "SOAPdenovo2: An Empirically Improved Memory-Efficient Short-Read de Novo Assembler." *GigaScience* 1 (1): 18.

Lynch, M., & Conery, J. S. (2000). The evolutionary fate and consequences of duplicate genes [Review of *The evolutionary fate and consequences of duplicate genes*]. *Science*, *290*(5494), 1151–1155.

Lynch, M., Walsh, B., & Others. (1998). *Genetics and analysis of quantitative traits* (Vol. 1). Sinauer Sunderland, MA.

Lynch, V. J. (2007). Inventing an arsenal: adaptive evolution and neofunctionalization of snake venom phospholipase A2 genes. *BMC Evolutionary Biology*, *7*, 2.

Mack, J. O., and Zhang, Y. V. (2021). A Rapid Food-Preference Assay in Drosophila. *J. Vis. Exp.* doi: 10.3791/62051.

Mackay, T. F. C., Richards, S., Stone, E. A., Barbadilla, A., Ayroles, J. F., Zhu, D., Casillas, S., Han, Y., Magwire, M. M., Cridland, J. M., Richardson, M. F., Anholt, R. R. H., Barrón, M., Bess, C., Blankenburg, K. P., Carbone, M. A., Castellano, D., Chaboub, L., Duncan, L., … Gibbs, R. A. (2012). The Drosophila melanogaster Genetic Reference Panel. *Nature*, *482*(7384), 173–178.

Manni, Mosè, Matthew R. Berkeley, Mathieu Seppey, Felipe A. Simão, and Evgeny M. Zdobnov. 2021. "BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes." *Molecular Biology and Evolution* 38 (10): 4647–54.

Marcellini, S., & Simpson, P. (2006). Two or four bristles: functional evolution of an enhancer of scute in Drosophilidae. *PLoS Biology*, *4*(12), e386.

Marino, P. C., & Cornell, H. V. (1992). Adult movement of the native holly leafminer, Phytomyza ilicicola Loew (Diptera: Agromyzidae): consequences for host choice within and between habitats. *Oecologia*, *92*(1), 76–82.

Marion-Poll, F., and van der Pers, J. (1996). "Un-filtered recordings from insect taste sensilla," in *Proceedings of the 9th International Symposium on Insect-Plant Relationships*, eds. E. Städler, M. Rowell-Rahier, and R. Bauer (Dordrecht: Springer Netherlands), 113–115.

Marvaldi, A. E., Sequeira, A. S., O'Brien, C. W., & Farrell, B. D. (2002). Molecular and morphological phylogenetics of weevils (coleoptera, curculionoidea): do niche shifts accompany diversification? *Systematic Biology*, *51*(5), 761–785.

Matsunaga, T., Reisenman, C. E., Goldman-Huertas, B., Brand, P., Miao, K., Suzuki, H. C., Verster, K. I., Ramírez, S. R., & Whiteman, N. K. (2022). Evolution of Olfactory Receptors Tuned to Mustard Oils in Herbivorous Drosophilidae. *Molecular Biology and Evolution*, *39*(2). https://doi.org/10.1093/molbev/msab362

Maunsell, S. C., Burwell, C. J., Morris, R. J., McDonald, W. J. F., Edwards, E. D., Oberprieler, R. G., & Kitching, R. L. (2017). Host-plants of leaf-miners in Australian subtropical rainforest. In *Austral Entomology* (Vol. 56, Issue 4, pp. 403–411). https://doi.org/10.1111/aen.12252

McBride, C. S. (2007). Rapid evolution of smell and taste receptor genes during host specialization in Drosophila sechellia. *Proc. Natl. Acad. Sci. U. S. A.* 104, 4996–5001.
McBride, C. S., Arguello, J. R., and O'Meara, B. C. (2007). Five Drosophila genomes reveal nonneutral evolution and the signature of host specialization in the chemoreceptor superfamily. *Genetics* 177, 1395–1416.
McQueen, E. W., Afkhami, M., Atallah, J., Belote, J. M., Gompel, N., Heifetz, Y., Kamimura, Y., Kornhauser, S. C., Masly, J. P., O'Grady, P., Peláez, J., Rebeiz, M., Rice, G., Sánchez-Herrero, E., Santos Nunes, M. D., Santos Rampasso, A., Schnakenberg, S. L., Siegal, M. L., Takahashi, A., … Yassin, A. (2022). A standardized nomenclature and atlas of the female terminalia of Drosophila melanogaster. *Fly*, *16*(1), 128–151.
Merritt, S. Z. (1996). Within-plant variation in concentrations of amino acids, sugar, and sinigrin in phloem sap of black mustard,Brassica nigra (L.) Koch (Cruciferae). *J. Chem. Ecol.* 22, 1133–1145.
Meunier, N., Marion-Poll, F., Rospars, J.-P., and Tanimura, T. (2003). Peripheral coding of bitter taste in Drosophila. *J. Neurobiol.* 56, 139–152.
Mitter, C., Farrell, B., & Wiegmann, B. (1988). The Phylogenetic Study of Adaptive Zones: Has Phytophagy Promoted Insect Diversification? *The American Naturalist*, *132*(1), 107–128.
Moon, S. J., Köttgen, M., Jiao, Y., Xu, H., and Montell, C. (2006). A taste receptor required for the caffeine response in vivo. *Curr. Biol.* 16, 1812–1817.
Münch, D., & Galizia, C. G. (2016). DoOR 2.0--Comprehensive Mapping of Drosophila melanogaster Odorant Responses. *Scientific Reports*, *6*, 21841.
Nagy, O., Nuez, I., Savisaar, R., Peluffo, A. E., Yassin, A., Lang, M., Stern, D. L., Matute, D. R., David, J. R., & Courtier-Orgogozo, V. (2018). Correlated Evolution of Two Copulatory Organs via a Single cis-Regulatory Nucleotide Change. *Current Biology: CB*, *28*(21), 3450–3457.e13.
Nayak, S. V., and Singh, R. N. (1985). Primary sensory projections from the labella to the brain of Drosophila melanogaster Meigen (Diptera : Drosophilidae). *Int. J. Insect Morphol. Embryol.* 14, 115–129.
Norga, K. K., Gurganus, M. C., Dilda, C. L., Yamamoto, A., Lyman, R. F., Patel, P. H., Rubin, G. M., Hoskins, R. A., Mackay, T. F., & Bellen, H. J. (2003). Quantitative analysis of bristle number in Drosophila mutants identifies genes involved in neural development. *Current Biology: CB*, *13*(16), 1388–1396.
Nylander, J. A. A., Ronquist, F., Huelsenbeck, J. P., & Nieves-Aldrey, J. L. (2004). Bayesian phylogenetic analysis of combined data. *Systematic Biology*, *53*(1), 47–67.
O'Grady, P. M., & DeSalle, R. (2018). Phylogeny of the Genus Drosophila. *Genetics*, *209*(1), 1–25.
Ohno, S. (1970). *Evolution by Gene Duplication*. https://doi.org/10.1007/978-3-642-86659-3
Okada, & Sasakawa. (1956). Leaf-mining species of Drosophilidae (Diptera). *Transactions of the Kyoto Entomological Society*, *5*(2), 25–28.

Orgogozo, V., Schweisguth, F., & Bellaïche, Y. (2001). Lineage, cell polarity and inscuteable function in the peripheral nervous system of the Drosophila embryo. *Development, 128*(5), 631–643.
Ozaki, K., Ryuda, M., Yamada, A., Utoguchi, A., Ishimoto, H., Calas, D., et al. (2011). A gustatory receptor involved in host plant recognition for oviposition of a swallowtail butterfly. *Nat. Commun.* 2, 542.
Pagel, M. (1999). Inferring the historical patterns of biological evolution. *Nature*, *401*(6756), 877–884.
Paradis, E., Claude, J., & Strimmer, K. (2004). APE: Analyses of Phylogenetics and Evolution in R language. In *Bioinformatics* (Vol. 20, Issue 2, pp. 289–290). https://doi.org/10.1093/bioinformatics/btg412
Paradis, E., Claude, J., & Strimmer, K. (2004). APE: Analyses of Phylogenetics and Evolution in R language. In *Bioinformatics* (Vol. 20, Issue 2, pp. 289–290). https://doi.org/10.1093/bioinformatics/btg412
Peláez, J. N., Gloss, A. D., Ray, J. F., Charboneau, J. L. M., Verster, K. I., & Whiteman, N. K. (2020). Evolution and genetic basis of the plant-penetrating ovipositor, a key adaptation in herbivorous Drosophilidae. In *bioRxiv*. bioRxiv. https://doi.org/10.1101/2020.05.07.083253
Poyet, M., Le Roux, V., Gibert, P., Meirland, A., Prévost, G., Eslin, P., & Chabrerie, O. (2015). The Wide Potential Trophic Niche of the Asiatic Fruit Fly Drosophila suzukii: The Key of Its Invasion Success in Temperate Europe? *PloS One*, *10*(11), e0142785.
Price, M. N., Dehal, P. S., & Arkin, A. P. (2010). FastTree 2--approximately maximum-likelihood trees for large alignments. *PloS One*, *5*(3), e9490.
Putnam, Nicholas H., Brendan L. O'Connell, Jonathan C. Stites, Brandon J. Rice, Marco Blanchette, Robert Calef, Christopher J. Troll, et al. 2016. "Chromosome-Scale Shotgun Assembly Using an in Vitro Method for Long-Range Linkage." *Genome Research* 26 (3): 342–50.
Raina, A. K., Kingan, T. G., & Mattoo, A. K. (1992). Chemical signals from host plant and sexual behavior in a moth. *Science*, *255*(5044), 592–594.
Rambaut, A., Suchard, M. A., Xie, D., & Drummond, A. J. (2014). *Tracer v1. 6. Computer program and documentation distributed by the author*. Retrieved from htp://beast. bio. ed. ac. uk/Tracer.
Rane, R. V., Clarke, D. F., Pearce, S. L., Zhang, G., Hoffmann, A. A., & Oakeshott, J. G. (2019). Detoxification Genes Differ Between Cactus-, Fruit-, and Flower-Feeding Drosophila. *The Journal of Heredity*, *110*(1), 80–91.
Rane, R. V., Walsh, T. K., Pearce, S. L., Jermiin, L. S., Gordon, K. H., Richards, S., & Oakeshott, J. G. (2016). Are feeding preferences and insecticide resistance associated with the size of detoxifying enzyme families in insect herbivores? *Current Opinion in Insect Science*, *13*, 70–76.
Revell, L. J. (2012). phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution / British Ecological Society*, *3*(2), 217–223.
Robertson, H. M. (2019). Molecular Evolution of the Major Arthropod Chemoreceptor Gene Families. *Annual Review of Entomology*, *64*, 227–242.

Ronquist, F., & Huelsenbeck, J. P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics, 19*(12), 1572–1574.
Rost, S., Fregin, A., Ivaskevicius, V., Conzelmann, E., Hörtnagel, K., Pelz, H.-J., Lappegard, K., Seifried, E., Scharrer, I., Tuddenham, E. G. D., Müller, C. R., Strom, T. M., & Oldenburg, J. (2004). Mutations in VKORC1 cause warfarin resistance and multiple coagulation factor deficiency type 2. *Nature*, *427*(6974), 537–541.
Saisawang, C., Wongsantichon, J., & Ketterman, A. J. (2012). A preliminary characterization of the cytosolic glutathione transferase proteome from Drosophila melanogaster. *Biochemical Journal*, *442*(1), 181–190.
Sang, J., Rimal, S., and Lee, Y. (2019). Gustatory receptor 28b is necessary for avoiding saponin in Drosophila melanogaster. *EMBO Rep.* 20. doi: 10.15252/embr.201847328.
Schaefer, M., Petronczki, M., Dorner, D., Forte, M., & Knoblich, J. A. (2001). Heterotrimeric G proteins direct two modes of asymmetric cell division in the Drosophila nervous system. *Cell*, *107*(2), 183–194.
Schaid, D. J. (2004). Evaluating associations of haplotypes with traits. *Genetic Epidemiology*, *27*(4), 348–364.
Schlötterer, C., Tobler, R., Kofler, R., & Nolte, V. (2014). Sequencing pools of individuals—mining genome-wide polymorphism data without big funding. *Nature Reviews. Genetics*, *15*(11), 749–763.
Schoonhoven, and Jermy (1998). Insect-plant biology: from physiology to evolution. *Insect-plant biology: from*. Available at: https://www.cabdirect.org/cabdirect/abstract/19981100796.
Schoonhoven, and Van Loon (2002). An inventory of taste in caterpillars: each species its own key. *Acta Zool. Academ. Sci. Hung.* Available at: https://library.wur.nl/WebQuery/wurpubs/fulltext/184891.
Schoonhoven, L. M., Van Loon, B., van Loon, J. J. A., & Dicke, M. (2005). *Insect-Plant Biology*. OUP Oxford.
Schoville, S. D., Chen, Y. H., Andersson, M. N., Benoit, J. B., Bhandari, A., Bowsher, J. H., Brevik, K., Cappelle, K., Chen, M.-J. M., Childers, A. K., Childers, C., Christiaens, O., Clements, J., Didion, E. M., Elpidina, E. N., Engsontia, P., Friedrich, M., García-Robles, I., Gibbs, R. A., … Richards, S. (2018). A model species for agricultural pest genomics: the genome of the Colorado potato beetle, Leptinotarsa decemlineata (Coleoptera: Chrysomelidae). *Scientific Reports*, *8*(1), 1931.
Schweisguth, F. (2015). Asymmetric cell division in the Drosophila bristle lineage: from the polarization of sensory organ precursor cells to Notch-mediated binary fate decision. In *Wiley Interdisciplinary Reviews: Developmental Biology* (Vol. 4, Issue 3, pp. 299–309). https://doi.org/10.1002/wdev.175
Scott, K. (2018). Gustatory Processing in Drosophila melanogaster. *Annu. Rev. Entomol.* 63, 15–30.
Seraj, A. A. (2000). Comparison of plant species as host for cabbage leaf miner in Khuzestan Province. *Journal of Agricultural Science and Technology: JAST*, *2*(2), 127–135.
Sham, P., Bader, J. S., Craig, I., O'Donovan, M., & Owen, M. (2002). DNA Pooling: a tool for large-scale association studies. *Nature Reviews. Genetics*, *3*(11), 862–871.

Shim, J., Lee, Y., Jeong, Y. T., Kim, Y., Lee, M. G., Montell, C., & Moon, S. J. (2015). The full repertoire of Drosophila gustatory receptors for detecting an aversive compound. *Nature Communications*, *6*, 8867.

Simão, Felipe A., Robert M. Waterhouse, Panagiotis Ioannidis, Evgenia V. Kriventseva, and Evgeny M. Zdobnov. 2015. "BUSCO: Assessing Genome Assembly and Annotation Completeness with Single-Copy Orthologs." *Bioinformatics* 31 (19): 3210–12.

Simon, J.-C., d'Alençon, E., Guy, E., Jacquin-Joly, E., Jaquiéry, J., Nouhaud, P., Peccoud, J., Sugio, A., & Streiff, R. (2015). Genomics of adaptation to host-plants in herbivorous insects. *Briefings in Functional Genomics*, *14*(6), 413–423.

Singh, S. P., Coronella, J. A., Benes, H., Cochrane, B. J., & Zimniak, P. (2001). Catalytic function of Drosophila melanogaster glutathione S-transferase DmGSTS1-1 (GST-2) in conjugation of lipid peroxidation end products. *European Journal of Biochemistry / FEBS*, *268*(10), 2912–2923.

Slone, J., Daniels, J., and Amrein, H. (2007). Sugar receptors in Drosophila. *Curr. Biol.* 17, 1809–1816.

Smadja, C., Shi, P., Butlin, R. K., & Robertson, H. M. (2009). Large gene family expansions and adaptive evolution for odorant and gustatory receptors in the pea aphid, Acyrthosiphon pisum. *Molecular Biology and Evolution*, *26*(9), 2073–2086.

Smit, A. F. A., Hubley, R., & Green, P. (2013). RepeatMasker Open-4.0. *Institute for Systems Biology*. https://www.repeatmasker.org/

Smit, AFA, Hubley, R & Green, P. 1996-2010. "RepeatMasker Open-3.0." 1996-2010. <http://www.repeatmasker.org>.

Southwood, T. R. E. (1972). insect/plant relationship--an evolutionary perspective. *Roy Entomol Soc London Symp*, *1972, 6*. http://agris.fao.org/agris-search/search.do?recordID=US201303250342

Stamatakis, A. (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics, 22*(21), 2688–2690.

Stanke, Mario, Mark Diekhans, Robert Baertsch, and David Haussler. 2008. "Using Native and Syntenically Mapped cDNA Alignments to Improve de Novo Gene Finding." *Bioinformatics* 24 (5): 637–44.

Steiner, J. E., Glaser, D., Hawilo, M. E., and Berridge, K. C. (2001). Comparative expression of hedonic impact: affective reactions to taste by human infants and other primates. *Neurosci. Biobehav. Rev.* 25, 53–74.

Stensmyr, M. C., Giordano, E., Balloi, A., Angioy, A.-M., & Hansson, B. S. (2003). Novel natural ligands for Drosophila olfactory receptor neurones. *The Journal of Experimental Biology*, *206*(Pt 4), 715–724.

Stern, D. L., & Orgogozo, V. (2008). The loci of evolution: how predictable is genetic evolution? *Evolution; International Journal of Organic Evolution*, *62*(9), 2155–2177.

Strong, D. R., Lawton, J. H., Southwood, S. R., & Others. (1984). *Insects on plants. Community patterns and mechanisms*. Blackwell Scientific Publicatons.

Sun, W., Margam, V. M., Sun, L., Buczkowski, G., Bennett, G. W., Schemerhorn, B., Muir, W. M., & Pittendrigh, B. R. (2006). Genome-wide analysis of phenobarbital-inducible genes in Drosophila melanogaster. *Insect Molecular Biology*, *15*(4), 455–464.

Sung, H. Y., Jeong, Y. T., Lim, J. Y., Kim, H., Oh, S. M., Hwang, S. W., Kwon, J. Y., & Moon, S. J. (2017). Heterogeneity in the Drosophila gustatory receptor complexes that detect aversive compounds. *Nature Communications*, *8*(1), 1484.

Suslak, T. (2015). *There and back again: A stretch receptor's tale* [PhD, The University of Edinburgh]. https://www.researchgate.net/profile/Thomas-Suslak/publication/292449835_There_and_back_again_a_stretch_receptor's_tale/links/591ecc15aca272d31bd0b4f7/There-and-back-again-a-stretch-receptors-tale.pdf

Suvorov, A., Kim, B. Y., Wang, J., Armstrong, E. E., Peede, D., D'Agostino, E. R. R., Price, D. K., Waddell, P., Lang, M., Courtier-Orgogozo, V., David, J. R., Petrov, D., Matute, D. R., Schrider, D. R., & Comeault, A. A. (2022). Widespread introgression across a phylogeny of 155 Drosophila genomes. *Current Biology: CB*, *32*(1), 111–123.e5.

Suzuki, H. C., Ozaki, K., Makino, T., Uchiyama, H., Yajima, S., & Kawata, M. (2018). Evolution of Gustatory Receptor Gene Family Provides Insights into Adaptation to Diverse Host Plants in Nymphalid Butterflies. *Genome Biology and Evolution*, *10*(6), 1351–1362.

Sved, J. A. (1971). Linkage disequilibrium and homozygosity of chromosome segments in finite populations. *Theoretical Population Biology*, *2*(2), 125–141.

Swain, T. (1977). Secondary Compounds as Protective Agents. In *Annual Review of Plant Physiology* (Vol. 28, Issue 1, pp. 479–501). https://doi.org/10.1146/annurev.pp.28.060177.002403

Swarup, S., Williams, T. I., & Anholt, R. R. H. (2011). Functional dissection of Odorant binding protein genes in Drosophila melanogaster. *Genes, Brain, and Behavior*, *10*(6), 648–657.

Swofford, D. L., & Sullivan, J. (2003). Phylogeny inference based on parsimony and other methods using PAUP*. *The Phylogenetic Handbook: A Practical Approach to DNA and Protein Phylogeny*, *7*, 160–206.

Taylor, B. J. (1989). Sexually dimorphic neurons in the terminalia of Drosophila melanogaster: I. Development of sensory neurons in the genital disc during metamorphosis. *Journal of Neurogenetics*, *5*(3), 173–192.

Thoen, M. P. M., Davila Olivas, N. H., Kloth, K. J., Coolen, S., Huang, P.-P., Aarts, M. G. M., Bac-Molenaar, J. A., Bakker, J., Bouwmeester, H. J., Broekgaarden, C., Bucher, J., Busscher-Lange, J., Cheng, X., Fradin, E. F., Jongsma, M. A., Julkowska, M. M., Keurentjes, J. J. B., Ligterink, W., Pieterse, C. M. J., … Dicke, M. (2017). Genetic architecture of plant stress resistance: multi-trait genome-wide association mapping. *The New Phytologist*, *213*(3), 1346–1362.

Thompson, J. N. (1988). Coevolution and Alternative Hypotheses on Insect/Plant Interactions. *Ecology*, *69*(4), 893–895.

Thompson, J. N. (1994). *The Coevolutionary Process*. University of Chicago Press.

Thomson, J. A., Jackson, M. J., & Bock, I. R. (1982). Contrasting resource utilisation in two Australian species of Drosophila Fallen (Diptera) feeding on the bracken fern Pteridium scopoli. *Australian Journal of Entomology*, *21*(1), 29–30.

Thurmond, J., Goodman, J. L., Strelets, V. B., Attrill, H., Gramates, L. S., Marygold, S. J., Matthews, B. B., Millburn, G., Antonazzo, G., Trovisco, V., Kaufman, T. C., Calvi,

B. R., & FlyBase Consortium. (2019). FlyBase 2.0: the next generation. *Nucleic Acids Research*, *47*(D1), D759–D765.
Trienens, M., Kraaijeveld, K., & Wertheim, B. (2017). Defensive repertoire of Drosophila larvae in response to toxic fungi. *Molecular Ecology*, *26*(19), 5043–5057.
Usui, K., Goldstone, C., Gibert, J.-M., & Simpson, P. (2008). Redundant mechanisms mediate bristle patterning on the Drosophila thorax. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(51), 20112–20117.
Van Doren, M., Bailey, A. M., Esnayra, J., Ede, K., & Posakony, J. W. (1994). Negative regulation of proneural gene activity: hairy is a direct transcriptional repressor of achaete. *Genes & Development*, *8*(22), 2729–2742.
Van Nynatten, A., Bloom, D., Chang, B. S. W., & Lovejoy, N. R. (2015). Out of the blue: adaptive visual pigment evolution accompanies Amazon invasion. *Biology Letters*, *11*(7). https://doi.org/10.1098/rsbl.2015.0349
Vaser, R., Sović, I., Nagarajan, N., & Šikić, M. (2017). Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Research*, *27*(5), 737–746.
Verschaffelt, E. (1910). The causu determing the selection of food in some herbivorous insect. *Proc. Acad. Sci. Amsterdam.* 1, 536–542.
Verster, K. I., Wisecaver, J. H., Karageorgi, M., Duncan, R. P., Gloss, A. D., Armstrong, E., Price, D. K., Menon, A. R., Ali, Z. M., & Whiteman, N. K. (2019). Horizontal transfer of bacterial cytolethal distending toxin B genes to insects. *Molecular Biology and Evolution*. https://doi.org/10.1093/molbev/msz146
Vertacnik, K. L., & Linnen, C. R. (2017). Evolutionary genetics of host shifts in herbivorous insects: insights from the age of genomics. *Annals of the New York Academy of Sciences*, *1389*(1), 186–212.
Vertacnik, K. L., Herrig, D. K., Keating Godfrey, R., Hill, T., Geib, S. M., Unckless, R. L., Nelson, D. R., & Linnen, C. R. (2021). Ecological correlates of gene family size in a pine-feeding sawfly genome and across Hymenoptera. In *bioRxiv* (p. 2021.03.14.435331). https://doi.org/10.1101/2021.03.14.435331
Vorojeikina, D., Broberg, K., Love, T. M., Davidson, P. W., van Wijngaarden, E., & Rand, M. D. (2017). Editor's Highlight: Glutathione S-Transferase Activity Moderates Methylmercury Toxicity During Development in Drosophila. *Toxicological Sciences: An Official Journal of the Society of Toxicology*, *157*(1), 211–221.
Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C. A., Zeng, Q., Wortman, J., Young, S. K., & Earl, A. M. (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PloS One*, *9*(11), e112963.
Wang, Y., Coleman-Derr, D., Chen, G., & Gu, Y. Q. (2015). OrthoVenn: a web server for genome wide comparison and annotation of orthologous clusters across multiple species. *Nucleic Acids Research*, *43*(W1), W78–W84.
Wanner, K. W., and Robertson, H. M. (2008). The gustatory receptor family in the silkworm moth Bombyx mori is characterized by a large expansion of a single lineage of putative bitter receptors. *Insect Mol. Biol.* 17, 621–629.

Watanabe, K., Toba, G., Koganezawa, M., & Yamamoto, D. (2011). Gr39a, a highly diversified gustatory receptor in Drosophila, has a role in sexual behavior. *Behavior Genetics*, *41*(5), 746–753.

Weadick, C. J., & Chang, B. S. W. (2012). Complex patterns of divergence among green-sensitive (RH2a) African cichlid opsins revealed by Clade model analyses. *BMC Evolutionary Biology*, *12*, 206.

Weiss, L. A., Dahanukar, A., Kwon, J. Y., Banerjee, D., & Carlson, J. R. (2011). The molecular and cellular basis of bitter taste in Drosophila. *Neuron*, *69*(2), 258–272.

Wells (1963). Taste by Touch: Some Experiments with Octopus. *J. Exp. Biol.*

Wentzell, A. M., and Kliebenstein, D. J. (2008). Genotype, age, tissue, and environment regulate the structural outcome of glucosinolate activation. *Plant Physiol.* 147, 415–428.

Whiteman, N. K., Groen, S. C., Chevasco, D., Bear, A., Beckwith, N., Gregory, T. R., Denoux, C., Mammarella, N., Ausubel, F. M., & Pierce, N. E. (2011). Mining the plant--herbivore interface with a leafmining Drosophila of Arabidopsis. *Molecular Ecology*, *20*(5), 995–1014.

Whiteman, Noah K., Andrew D. Gloss, Timothy B. Sackton, Simon C. Groen, Parris T. Humphrey, Richard T. Lapoint, Ida E. Sønderby, et al. 2012. “Genes Involved in the Evolution of Herbivory by a Leaf-Mining, Drosophilid Fly.” *Genome Biology and Evolution* 4 (9): 900–916.

Whitworth, A. J., Theodore, D. A., Greene, J. C., Benes, H., Wes, P. D., & Pallanck, L. J. (2005). Increased glutathione S-transferase activity rescues dopaminergic neuron loss in a Drosophila model of Parkinson’s disease. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(22), 8024–8029.

Wieczorek, H., and Wolff, G. (1989). The labellar sugar receptor ofDrosophila. *J. Comp. Physiol. A Neuroethol. Sens. Neural Behav. Physiol.* 164, 825–834.

Wiegmann, B. M., Trautwein, M. D., Winkler, I. S., Barr, N. B., Kim, J.-W., Lambkin, C., Bertone, M. A., Cassel, B. K., Bayless, K. M., Heimberg, A. M., Wheeler, B. M., Peterson, K. J., Pape, T., Sinclair, B. J., Skevington, J. H., Blagoderov, V., Caravas, J., Kutty, S. N., Schmidt-Ott, U., … Yeates, D. K. (2011). Episodic radiations in the fly tree of life. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(14), 5690–5695.

Wiens, J. J., Lapoint, R. T., & Whiteman, N. K. (2015). Herbivory increases diversification across insect clades. *Nature Communications*, *6*, 8370.

Willoughby, L., Chung, H., Lumb, C., Robin, C., Batterham, P., & Daborn, P. J. (2006). A comparison of Drosophila melanogaster detoxification gene induction responses for six insecticides, caffeine and phenobarbital. *Insect Biochemistry and Molecular Biology*, *36*(12), 934–942.

Xu, K., DiAngelo, J. R., Hughes, M. E., Hogenesch, J. B., & Sehgal, A. (2011). The circadian clock interacts with metabolic physiology to influence reproductive fitness. *Cell Metabolism*, *13*(6), 639–654.

Xu, L., Dong, Z., Fang, L., Luo, Y., Wei, Z., Guo, H., Zhang, G., Gu, Y. Q., Coleman-Derr, D., Xia, Q., & Wang, Y. (2019). OrthoVenn2: a web server for whole-genome comparison and annotation of orthologous clusters across multiple species. *Nucleic Acids Research*, *47*(W1), W52–W58.

Xu, W., Papanicolaou, A., Zhang, H.-J., & Anderson, A. (2016). Expansion of a bitter taste receptor family in a polyphagous insect herbivore. *Scientific Reports*, *6*, 23666.

Yang, C.-H., Belawat, P., Hafen, E., Jan, L. Y., & Jan, Y.-N. (2008). Drosophila egg-laying site selection as a system to study simple decision-making processes. *Science*, *319*(5870), 1679–1683.

Yang, J., Guo, H., Jiang, N.-J., Tang, R., Li, G.-C., Huang, L.-Q., et al. (2021). Identification of a gustatory receptor tuned to sinigrin in the cabbage butterfly Pieris rapae. *PLoS Genet.* 17, e1009527.

Yang, J., Weedon, M. N., Purcell, S., Lettre, G., Estrada, K., Willer, C. J., Smith, A. V., Ingelsson, E., O'Connell, J. R., Mangino, M., Mägi, R., Madden, P. A., Heath, A. C., Nyholt, D. R., Martin, N. G., Montgomery, G. W., Frayling, T. M., Hirschhorn, J. N., McCarthy, M. I., … GIANT Consortium. (2011). Genomic inflation factors under polygenic inheritance. *European Journal of Human Genetics: EJHG*, *19*(7), 807–812.

Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*, *24*(8), 1586–1591.

Yang, Z., & Nielsen, R. (2002). Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Molecular Biology and Evolution*, *19*(6), 908–917.

Yang, Z., Bengtsson, M., & Witzgall, P. (2004). Host plant volatiles synergize response to sex pheromone in codling moth, Cydia pomonella. *Journal of Chemical Ecology*, *30*(3), 619–629.

Yassin, A., Debat, V., Bastide, H., Gidaszewski, N., David, J. R., & Pool, J. E. (2016). Recurrent specialization on a toxic fruit in an island Drosophila population [Review of *Recurrent specialization on a toxic fruit in an island Drosophila population*]. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(17), 4771–4776.

Yehuda, B.-S. (2012). The Role of DEG/ENaC Subunit ppk8 in Regulating Neuronal Excitability in Drosophila melanogaster. In *Frontiers in Behavioral Neuroscience* (Vol. 6). https://doi.org/10.3389/conf.fnbeh.2012.27.00333

You, M., Yue, Z., He, W., Yang, X., Yang, G., Xie, M., Zhan, D., Baxter, S. W., Vasseur, L., Gurr, G. M., Douglas, C. J., Bai, J., Wang, P., Cui, K., Huang, S., Li, X., Zhou, Q., Wu, Z., Chen, Q., … Wang, J. (2013). A heterozygous moth genome provides insights into herbivory and detoxification. *Nature Genetics*, *45*(2), 220–225.

Zdobnov, Evgeny M., Fredrik Tegenfeldt, Dmitry Kuznetsov, Robert M. Waterhouse, Felipe A. Simão, Panagiotis Ioannidis, Mathieu Seppey, Alexis Loetscher, and Evgenia V. Kriventseva. 2017. "OrthoDB v9.1: Cataloging Evolutionary and Functional Annotations for Animal, Fungal, Plant, Archaeal, Bacterial and Viral Orthologs." *Nucleic Acids Research* 45 (D1): D744–49.

Zhang, C., Rabiee, M., Sayyari, E., & Mirarab, S. (2018). ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics*, *19*(Suppl 6), 153.

Zhang, G., Cao, S., Guo, T., Wang, H., Qi, X., Ren, X., et al. (2022). Identification and expression profiles of gustatory receptor genes in Bactrocera minax larvae

(Diptera: Tephritidae): Role of BminGR59f in larval growth. *Insect Sci.* doi: 10.1111/1744-7917.13014.
Zhang, J., Nielsen, R., & Yang, Z. (2005). Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Molecular Biology and Evolution*, *22*(12), 2472–2479.
Zhou, J.-J., Kan, Y., Antoniw, J., Pickett, J. A., & Field, L. M. (2006). Genome and EST analyses and expression of a gene family with putative functions in insect chemoreception. *Chemical Senses*, *31*(5), 453–465.

Appendices

Appendix A: Supplemental methods from Chapter 1.

Alignment partitioning and model implementation for phylogeny reconstruction.

The concatenated alignment was partitioned by codon and gene, with ribosomal genes given single partitions. The ML analysis used a GTR+gamma model on all partitions. To evaluate consistency across runs, five independent runs were performed, with distinct starting seeds. Each run included 1,000 bootstrap replicates, and a slow ML search on every 5th tree. The phylogeny with the highest likelihood was used for ancestral character estimation. For the Bayesian analysis, models of sequence evolution were selected for each partition with the Akaike information criterion (AIC), using MrModeltest2 v.2.3 (Nylander et al., 2004) and PAUP* v.4.0a (Swofford & Sullivan, 2003). To infer a phylogeny and divergence times, a Markov-Chain Monte-Carlo (MCMC) analysis was performed as previously described using the same time calibrations points and run parameters (Katoh et al., 2017). To ensure that the Markov chain adequately converged to a stationary distribution, forty replicate runs of 10 million generations each were performed and implemented in BEAST v.2.4.6 (Bouckaert et al., 2014) with BEAGLE (Ayres et al., 2012) for multicore processing. The first ten percent of samples were discarded as burn-in. Trees were re-sampled every 250,000 generations, and combined using LogCombiner v.2.4.6 (Bouckaert et al., 2014). Tracer v.1.6.0 (Rambaut et al., 2014) was used to confirm that ESS values were sufficient for reliable parameter estimates (ESS >200).

Narrow-sense heritability estimates from mother-daughter regression.

Fifty single male-female pairs (virgin females) from the combined NH1 and NH2 colonies were individually mated on *Turritis* (formerly *Arabis*) *glabra* in Magenta boxes (Sigma-Aldrich) with mesh covers. Each box was provisioned with a cotton ball soaked in 10% honey solution. For 35 mate pairs that produced daughters, ovipositor length and bristle number were profiled for every mother and at least one and up to four of her daughters (mean n = 3.4). To interrogate bristle number independently of ovipositor length, we extracted residual bristle number from a linear regression of bristle number against ovipositor length; because the interaction between ovipositor length and generation (mother or daughters) was not significant, we included both generations in a single regression model. Narrow-sense heritability (h^2) of residual ovipositor bristle number and ovipositor length – the proportion of phenotypic variation due to additive genetic effects – were each estimated by regressing the phenotype of each mother against the average phenotype of her daughters. Following convention when a trait can only be measured in parents of a single sex (Lynch et al., 1998), h^2 was defined as twice the slope of the parent-offspring regression. A one-tailed *p*-value was used to test the hypothesis that $h^2 > 0$.

Read mapping and pool-GWAS.

Reads were trimmed using Trimmomatic v0.32 with the parameters “TRAILING:3 HEADCROP:2 SLIDINGWINDOW:6:15 MINLEN:50” and mapped to an *S. flava* reference genome assembly (GenBank accession no. GCA_003952975.1) using bwa v0.7.12 (Bolger et al., 2014; H. Li & Durbin, 2009; Schlötterer et al., 2014) with the following parameters: for bwa aln, “-o 3 –d 15 -l 100”; for bwa sampe, “-a 1000”. PCR and optical duplicate reads were removed using Picard Tools v1.107 (http://picard.sourceforge.net). Unpaired and low-quality reads were removed using the View command in Samtools v1.3.1 (H. Li et al., 2009) with parameters “-q 20 –f 0x0002 –F 0x0004 –F 0x0008”. Low quality bases were removed using the Samtools mpileup command with parameters “-B –Q 17”. Repeat regions and 5 bp windows flanking indels (minimum count > 4) were filtered using Popoolation2 (Kofler et al., 2011).

Statistical significance of allele frequency differences per site was estimated using the Cochran-Mantel-Haenszel test using Popoolation2 and custom scripts for sites with a minimum minor allele count of 10, coverage depth of 100, and minor allele frequency of 8% across all pools combined, and a minimum coverage depth of 15 and maximum coverage depth 200 in each pool.

We adjusted for a minor observed inflation of *p*-values. Systematic inflation of GWAS test statistics (termed genomic inflation) – which is typically assumed to arise due to unmodeled relatedness among individuals, biased test implementation, or errors in genotyping – can result in overly confident *p*-values. Genomic inflation is also expected to arise under polygenic control of a trait, even in the absence of population structure and other technical artifacts (Yang et al., 2011). We identified and conservatively sought to correct for a slight inflation of *p*-values in our pool-GWA analysis (Devlin & Roeder, 1999). However, because the distribution of *p*-values was non-uniform with an excess of both higher and lower values, typical corrections based on the observed versus median test statistic gave unsuitable inflation factors. Following Thoen et al., we therefore regressed observed against expected $-\log_{10}(P)$ values with the intercept constrained to 0, and divided each $-\log_{10}(P)$ value by the slope of the regression line (Thoen et al., 2017). We excluded from our regression model the 1% most significant SNPs from our pool-GWAS (which are likely to be enriched for true associations) and SNPs that failed the stringent filtering described in our GO enrichment analysis (which may have overly conservative or liberal *p*-values due to biases or errors in genome assembly, sequencing and read mapping, and SNP genotyping, given the nature of our filters). Our approach for *p*-value adjustment was conservative, yielding a median *p*-value of 0.597.

Linkage disequilibrium and population structure analyses:

Sampling and pooled genome re-sequencing. Linkage disequilibrium (LD) caused by population structure, mating systems, selection and drift, among many other factors, could confound the results of our GWAS. Therefore, we explored levels of linkage disequilibrium and population structure (Fst) in several wild populations, including from different host plants within sites, across several sites, and within one laboratory population (the offspring of individuals used in the GWAS analysis). This sampling allowed us to ask whether there was any significant population structure caused by

geographic separation, host plant usage, or by our mating scheme. The populations were sampled from as follows:

(1) “Belmont, MA”: A total of 45 *S. flava* larvae were collected from *Turritis glabra* from a large field (~50,000 m^2) in Belmont, MA, USA, which contained thousands of individual *T. glabra* plants with heavy *S. flava* mining damage, between June 22 and July 2, 2013. (Linkage disequilibrium from this population is shown in Fig. 4b.)

(2) “PS-T" and “PS-B” populations: larvae were collected from *T. glabra* and *Barbarea vulgaris* from Pease Air Base in Portsmouth, NH, USA in 2014.

(3) “RM-T” and “RM-B” populations: larvae were collected from *T. glabra* and *B. vulgaris* from Rock Meadow, in Belmont, MA, USA in 2014.

(4) “FH1A1” and “FLOB1” populations: flies were generated from a mix of NH1 and NH2 flies used in the pool-GWAS. Originally collected from wild-caught larvae from Dover, NH, USA in 2014 and bred as two separate populations for two generations (see main methods and Fig. S3), the third generation of this lab colony (G3) was reared on *T. glabra* and *B. vulgaris* (G2 was used for the pool-GWAS).

Each larva was collected from a separate plant to minimize relatedness. Samples were preserved in 95% ethanol at -80C. DNA was extracted from the pool of larvae using a DNeasy Blood and Tissue Kit (Qiagen). 100 bp paired-end sequencing was conducted on half of a lane on an Illumina HiSeq 2500 at the University of Arizona in January 2014.

Read mapping. Raw sequencing reads were processed with the help of the Snakemake workflow *grenepipe* (https://github.com/moiexpositoalonsolab/grenepipe). Reads were trimmed of adapters, and trimmed with Trimmomatic v. 0.32 (Bolger et al., 2014) using the following settings: LEADING:3, TRAILING:3, HEADCROP:2, SLIDINGWINDOW:6:15, and MINLEN:50. The trimmed reads were then filtered using *samtools view* (H. Li et al., 2009) using a filtering cutoff (-q) of 60. Retained reads were then mapped to the chromosome-level genome assembly of *Scaptomyza flava* (GenBank accession no. GCA_003952975.1) using *bwa-mem* (H. Li & Durbin, 2010) with the default parameters. We hardfiltered the *S. flava* genome assembly for repeat content inferred from *RepeatMasker* prior to using it as a reference for read mapping with *bwa-mem*. We then used *Picard Tools* to remove duplicate reads using default parameters. We conducted variant calling using *GATK HaplotypeCaller* with default parameters.

Linkage Disequilibrium estimates. LD was estimated from the 15 largest autosomal scaffolds. We used Tabix version 0.2.6 to split the variant calling file (vcf) file by 15 scaffolds. We then used the LDx.pl script from the LDx program (Feder et al., 2012) to calculate genome-wide linkage disequilibrium for each scaffold separately. We ran LDx with the following parameter values, which remained the same for all populations: maximum read depth of 35, minimum read depth of 15, and minimum intersection depth of three. The parameter for the distance between the start of the first read to the end of

the second read was specified for each population based on the paired-read insert sizes as follows: AV-B = 283 bp, AV-T = 287 bp, FH1A1 = 338 bp, FLOB1 = 306 bp, PS-B = 296 bp, PS-T = 298 bp, RM-B = 281 bp, RMT = 267 bp, and Belmont, MA = 422 bp.

Fst estimates. We used the program grenedalf (https://github.com/lczech/grenedalf) to calculate genome-wide pairwise Fst for all given population pairs. Many commands in grenedalf are re-implementations of commands of the PoPoolation and PoPoolation2 tools (Kofler et al., 2011). To do this, we used the merged mpileup file which was the output from the grenepipe program (above). We specified custom pool sizes for each population. We then calculated Fst using the Spence-nei statistics for 10,000 bp windows across 15 scaffolds representing the autosomes (as above).

Gene ontology enrichment analysis on candidate SNPs:

To test whether genes intersecting the top GWAS associations were enriched for particular predicted functions, we assigned functional annotations using orthology relationships among protein-coding genes in *S. flava* and other *Drosophila* genomes. Orthology was inferred by similarity clustering using orthoMCL v2.0.9 (L. Li et al., 2003), with default parameters and an inflation value of 1.5, among proteomes for *S. flava,* all *Drosophila* species from the Drosophila 12 Genomes Consortium (Drosophila 12 Genomes Consortium et al., 2007) (retrieved from FlyBase release 2013_06) except *D. willistoni*, and a draft genome assembly of *S. pallida* (unpublished data). Each *S. flava* gene was then annotated with the gene ontology (GO) terms assigned to its predicted ortholog(s) in *D. melanogaster* in FlyBase (release 2020_02). Parental GO terms that were implied but not directly listed, which were necessary for downstream analyses, were retrieved using GO.db v3.7.0 (Carlson et al., 2017; Gentleman et al., 2004).

We used GOWINDA (Kofler & Schlötterer, 2012) to test for enrichments of gene ontology (GO) terms among the genes that intersected SNPs with the strongest pool-GWAS associations (top 0.1% and 0.005% of *p*-values). We conservatively assumed that SNPs within the same gene were in LD and thus were not independent associations (–mode gene), and significance was determined from one million permutations. All gene models were extended by 200 bp (–gene-definition updownstream200) to account for the fact that genotyped SNPs may tag non-genotyped causal variants that are proximal, but the rapid decay of LD in *S. flava* (Fig. 4b) makes this unlikely over long physical distances. To capture both protein-coding and regulatory effects, SNPs were assigned to a given gene if they fell within its exons, introns, or the adjacent upstream or downstream intergenic region. Intergenic regions were extended from a focal gene's UTR boundary until reaching the boundary of the adjacent gene's UTR, up to a maximum of 2 kb. Intergenic regions were included because most cis-regulatory elements in *D. melanogaster* are located within or adjacent to the genes they regulate, but are rarely separated by an intervening gene (Kvon et al., 2014). To avoid diluting statistical power by the inclusion of redundant, nested GO terms or terms with few member genes, terms were only considered if they were assigned to at least 20 genes in *S. flava,* and we focused on only a single level of the GO hierarchy (where level refers to the number of

edges from the focal GO term to the root of the acyclic graph of GO term relationships). Level two was used for Molecular Function, which maximized the number of terms considered, and level four was used for Biological Process because it contained a similar number of terms.

Prior to the GO enrichment analysis, we performed a more stringent SNP filtering step, excluding tri-allelic sites (having a frequency > 0.08 for the third most common allele) and SNPs located within 300 bp of a scaffold edge.

Reproducing pool-GWAS association for a candidate SNP upstream of *Gai*:

DNA was extracted from whole adult flies with the Qiagen DNeasy blood and tissue kit, following the provided protocol. PCR primers for the *Gαi* region were designed in Geneious, using default settings and a target region size of 500 bp. The primer sequences are as follows: gai4F: CATTTCTGTCCATGGCGTCG; gai9R: GCCGTTAGACAAAGCGCATT. PCR methods were the same as those previously described (Gloss et al., 2014). PCR clean up and sequencing were performed at the UC Berkeley DNA Sequencing Facility.

For trimming, alignment, and base calling, we used Geneious v.10.0.5 (Biomatters Ltd.) Using the Trimming Tool in Geneious, regions with more than a 5% chance of an error per base were trimmed. Sequences were aligned with the default "Geneious Alignment" setting (Cost matrix: 65% similarity, Gap open penalty: 12, Gap extension penalty: 3, Refinement Iterations: 2, Alignment type: Global alignment with free end gaps). The final alignment was 331 base pairs long. A significant fraction of sequences had convoluted regions upstream of the SNP, likely due to heterozygous indels. Manual base-calling was therefore performed on all sequences. Double peaks were considered heterozygotes. Sequences with convoluted regions were analyzed with Indelligent v.1.2, which identified insertion-deletion polymorphisms (indels) (Dmitriev & Rakitov, 2008). In these cases, most of the reverse strand sequence was rendered unreadable, so all sequences (convoluted or not) were based on only forward strand nucleotide base calls. Variant sites, including the candidate SNP, were identified in Geneious, using a minimum variant frequency cutoff of 0.05.

The regression model tested for significant effects of the major allele, collection locality, lab-rearing host plant species, ovipositor length, and whether they were included in the high or low bristle number pools (to account for polygenic effects of the genomic background). Effect size (β) was estimated as the shift in bristle number in standard deviations expected from a single allelic substitution. A second model tested the effects of these factors on ovipositor length.

Pairwise measures of LD between the SNP and other variant sites (minimum frequency >0.05) were calculated using the *LD2* function from the pegas package in R (Paradis et al., 2004). Correlation among alleles is given by δ (Schaid, 2004) with strong LD indicated by $|\delta| > 0.5$ (p-value < 0.01).

We performed a test for Hardy-Weinberg equilibrium (HWE) for the focal SNP among all individuals using the *hw.test* function from the pegas package. The SNP was found to be in HWE ($\chi 2 = 3.15$, $P > 0.05$).

Appendix B: Supplemental tables from Chapter 1.

Table S1. Species included in this study and GenBank accession numbers for sequences included. Sequences obtained in this study are in underlined bold font.

Species	*16S*	*COI*	*COII*	*ND2*	*28S*	*Adh*	*Cad-r*	*Gpdh*	*Gstd1*	*Marf*	*N(l)tid*
S. abrupta	KC609618	KC609723	KC609678	–	–	–	KC609590	–	–	KC609521	–
S. akalae	HQ171045	HQ170854	HQ170736	–	–	–	–	–	–	KC609487	KC609541
S. albovittata	–	KC609725	KC609680	–	–	M80925	KC609592	–	–	KC609523	–
S. ampliloba	KC609616	KC609721	KC609676	KC609640	–	–	KC609588	–	–	KC609519	KC609557
S. anomala	–	–	HQ170748	–	–	AB033646–48	–	–	–	–	–
S. apicata	KC609623	JX160024	JX160028	KC609646	–	–	JX160020	–	KC609534	JX160038	KC609561
S. apiciguttula	KC609601	KC609693	KC609650	KC609631	–	–	–	–	–	KC609492	–
S. baechlii	LC061479	LC061490	LC061501	LC061512	LC061523	LC061639	–	LC061649	–	LC061888	LC061877
S. boninensis	LC061480	LC061491	LC061502	LC061513	LC061524	LC061640	LC061867	LC061650	–	LC061889	LC061878
S. bryani	KC609620	KC609726	KC609681	–	–	–	–	–	KC609531	KC609524	KC609558
S. buccata	KC609600	KC609690	–	KC609630	–	–	–	–	–	KC609488	–
S. caliginosa	–	–	EU493805	–	–	–	KC609589	–	KC609530	KC609520	–
S. chauliodon	–	–	KC609684	–	–	–	–	–	–	–	–
S. connata	–	KC609701	KC609656	–	–	–	–	–	–	KC609497	KC609545
S. consimilis	LC061481	LC061492	LC061503	LC061514	LC061525	LC061641	LC061868	LC061651	–	LC061890	LC061879
S. crassifemur	KC609614	–	EU493806	EU493547	–	M60790	–	–	–	KC609517	KC609556
S. ctenophora	–	KC609715	KC609670	–	–	–	KC609581	–	–	KC609511	–
S. cuspidata	KC609610	KC609714	KC609669	–	–	–	KC609580	–	–	KC609510	–
S. cyrtandrae	HQ171046	KC609694	HQ170737	HQ170936	–	–	–	–	KC609529	KC609493	KC609542
S. decepta	–	KC609712	KC609667	KC609638	–	–	KC609578	–	–	KC609508	KC609551
S. dentata	KC609607	KC609703	KC609658	–	–	–	–	–	–	KC609499	–
S. devexa	–	KC609716	KC609671	–	–	–	KC609582	–	–	KC609512	–
S. elmoi(HW)	–	–	HQ170735	–	–	–	–	–	–	–	–
S. elmoi(JP)	LC061482	LC061493	LC061504	LC061515	LC061526	LC061642	LC061869	LC061652	–	–	LC061880
S. exigua	KC609603	KC609697	KC609652	KC609634	–	–	–	–	–	KC609495	KC609543
S. fastigata	–	KC609706	KC609661	–	–	–	KC609573	–	–	KC609502	–
S. flava(NA)	KC609621	JX160022		KC609644	–	–	JX160018	–	KC609532	JX160036	KC609559
S. flava(NZ)	KC609599	HQ170855	HQ170738	–	–	–	–	–	KC609527	KC609486	–
S. frustulifera	–	–	–	EU493549	–	–	–	–	–	–	–
S. graminum	LC061483	LC061494	LC061505	LC061516	LC061527	LC061643	LC061870	LC061653	–	LC061891	LC061881
S. hackmani	–	KC609708	KC609663	–	–	–	KC609575	–	–	KC609504	–

S. hamata	–	–	KC609685	–	–	–	–	–	–	–	–
S. hennigi	LC061484	LC061495	LC061506	LC061517	LC061528	–	–	LC061654	LC061875	LC061892	LC061882
S. hexasticha	LC061485	LC061496	LC061507	LC061518	LC061529	LC061644	LC061871	LC061655	–	–	LC061883
S. himalayana	LC061486	LC061497	LC061508	LC061519	LC061530	LC061645	LC061872	LC061656	–	LC061893	LC061884
S. hsui	KC609626	KC609729	KC609687	–	–	–	KC609594	–	–	KC609480	KC609565
S. intricata	–	KC609709	KC609664	KC609637	–	–	KC609576	–	–	KC609505	KC609548
S. latitergum	–	KC609710	KC609665	–	–	–	–	–	–	KC609506	–
S. lobifera	KC609604	–	KC609654	–	–	–	–	–	–	–	–
S. longisetosa	–	HQ170856	HQ170739	HQ170938	–	–	KC609567	–	–	KC609482	KC609538
S. magnipalpa	HQ171048	HQ170857	HQ170740	HQ170947	–	–	KC609566	–	–	KC609481	KC609537
S. montana	–	**MH938266**	–	–	–	–	–	–	–	–	–
S. nasalis	KC609617	KC609722	KC609677	KC609641	–	–	–	–	–	–	–
S. neocyrtandrae	HQ171049	KC609689	HQ170741	HQ170939	–	–	–	–	–	–	KC609540
S. neosilvicola	–	HQ170858	–	KC609629	–	–	KC609569	–	–	KC609485	–
S. nigrita	KC609624	JX160025	JX160029	KC609647	–	–	JX160021	–	KC609535	JX160039	KC609562
S. nr. nigrita	–	**MH938268**	**MH938269**	–	–	–	**MH938267**	–	–	**MH938270**	–
S. okadai	LC061487	LC061498	LC061509	LC061520	LC061531	LC061646	LC061873	LC061657	–	LC061894	LC061885
S. palata	–	–	KC609686	–	–	–	–	–	–	–	–
S. pallida(NA)	KC609622	JX160023	HQ110571	KC609645	–	–	JX160019	–	KC609533	JX160037	KC609560
S. pallida(JP)	LC061488	LC061499	LC061510	LC061521	LC061532	LC061647	LC061874	LC061658	LC061876	–	LC061886
S. pallifrons	–	KC609713	KC609668	–	–	–	KC609579	–	–	KC609509	KC609552
S. palmae	–	–	EU493809	EU493550	–	AB033649–51	KC609571	–	KC609528	KC609490	–
S. polygonia	LC061489	LC061500	LC061511	LC061522	LC061533	LC061648	–	LC061659	–	LC061895	LC061887
S. protensa	KC609606	KC609702	KC609657	–	–	–	–	–	–	–	–
S. pusilla	KC609605	KC609699	KC609655	–	–	–	–	–	–	KC609496	–
S. recava	–	KC609696	–	KC609633	–	–	–	–	–	–	–
S. recta	KC609611	KC609717	KC609672	KC609639	–	–	KC609583	–	–	KC609513	KC609553
S. reducta	KC609613	KC609719	KC609674	–	–	–	KC609586	–	–	KC609516	KC609555
S. scoloplichas	–	KC609698	KC609653	–	–	–	–	–	–	–	–
S. setosiscutellum	–	KC609727	KC609682	KC609643	–	–	KC609593	–	–	KC609525	–
S. silvicola	–	–	–	HQ170940	–	–	–	–	–	–	–
***S. sp.* (CA)**	–	**MH938263**	**MH938264**	–	–	–	**MH938262**	–	–	**MH938265**	–
S. spnr anomala	–	KC609692	KC609649	–	–	–	KC609572	–	–	KC609491	–
S. spnr cuspidata	–	KC609718	KC609673	–	–	–	KC609584	–	–	KC609514	–
S. spnr decepta	–	KC609707	KC609662	–	–	–	KC609574	–	–	KC609503	–
S. spnr inflatus	KC609615	KC609720	KC609675	–	–	–	KC609587	–	–	KC609518	–
S. spnr longipecten	–	KC609711	KC609666	–	–	–	KC609577	–	–	KC609507	KC609550

S. spnr longisetosa	–	KC609691	KC609648	–	–	–	KC609570	–	–	KC609489	–
S. spnr mitchelli	KC609619	KC609724	KC609679	KC609642	–	–	KC609591	–	–	KC609522	–
S. trivittata	–	KC609704	KC609659	KC609635	–	–	–	–	–	KC609500	KC609546
S. tumidula	KC609619	KC609688	HQ170743	HQ170941	–	–	–	–	–	KC609483	–
S. umbrosa	EU494407	KC609695	KC609651	KC609632	–	–	–	–	–	KC609494	–
S. undulata	EU494407	–	EU493810	–	–	–	KC609585	–	–	KC609515	KC609554
S. vagabunda	KC609608	KC609705	KC609660	KC609636	–	–	–	–	–	KC609501	KC609547
S. varia	–	KC609700	–	–	–	–	–	–	–	–	–
S. varifrons	KC609598	HQ170859	HQ170744	KC609627	–	–	–	–	–	–	–
S. varipicta	HQ171051	HQ170860	HQ170745	KC609628	–	–	KC609568	–	KC609526	KC609484	KC609539
I. ancyla	HQ170952	HQ170749	HQ170632	HQ170861	–	–	–	–	–	–	–
I. dissita	HQ170964	HQ170763	HQ170649	HQ170876	–	–	–	–	–	–	–
I. grimshawi	BK006341	BK006341	BK006341	BK006341	GU597390	XM_001993474	XM_001991642	XM_001988117	XM_001993564	XM_001992688	XM_001986304
I. malele	HQ170978	HQ170776	HQ170668	HQ170891	–	–	–	–	–	–	–
I. melanocephala	HQ170980	HQ170778	HQ170670	–	–	–	–	–	–	–	–
I. melanoloma	EU494391	–	EU493791	EU493536	–	–	–	–	–	–	–
I. mimica	HQ170982	HQ170780	HQ170672	HQ170950	–	M60792	–	–	–	–	–
I. mitchelli	HQ170983	HQ170781	HQ170673	HQ170894	–	–	–	–	–	–	–
I. nigra	EU494394	–	EU493796	EU493540	–	M60793	–	–	–	–	–
I. ochropleura	KC609625	KC609728	KC609683	–	–	–	–	–	KC609536	–	KC609564
I. percnosoma	HQ171022	HQ170819	HQ170715	HQ170929	–	–	–	–	–	–	–
I. tanythrix	–	–	HQ170726	–	–	–	–	–	–	–	KC609563
D. hydei	–	JQ679112	EU493736	DQ471529	HQ110530	X58694	–	L41650	–	–	–
D. mojavensis	BK006339	BK006339	BK006339	BK006339	CH934023	XM_002002893	XM_002009405	XM_002002668	XM_002000706	XM_002009489	XM_002004145
D. virilis	BK006340	BK006340	BK006340	BK006340	HQ110543	XM_002057548	XM_002055577	XM_002051332	XM_002054301	XM_002057373	XM_002059275

Table S2. Ovipositor bristle number, length and larval feeding ecology data for 95 species included in phylogenetic analyses to estimate the ancestral character states and perform phylogenetic generalized least squares. Larval diet data were from Lapoint et al. 2013.

Species	Larval Diet	Source for Bristle Counts	Source Type	Ovipositor Bristle Counts	Ovipositor Length (µm)
D. ancyla	saprophagy				
D. dissita	saprophagy	Throckmorton 1966	illustration	18	
D. grimshawi	saprophagy	Craddock et al. 2018	SEM	36,42, 43	690, 616, 816
D. hydei	saprophagy	Baechli et al. 2004	illustration	23	
D. malele	unknown				
D. melanocephala	saprophagy	Craddock et al. 2018	SEM	39	789
		Hardy 1967	illustration	31	
D. melanoloma	fungus feeding	Throckmorton 1966	illustration	23	
D. mimica	saprophagy	Craddock et al. 2018	SEM	30, 30	393, 394
D. mitchelli	unknown				
D. mojavensis	saprophagy				
D. nigra	unknown	Hardy et al. 2001	illustration	28	
		Throckmorton 1966	illustration	27	
		Craddock et al. 2018	SEM	33, 33	509, 537
D. ochropleura	fungus feeding	Hardy et al. 2001	illustration	21	
		Craddock et al. 2018	SEM	19, 23, 24	368, 310, 328
D. percnosoma	saprophagy				
D. tanythrix	saprophagy	Craddock et al. 2018	SEM	28	663
D. virilis	saprophagy	Sturtevant 1921	illustration	20	
S. abrupta	unknown				
S. akalae	unknown				
S. albovittata	unknown	Craddock & Kambysellis 1997	SEM	12	
		Craddock et al. 2018	SEM	10	90
S. ampliloba	unknown				
S. anomala	unknown				
S. nr. anomala	unknown				
S. apicata	saprophagy	Mount Hood National Forest, Camp Creek	light microscopy	11	261.4
		Mount Hood National Forest, Trout Creek	light microscopy	13, 13	241.8, 243.9
		Lab colony reared from wild caught individuals from Strawberry Creek, Berkeley, California	light microscopy	12, 14, 15	223.4, 224.9, 243.9
S. apiciguttula	saprophagy				
S. baechlii	unknown	Sidorenko 1993	illustration	15	
S. boninensis	unknown				
S. bryani	spider predation				
S. buccata	unknown				
S. caliginosa	flower specialist	Craddock et al. 2018	SEM	16	209
S. chauliodon	spider predation				
S. connata	unknown				
S. consimilis	saprophagy	Baechli et al. 2004	illustration	45	351.7
S. crassifemur	unknown	Throckmorton 1966	illustration	20	
		Craddock et al. 2018	SEM	18	238
S. ctenophora	unknown				
S. cuspidata	saprophagy				

S. nr. cuspidata	saprophagy				
S. cyrtandrae	saprophagy				
S. decepta	unknown				
S. nr. decepta	unknown				
S. dentata	unknown				
S. devexa	unknown				
S. elmoi (Hawaii)	saprophagy				
S. elmoi (Japan)	saprophagy				
S. exigua	saprophagy				
S. fastigata	unknown				
S. flava (North America)	leaf-mining	Lab colony, reared from wild caught individuals from New Hampshire	light microscopy	34, 35, 38, 38, 40, 40, 40, 40, 41, 41, 41, 44	269.9, 303.7, 278.2, 292.6, 292.1, 296.6, 296.2, 281.4, 332.4, 312.3, 313.3, 308
S. flava (New Zealand)	leaf-mining				
S. montana (California)	leaf-mining	Strawberry Creek, Berkeley, California	light microscopy	37, 39	306.3, 285.1
		Lab colony, reared from wild caught individuals from Strawberry Creek, Berkeley, CA	light microscopy	27, 36, 37, 37, 38, 40, 40, 42	NA, 258.1, 310.1, 306.3, 271.9, 309.8, 295.2, 283.6
S. montana (Arizona)	leaf-mining	Lab colony, reared from wild caught individuals from Arizona	light microscopy	37, 37, 37, 39, 40, 41, 42, 42, 44, 46	306.5, 288.1, 298.9, 290.5, 276.4, 278.3, 287.6, 300.4, 302.6
S. sp.	leaf-mining	Santa Cruz, California	light microscopy	39, 43	358.7, 309.8
S. frustulifera	unknown				
S. graminum	leaf-mining	Baechli et al. 2004	illustration	39	
S. hackmani	saprophagy				
S. hamata	unknown				
S. hennigi	unknown				
S. hexasticha	unknown	Okada 1973	illustration	14	173.1
S. himalayana	unknown				
S. hsui	saprophagy	Thomas Creek, Nevada	light microscopy	11, 11, 12, 12, 12, 12, 13	250.4, 247.3, 234, 244, 243, 236.8, 241
		Mount Hood National Forest, Still Creek, Oregon	light microscopy	12	158.2
		Rocky Mountain Biological Lab, near Gothic, Colorado	light microscopy	12, 15	248.3, 254
S. nr. inflatus	unknown				
S. intricata	saprophagy				
S. latitergum	saprophagy	Throckmorton 1966	illustration	25	
S. lobifera	unknown				
S. nr. longipecten	saprophagy				
S. longisetosa	unknown				
S. nr. longisetosa	unknown				
S. magnipalpa	spider predation				
S. nr. mitchelli	unknown				
S. nasalis	unknown				
S. neocyrtandrae	saprophagy				

S. neosilvicola	spider predation				
S. nigrita (Colorado)	leaf-mining	Rocky Mountain Biological Lab, near Gothic, Colorado	light microscopy	37, 38, 39, 41, 41, 42, 42	394.6, 396, 422.5, 421, 434.2, 395.7, 309.9
S. nr. nigrita (California)	leaf-mining				
S. nr. nigrita (Nevada)	leaf-mining	Lab colony, reared from wild caught individuals from Lake Tahoe, Nevada	light microscopy	38, 38, 44	286.1, 288.5, 306
S. okadai	unknown	Hackman 1959	illustration	15	
S. palata	unknown				
S. pallida (Japan)	saprophagy				
S. pallida (North American)	saprophagy	Mount Hood National Forest, Still Creek, Oregon	light microscopy	8, 8, 9, 10, 11, 12, 12	197, 184.2, 193, 208.9, 181.6, 187, 213.8
		Mount Hood National Forest, Camp Creek, Oregon	light microscopy	13	195.2
		Baechli et al. 2004	illustration	17	208.9
S. pallifrons	unknown				
S. palmae	flower specialist				
S. polygonia	saprophagy				
S. protensa	unknown				
S. pusilla	unknown				
S. recava	unknown				
S. recta	unknown				
S. reducta	unknown				
S. scoloplichas	saprophagy				
S. setosiscutellum	spider predation	Hardy 1965	illustration	8	294.6
S. silvicola	unknown				
S. trivittata	unknown				
S. tumidula	saprophagy				
S. umbrosa	unknown				
S. undulata	unknown				
S. vagabunda	unknown				
S. varia	saprophagy				
S. varipicta	saprophagy				
S. varifrons	unknown				

Table S3. Comparison of models of evolution for ovipositor bristle number.

Model	log likelihood	AIC	AICc	dAICc	AICc Weights	Parameter Estimates
Brownian Motion	-104.694	213.388	213.849	0	0.622	σ2 = 4.82
Ornstein-Uhlenbeck	-104.586	215.172	216.131	2.282	0.199	α = 0.014, σ2 = 6.066
Early Burst	-104.694	215.388	216.348	2.499	0.178	α = -0.000001, σ2 = 4.82
White Noise	-111.734	227.468	227.93	14.081	0.001	σ2 = 130.051

Table S4. Comparison of models of evolution for larval diet.

Model	log likelihood	AIC	AICc	dAICc	AICc Weights
Equal rates	-23.39	48.78	48.864	0	0.999
Symmetric	-22.645	65.29	70.93	22.066	1.62E-05
All rates different	-22.389	84.779	113.744	64.88	8.16E-15

Table S5. Full dataset used to evaluate whether SNP effect sizes identified from the pool-GWAS on ovipositor bristle number in *S. flava* could be replicated through individual sequencing. SNP214 represents the candidate SNP identified from the pool-GWAS. Other SNPs and indels that were present at a minimum variant frequency of 0.05 within the sequenced region are also presented. Ovipositor length is given in μm. Raw sequence data is available on GenBank (accession no. MH884655-MH884734).

ID	Population / Host plant	Population	Bristle number pool	Bristle number	Ovipositor length	SNP 214	SNP 214 additive	SNP 49	Indel 213	SNP 215	SNP 235	SNP 236
P4RAC11	NH1_TurrBarb	NH1	low	35	0.252	AA	0	AA	TT	TT	TT	CC
P4RAC8	NH1_TurrBarb	NH1	high	42	0.254	CA	1	AA	T-	GG	NA	NA
P4RBC1	NH1_TurrBarb	NH1	high	44	0.241	CC	2	TT	--	GG	CC	TT
P4RBC2	NH1_TurrBarb	NH1	low	34	0.26	AA	0	AA	TT	TT	TT	CC
P4RBC4	NH1_TurrBarb	NH1	high	45	0.234	CC	2	TT	--	GG	CC	TT
P4RBC5	NH1_TurrBarb	NH1	low	35	0.237	AA	0	AA	TT	TT	TT	CC
P4RBC6	NH1_TurrBarb	NH1	high	46	0.246	CA	1	AA	T-	GG	NA	NA
P4RBC7	NH1_TurrBarb	NH1	high	41	0.233	CC	2	AA	--	TT	CC	TT
P4RCC10	NH1_TurrBarb	NH1	high	44	0.268	CC	2	TT	--	GG	CC	TT
P4RCC11	NH1_TurrBarb	NH1	high	42	0.25	CA	1	AA	T-	TT	TT	CC
P4RCC12	NH1_TurrBarb	NH1	high	44	0.251	CC	2	TT	--	GG	CC	TT
P4RCC3	NH1_TurrBarb	NH1	high	44	0.269	AA	0	AA	TT	TT	TT	CC
P4RCC9	NH1_TurrBarb	NH1	high	42	0.232	CA	1	AA	TT	GT	CT	CT
P4RDC1	NH1_TurrBarb	NH1	low	36	0.25	CC	2	AA	TT	GG	CC	TT
P4REC3	NH1_TurrBarb	NH1	low	34	0.236	AA	0	AA	TT	TT	TT	CC
P4REC6	NH1_TurrBarb	NH1	high	43	0.264	CC	2	TT	--	GG	CC	TT
P4REC7	NH1_TurrBarb	NH1	low	34	0.246	AA	0	AA	TT	TT	TT	CC
P4REC8	NH1_TurrBarb	NH1	low	34	0.25	CC	2	AA	--	GG	CC	TT
P4RFC1	NH1_TurrBarb	NH1	low	34	0.228	CA	1	AA	T-	TT	TT	CC
P4RFC11	NH1_TurrBarb	NH1	low	37	0.263	CC	2	TT	--	GG	CC	TT
P4RFC3	NH1_TurrBarb	NH1	low	37	0.26	CA	1	AA	T-	TT	TT	CC
P4RGC2	NH1_TurrBarb	NH1	high	44	0.27	CC	2	TT	--	GG	CC	TT
P4RGC4	NH1_TurrBarb	NH1	high	42	0.243	AA	0	AA	TT	TT	TT	CC
P4RGC5	NH1_TurrBarb	NH1	high	43	0.236	CC	2	TT	--	GG	CC	TT

P4RGC8	NH1_TurrBarb	NH1	low	34	0.247	AA	0	AA	TT	TT	TT	CC
P4RGC9	NH1_TurrBarb	NH1	low	34	0.245	CA	1	AA	T-	TT	TT	CC
P4RHC1	NH1_TurrBarb	NH1	low	36	0.268	CA	1	AA	T-	GG	NA	NA
P4RHC11	NH1_TurrBarb	NH1	high	46	0.259	CC	2	TT	--	GG	CC	TT
P4RHC3	NH1_TurrBarb	NH1	high	44	0.255	CA	1	AA	T-	TT	TT	CC
P4RHC4	NH1_TurrBarb	NH1	low	33	0.253	AA	0	AA	TT	TT	TT	CC
P4RHC6	NH1_TurrBarb	NH1	high	45	0.254	AA	0	AA	TT	TT	TT	CC
P5RAC4	NH1_TurrBarb	NH1	low	37	0.268	CC	2	TT	--	GG	CC	TT
P5RAC6	NH1_TurrBarb	NH1	low	35	0.244	AA	0	AA	TT	TT	TT	CC
P5RBC1	NH1_TurrBarb	NH1	low	33	0.244	AA	0	AA	TT	TT	TT	CC
P5RBC10	NH1_TurrBarb	NH1	low	37	0.228	CC	2	AA	--	GG	CC	TT
P5RBC2	NH1_TurrBarb	NH1	high	46	0.24	CC	2	TT	--	GG	CC	TT
P5RBC3	NH1_TurrBarb	NH1	low	33	0.251	CA	1	AA	T-	GG	NA	NA
P6RAC10	NH2_Turr	NH2	high	45	0.226	CA	1	AA	T-	TT	NA	NA
P6RAC11	NH2_Turr	NH2	low	35	0.236	CA	1	AA	T-	TT	NA	NA
P6RAC4	NH2_Turr	NH2	low	38	0.259	CC	2	AA	TT	GG	CC	TT
P6RAC5	NH2_Turr	NH2	low	39	0.269	CA	1	AA	T-	GG	NA	NA
P6RBC11	NH2_Turr	NH2	low	37	0.254	CC	2	TT	--	GG	CC	TT
P6RBC2	NH2_Turr	NH2	high	46	0.267	CC	2	AA	--	GG	CC	TT
P6RBC3	NH2_Turr	NH2	high	43	0.245	CC	2	AA	TT	GG	CC	TT
P6RBC4	NH2_Turr	NH2	high	45	0.259	CC	2	TT	--	GG	CC	TT
P6RBC7	NH2_Turr	NH2	low	33	0.242	CA	1	AA	TT	GT	CT	CC
P6RCC1	NH2_Turr	NH2	low	37	0.252	CC	2	AA	--	GG	CC	TT
P6RDC1	NH2_Turr	NH2	low	37	0.264	CA	1	AA	T-	TT	NA	NA
P6RDC11	NH2_Turr	NH2	high	41	0.247	AA	0	AA	TT	TT	TT	CC
P6RDC4	NH2_Turr	NH2	high	43	0.25	CC	2	TT	--	GG	CC	TT
P6RDC5	NH2_Turr	NH2	low	36	0.255	CA	1	AA	T-	TT	NA	NA
P6REC4	NH2_Turr	NH2	high	44	0.254	CC	2	AA	--	GG	CC	TT
P6REC5	NH2_Turr	NH2	high	42	0.252	AA	0	AA	TT	TT	TT	CC
P6REC8	NH2_Turr	NH2	low	37	0.264	AA	0	AA	TT	TT	TT	CC
P7RAC1	NH2_Barb	NH2	high	43	0.265	CC	2	TT	--	GG	CC	TT
P7RAC11	NH2_Barb	NH2	high	45	0.256	CA	1	AA	T-	GG	NA	NA

P7RAC5	NH2_Barb	NH2	high	43	0.266	CC	2	AA	--	GG	CC	TT
P7RAC6	NH2_Barb	NH2	low	35	0.255	AA	0	AA	TT	TT	TT	CC
P7RAC8	NH2_Barb	NH2	high	43	0.26	CC	2	AA	--	GT	CC	TT
P7RBC1	NH2_Barb	NH2	low	34	0.238	CA	1	AA	T-	TT	TT	CC
P7RBC10	NH2_Barb	NH2	high	41	0.236	CC	2	AA	--	TT	CC	TT
P7RBC2	NH2_Barb	NH2	low	36	0.259	CA	1	AA	T-	GG	NA	NA
P7RBC3	NH2_Barb	NH2	high	41	0.246	CC	2	TT	--	GG	CC	TT
P7RBC6	NH2_Barb	NH2	high	44	0.265	CA	1	AA	TT	GT	TT	CC
P7RCC6	NH2_Barb	NH2	low	35	0.264	CC	2	AA	--	TT	CC	TT
P7RCC8	NH2_Barb	NH2	low	33	0.255	CC	2	TT	--	GG	CC	TT
P7RDC3	NH2_Barb	NH2	high	45	0.268	CC	2	AA	T-	GG	NA	NA
P7RDC5	NH2_Barb	NH2	low	36	0.254	CC	2	AA	T-	GG	NA	NA
P7REC1	NH2_Barb	NH2	high	43	0.251	CC	2	AA	TT	GG	CC	TT
P7REC10	NH2_Barb	NH2	low	34	0.266	AA	0	AA	TT	TT	TT	CC
P7REC5	NH2_Barb	NH2	high	44	0.262	CA	1	AA	T-	TT	TT	CC
P7REC6	NH2_Barb	NH2	high	44	0.268	CC	2	AA	TT	GG	CC	TT
P7REC7	NH2_Barb	NH2	low	34	0.263	AA	0	AA	TT	TT	TT	CC
P7RGC3	NH2_Barb	NH2	high	44	0.276	CA	1	AA	TT	GT	TT	CC

Table S6. Summary of the phylogenetic least squares regression model testing for the effects of larval diet (herbivorous versus non-herbivorous), ovipositor length and source of the data (this study versus literature sources) on ovipositor bristle number per valve. This analysis includes only species that have both larval diet and ovipositor bristle counts available. Changes in larval diet are significantly associated with changes in bristle number ($F_{1,19} = 5.801$, $p = 0.028$, n = 20 species, $\lambda = 1.01$).

term	estimate	s.e.m	*t*	*p*
(Intercept)	34.445	11.008	3.129	0.007
Diet (non-herbivorous)	-20.561	8.344	-2.464	0.025
Source (this study)	-4.527	4.967	-0.912	0.376
Ovipositor length	0.027	0.013	2.158	0.047

Table S7. Top SNPs associated with variation in ovipositor bristle number in *S. flava* from the pool genome-wide association study are shown in descending *p*-value ranking.

P-value ranking	Scaffold [scaffold length]	Position on scaffold	*P*-value	FDR q value	Nearby gene(s)	SNP location relative to gene		Gene function[a]
1 **6** **16**	scaffold00465 [104,011]	184 188 187	4.66E-09 2.94E-07 7.56E-07	0.008 0.069 0.076	**Muscle-specific protein**	453 bp 449 bp 450 bp	downstream downstream downstream	Actin binding; cytoskeleton organization; required for proper positioning of muscle nuclei, mitochondria, and neuromuscular junction
2	scaffold00015 [769,991]	186,025	4.26E-08	0.036	**heavyweight**	8,623 bp	downstream	Predicted to have phosphotyrosine residue binding activity; polymorphisms associated with body mass and starvation resistance
					cuticular protein 11B	1,625 bp	downstream	Chitin-based cuticle development
3 **21**	scaffold00063 [441,164]	140,206 140,207	6.69E-08 1.34E-06	0.037 0.107	**G protein alpha i subunit**	2 bp 3 bp	upstream upstream	Asymmetric neuroblast division and asymmetric protein localization involved in cell fate determination; cytoskeleton organization; and nervous system development
4	scaffold00053 [434,434]	135,098	1.20E-07	0.048	**sloppy paired 2**	560 bp	upstream	Transcription factor that regulates embryonic segment polarity and neural fate specification by temporal patterning of medulla neuroblasts
					CG11018	1,809 bp	upstream	Unknown
5	scaffold00071 [361,359]	290,082	1.42E-07	0.048	**CG32655**	9,932 bp	downstream	Unknown
					tenascin accessory	42,315 bp	downstream	Nervous system development; regulation of cell-cell adhesion; and synapse organization
7	scaffold00004 [1,460,103]	1,184,666	4.33E-07	0.069	**bacchus**	within 1st intron		Genetic regulator of behavioral response to ethanol and aggressive behavior
8	scaffold00298 [138,236]	48,050	4.78E-07	0.069	**Site-1 protease**	within 3rd intron		Serine endopeptidase involved in proteolysis
9	scaffold00088 [305,641]	170,647	5.23E-07	0.069	**CG42404**	within 3rd exon		Unknown

10	scaffold00098 [290,702]	51,932	5.38E-07	0.069	**Ecdysone-induced protein 75B**	4,151 bp	upstream	Encodes a nuclear receptor involved in ecdysis, chitin-based cuticle; regulation of ecdysteroid metabolic process; controls neuronal remodeling
11	scaffold00069 [371,200]	339,968	5.42E-07	0.069	**CG3216**	1,210 bp	upstream	Membrane receptor guanylyl cyclase
					proteasome alpha3 subunit	260 bp	upstream	Degrades polyubiquitinated proteins in the cytoplasm and nucleus
12	scaffold00045 [452,822]	130,642	5.54E-07	0.069	**pyramus**	52,851 bp	downstream	Fibroblast growth factor involved in larval development; mesoderm formation; and neurogenesis
					thisbe	40,381 bp	upstream	Fibroblast growth factor involved in glial cell development; larval muscle development; mesodermal cell migration and cell fate specification
13	scaffold00046 [471,014]	90,481	5.75E-07	0.069	**Ets at 97D**	89 bp	downstream	Transcription factor regulates mitochondrial mass; cellular response to starvation; egg chamber differentiation
						82 bp	downstream	
20		90,474	1.31E-06	0.107	**transcript 48**	410 bp	downstream	Ventral furrow formation during gastrulation
						417 bp	downstream	
14	scaffold00001 [2,846,614]	664,693	5.75E-07	0.069	**CG11560**	85,795 bp	downstream	Transcription factor expressed in the developing nervous system
					adenosine deaminase	6,126 bp	downstream	Predicted to catalyze deamination of adenosine and deoxyadenosine to inosine and deoxyinosine, but likely not an active enzyme
15	scaffold00080 [330,737]	162,285	6.37E-07	0.071	**kekkon 5**	18,362 bp	upstream	Transmembrane protein enriched in embryonic central nervous system and regulates the bone morphogenetic protein (BMP) signaling pathway involved in tissue patterning and growth
17	scaffold00107 [273,675]	220,388	7.74E-07	0.076	**CG17786**	4,651 bp	downstream	Unknown
					twin	6,663 bp	upstream	Degrades mRNA poly(A) tails; involved in female germ-line stem cell asymmetric division
18	scaffold00118 [259,145]	234,065	9.54E-07	0.084	**tubulin tyrosine ligase-like 1B**	within exon		Catalyzes ligations of amino acids to tubulins; predicted to be involved in microtubule cytoskeleton organization

19	scaffold00121 [259,553]	226,706	9.57E-07	0.084	**outsiders**	194 bp	upstream	Programmed cell death of primordial germ cell/pole cells
					CG8051	11,337 bp	upstream	Dorsal thorax formation and bristle development; expressed in blood-brain barrier surface glia
22	scaffold00065 [398,966]	19,097	1.46E-06	0.107	**nudC**	5,354 bp		Nucleus localization and positive regulation of dendrite morphogenesis
					CG9674	3,104 bp		Predicted to be involved in ammonia assimilation cycle and glutamate biosynthetic process
23	scaffold00047 [463,252]	34,694	1.51E-06	0.107	**nuclear fallout**	within intron		Actin cytoskeleton reorganization during furrow formation; sensory organ precursor cell fate determination

[a] Inferences about gene function were based on orthologous function in *D. melanogaster* from Flybase FB2020_01.

Table S8. Genes annotated with enriched gene ontology (GO) terms that intersected the most significant pool-GWAS SNPs.

Scaffold	Position	-log10(P)	Genome-wide Rank[b]	Gene ID (*S. flava*)	Ortholog ID (*D. mel.*)	Ortholog Name
DNA-binding transcription repressor activity, RNA polymerase II-specific (GO:0001227)						
scaffold00328	130182	5.763	23	scaffold00328-augustus-gene-0.67	FBgn0001168	*h*
scaffold00222	139175	5.364	39	scaffold00222-augustus-gene-0.55	FBgn0038244	*CG7987*
scaffold00521	31530	4.429	237	scaffold00521-augustus-gene-0.30	FBgn0008646	*E5*
scaffold00355	77106	4.414	244	scaffold00355-augustus-gene-0.40	FBgn0035137	*CG1233*
scaffold00751	23120	4.381	266	scaffold00751-augustus-gene-0.20	FBgn0011648	*Mad*
scaffold00332	46880	4.252	351	scaffold00332-augustus-gene-0.52	FBgn0034520	*lms*
scaffold00196	39864	4.217	369	scaffold00196-processed-gene-0.17	FBgn0002733	*E(spl)mβ-HLH*
scaffold00096	260245	4.190	379	scaffold00096-augustus-gene-1.80	FBgn0015371	*chn*
scaffold00115	115881	4.157	405	scaffold00115-augustus-gene-1.53	FBgn0038852	*HHEX*
scaffold00160	26969	4.023	509	scaffold00160-augustus-gene-0.20	FBgn0010109	*dpn*
scaffold00302	38793	3.696	943	scaffold00302-augustus-gene-0.63	FBgn0011655	*Med*
scaffold00012	430878	3.521	1269	scaffold00012-augustus-gene-4.41	FBgn0027788	*Hey*
scaffold00394	52743	3.442	1466	scaffold00394-processed-gene-0.15	FBgn0032295	*CG12299*
scaffold00216	115550	3.420	1520	scaffold00216-augustus-gene-0.117	FBgn0261983	*l(2)gd1*
scaffold00001	1171740	3.418	1528	scaffold00001-processed-gene-11.15	FBgn0038390	*Rbf2*
scaffold00363	49064	3.417	1531	scaffold00363-augustus-gene-0.72	FBgn0000097	*aop*
Phosphatidylinositol biosynthetic process (GO:0006661)						
scaffold00007	1028925	5.436	36	scaffold00007-augustus-gene-9.68	FBgn0015278[c]	*Pi3K68D*
scaffold00482	46732	5.374	37	scaffold00482-processed-gene-0.12	FBgn0034789[d]	*PIP5K59B*
scaffold00977	3771	4.408	246	scaffold00977-augustus-gene-0.24	FBgn0037916	*CG5342*
scaffold00172	34359	4.176	390	scaffold00172-augustus-gene-0.45	FBgn0034346	*PIG-O*

scaffold00006	144520	3.714	910	scaffold00006-augustus-gene-1.43	FBgn0029818	*GAA1*
scaffold00158	47689	3.704	929	scaffold00158-augustus-gene-0.29	FBgn0266438	*PIG-Z*
scaffold00012	159183	3.562	1184	scaffold00012-snap-gene-1.49	FBgn0034789	*PIP5K59B*
scaffold00046	138764	3.479	1376	scaffold00046-processed-gene-1.2	FBgn0265190	*PIG-S*
Establishment of cell polarity (GO:0030010)						
scaffold00063	140206	7.175	2	scaffold00063-augustus-gene-1.49	FBgn0001104	*Gαi*
scaffold00007	1028925	5.436	36	scaffold00007-augustus-gene-9.68	FBgn0036206[c]	*CG5964*
scaffold00482	46732	5.374	37	scaffold00482-processed-gene-0.12	FBgn0016984[d]	*sktl*
scaffold00297	77172	5.354	40	scaffold00297-augustus-gene-0.26	FBgn0263289	*scrib*
scaffold01112	22724	5.253	50	scaffold01112-augustus-gene-0.15	FBgn0019968	*Khc-73*

[a] These correspond to gene: SNP intersections summarized in Table 2.

[b] Different SNP filters were applied prior to candidate gene list construction and GO enrichment analyses (see Methods). The filter applied prior to the latter analysis was more stringent, resulting in the exclusion of more SNPs and, consequently, different p-value rankings (e.g., Gαi's p-value rank among all considered SNPs is #2 in the GO Enrichment analysis above, but #3 in the candidate gene investigation in Table 1).

[c] These two gene models physically overlap in *D. melanogaster* and consequently were assigned orthology to the same *S. flava* gene.

[d] scaffold00482-processed-gene-0.12 was assigned orthology to both FBgn0034789 and FBgn0016984, which encode proteins with a Phosphatidylinositol-4-phosphate 5-kinase core domain.

Table S9. Summary of the regression model testing for the additive effects of a candidate SNP on ovipositor bristle number from individually sequenced *S. flava* flies. The SNP was identified from the pool-GWAS on ovipositor bristle number and is located upstream of the neural development gene *G alpha i subunit* (*Gαi*). The model also tests for the effects of peg pool (high or low), ovipositor length, population/host plant (which could not be disentangled; NH1/*Barbarea vulgaris*/*Turritis glabra*, NH2/*B. vulgaris*, or NH2/*T. glabra*). For peg pool categorization, the reference category is 'high,' and for host plant/population, it is 'NH1/*Barbarea vulgaris*/*Turritis glabra*'. Sequence data for this analysis has been deposited on Genbank (accession no. MH884655-MH884734). $R^2 = 0.919$, $F_{5,68} = 154.9$, $p < 0.001$, n = 74 individuals.

term	estimate	s.e.m	*t*	*p*
intercept	29.983	3.446	8.702	<0.001
SNP	0.583	0.203	2.876	0.005
Bristle pool (low)	-8.11	0.325	-24.95	<0.001
Ovipositor length	50.972	13.75	3.708	<0.001
Population/Host plant (NH2/*B. vulgaris*)	-0.889	0.392	-2.27	0.026
Population/Host plant (NH2/*T. glabra*)	0.556	0.391	1.422	0.16

Table S10. Summary of the regression model testing for the effects of a candidate SNP (additive), peg pool (high or low), and population/host plant (NH1/*Barbarea vulgaris*/*Turritis glabra*, NH2/*B. vulgaris*, or NH2/*T. glabra*) on ovipositor length from individually sequenced *S. flava* flies. The candidate SNP was identified from the pool-GWAS on ovipositor bristle number and is located upstream of the neural development gene *G alpha i subunit* (*Gαi*). For peg pool categorization, the reference category is 'high,' and for host plant/population, it is 'NH1/*Barbarea vulgaris*/*Turritis glabra*'. Sequence data for this analysis has been deposited on Genbank (accession no. MH884655-MH884734). $R^2 = 0.108$, $F_{4, 69} = 2.085$, $p = 0.092$, n = 74 species.

term	estimate	s.e.m	*t*	*p*
intercept	0.249	0	74.43	<0.001
SNP	<0.001	0	0.177	0.86
Bristle pool (low)	<0.001	0	0.013	0.99
Population/Host plant (NH2/*B. vulgaris*)	0.009	0	2.81	0.006
Population/Host plant (NH2/*T. glabra*)	0.003	0	0.956	0.342

Table S11. Linkage disequilibrium estimates between the focal SNP identified from the pool-GWAS (locus 214), located upstream of the neural development gene *Gai*, and neighboring variants identified from individual re-sequencing. Delta values estimate correlations among unphased alleles. Strong LD values (| δ | > 0.5 (*p*-value < 0.01) are shown in bold italics.

	Delta δ						
Variant locus	**AA**	**Aa**	**aA**	**aa**	**T2**	**df**	***p*-value**
49	0.156	***0.56***	***-0.872***	0.156	230.27	1	5.20E-52
213	-0.23	***0.608***	-0.149	-0.23	106.03	1	7.26E-25
215	-0.253	0.253	0.253	-0.253	48.64	1	3.08E-12
235	-0.291	0.291	0.291	-0.291	65.01	1	7.47E-16
236	0.285	-0.285	-0.285	0.285	68.1	1	1.55E-16

Appendix C: Supplemental figures from Chapter 1.

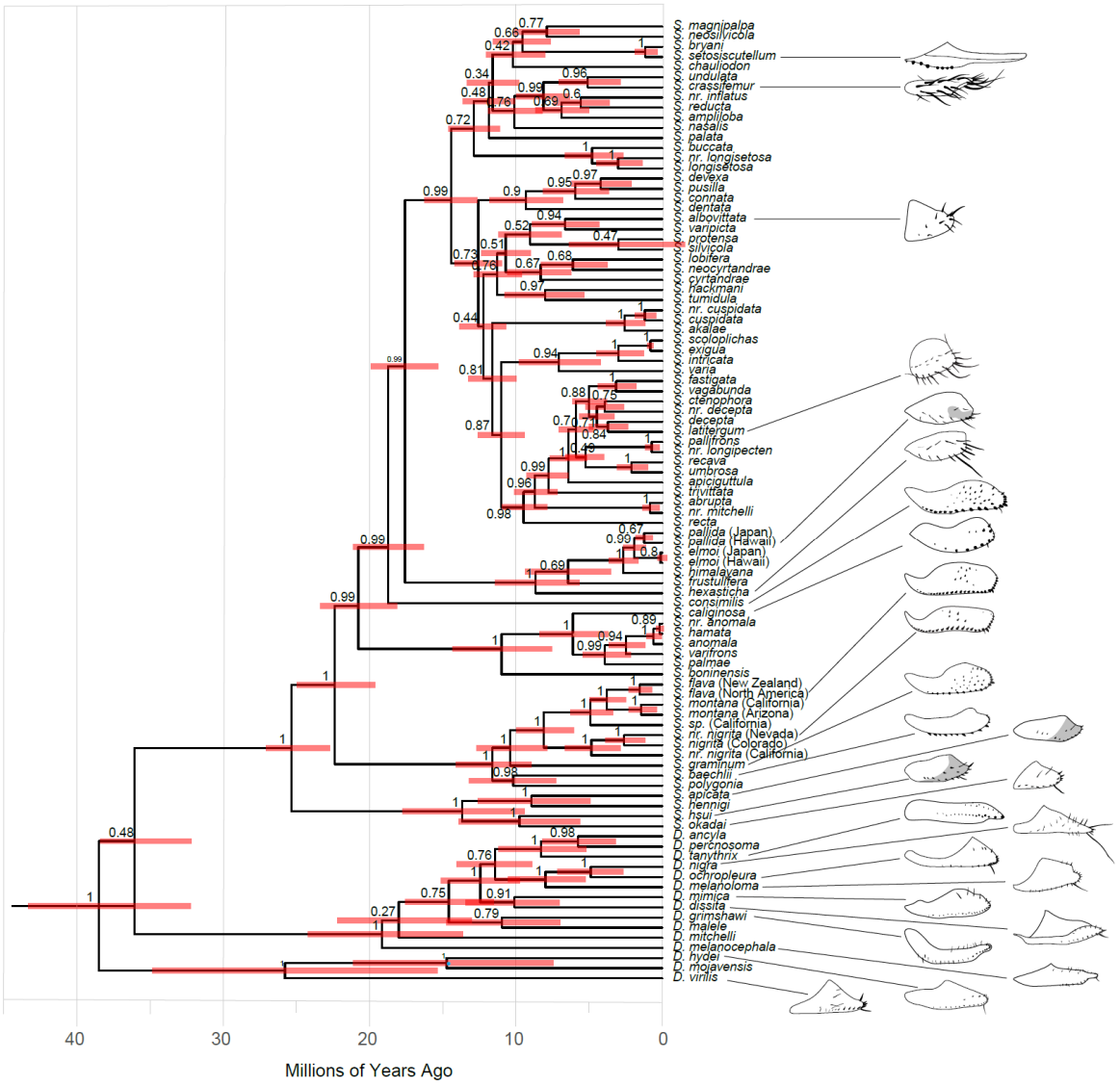


Figure S1. Time calibrated phylogeny of the subgenera *Drosophila* and *Scaptomyza* inferred from Bayesian analysis, using 11 genes (16S, 28S, *Adh*, *Cad-r* (rudimentary), *COI*, *COII*, *gstd1*, *gpdh*, *marf*, *ND2*, *n(l)tid)*, *Orco*) and fossil and biogeographic time calibrations. Nodes indicate posterior probabilities; red bars represent 95% highest posterior density intervals around the mean node age. Illustrations of female ovipositors, indicating bristle number, were drawn from sources indicated in Table S1.

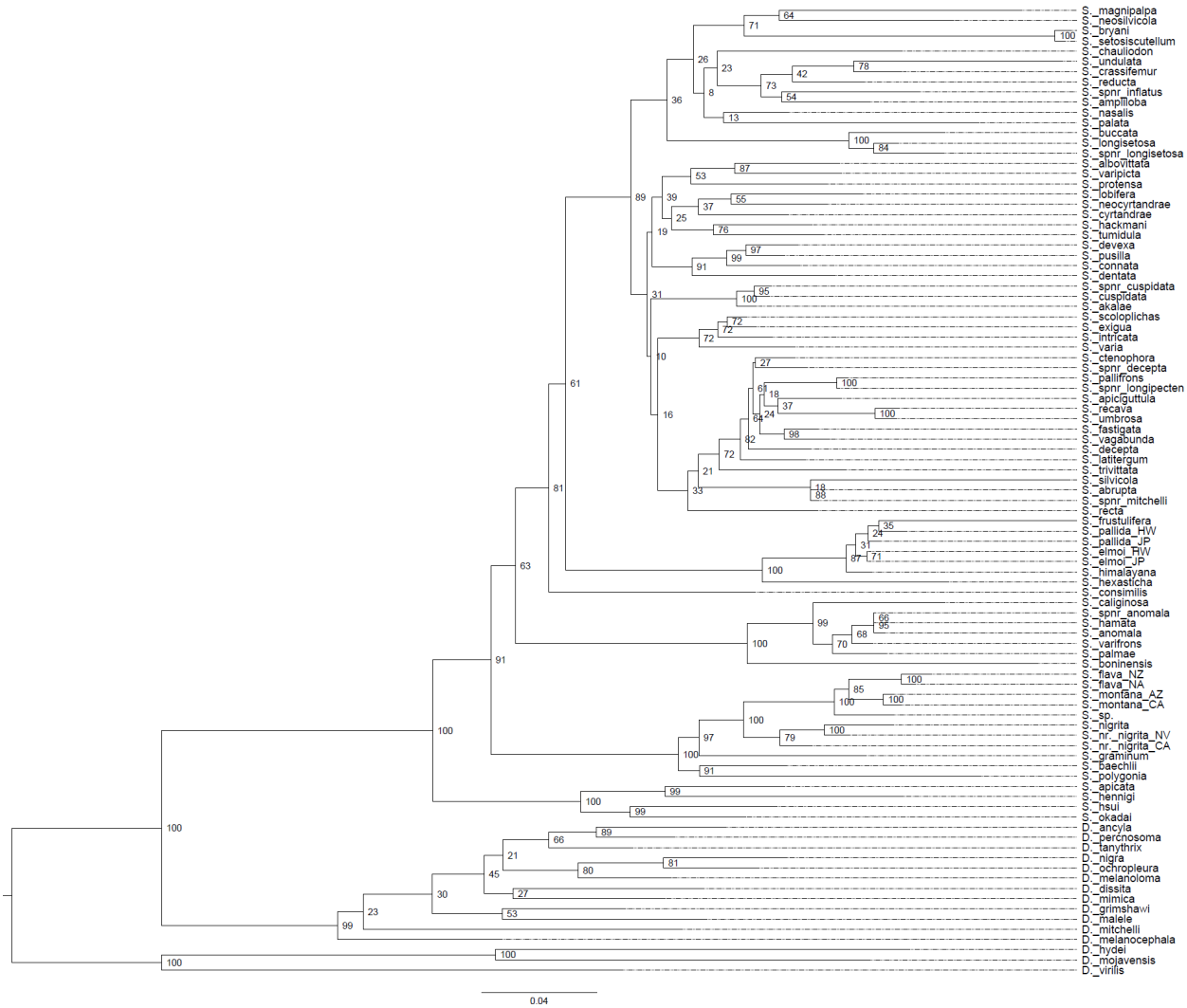


Figure S2. Phylogeny of subgenera *Drosophila* and *Scaptomyza* inferred from maximum likelihood (ML), using 11 genes (16S, 28S, *Adh*, *Cad-r* (rudimentary), *COI*, *COII*, *gstd1*, *gpdh*, *marf*, *ND2*, *n(l)tid)*, *Orco*) and fossil and biogeographic time calibrations. Nodes represent bootstrap values (≥50%) from the ML analysis.

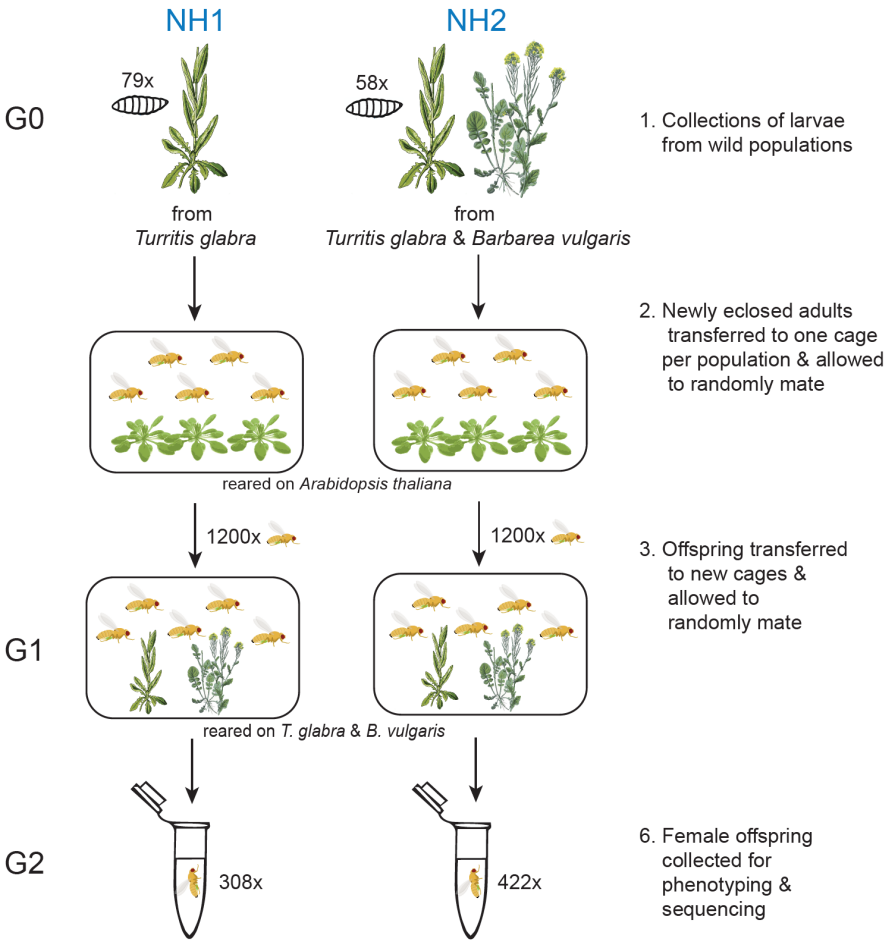
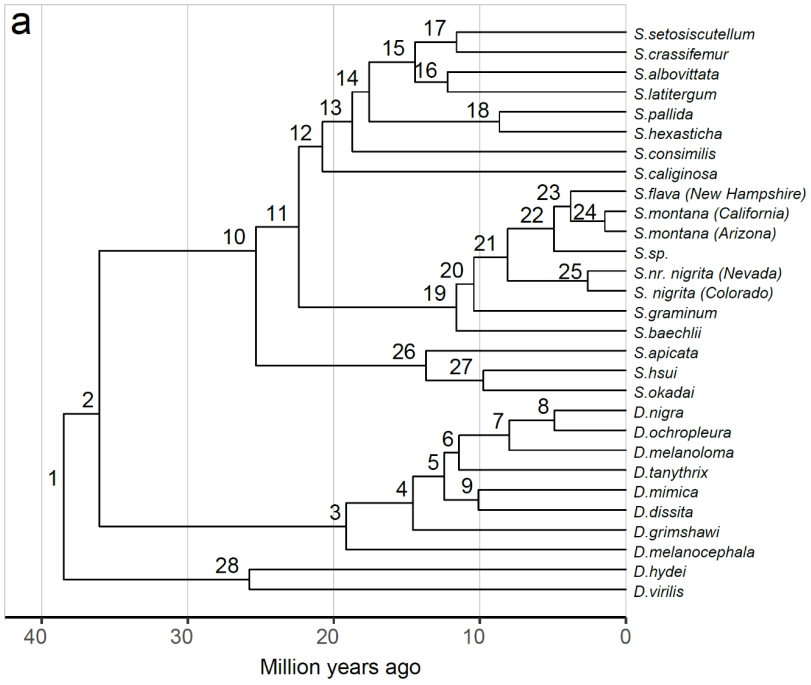


Figure S3. Mating and rearing scheme of fly lineages used for the pool-GWAS on ovipositor bristle number in *Scaptomyza flava*.

a



b

Node	Lower bound	Ovisenislla number ASR	Upper bound
1	15.32	17.59	32.91
2	16.11	16.41	32.52
3	22.44	13.58	36.02
4	23.74	10.88	34.63
5	22.80	9.45	32.25
6	22.49	9.46	31.95
7	21.34	9.25	30.59
8	21.83	8.31	30.14
9	21.33	10.15	31.49
10	15.71	12.74	28.46
11	17.34	11.28	28.63
12	16.92	10.92	27.84
13	17.00	10.49	27.50
14	15.62	10.34	25.95
15	13.60	10.22	23.82
16	13.07	10.89	23.96
17	11.60	10.96	22.56
18	9.69	11.10	20.79
19	25.16	10.23	35.39
20	27.90	9.58	37.49
21	31.51	8.79	40.30
22	34.70	7.36	42.06
23	35.30	6.74	42.04
24	36.42	4.86	41.27
25	35.93	6.58	42.50
26	9.85	12.53	22.38
27	9.26	11.47	20.73
28	14.07	18.00	32.06

Figure S4. Confidence intervals for ancestral state reconcstruction of ovipositor bristle number among *Drosophila* and *Scaptomyza* species. Values in the table correspond to respective nodes given in the phylogeny.

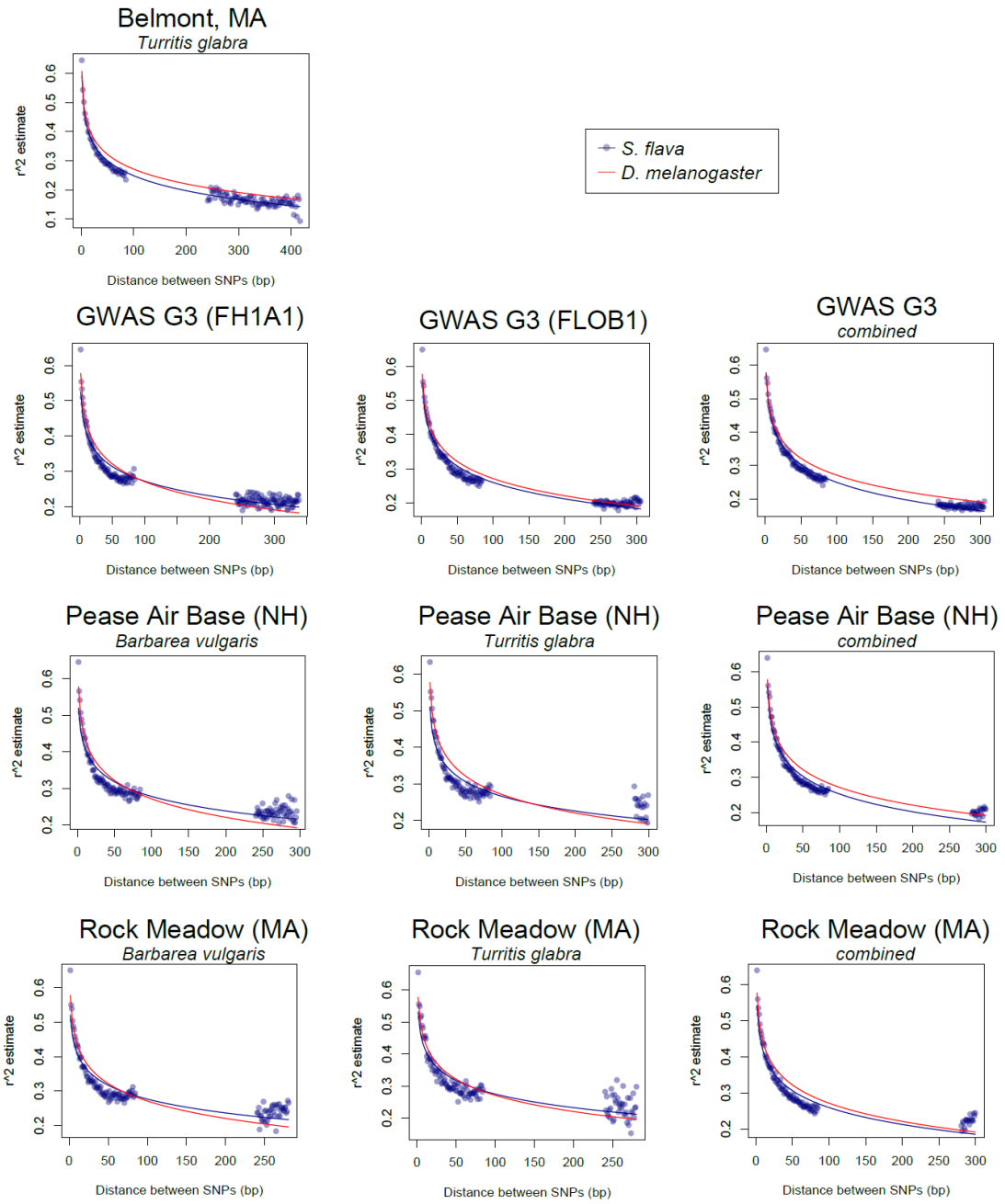


Figure S5. Linkage disequilibrium across multiple populations of *S. flava*. LD, inferred from pooled sequencing of wild and lab-reared *S. flava*, decays at a similar rate across different sites from Massachusetts (Belmont and Rock Meadow) and New Hampshire (Pease Air Base), across different host plants (*Turritis glabra* and *Barbarea vulgaris*), and among the lineage of flies used for the pooled-GWAS (G3). This rate is similar to that seen in *D. melanogaster*, shown in red.

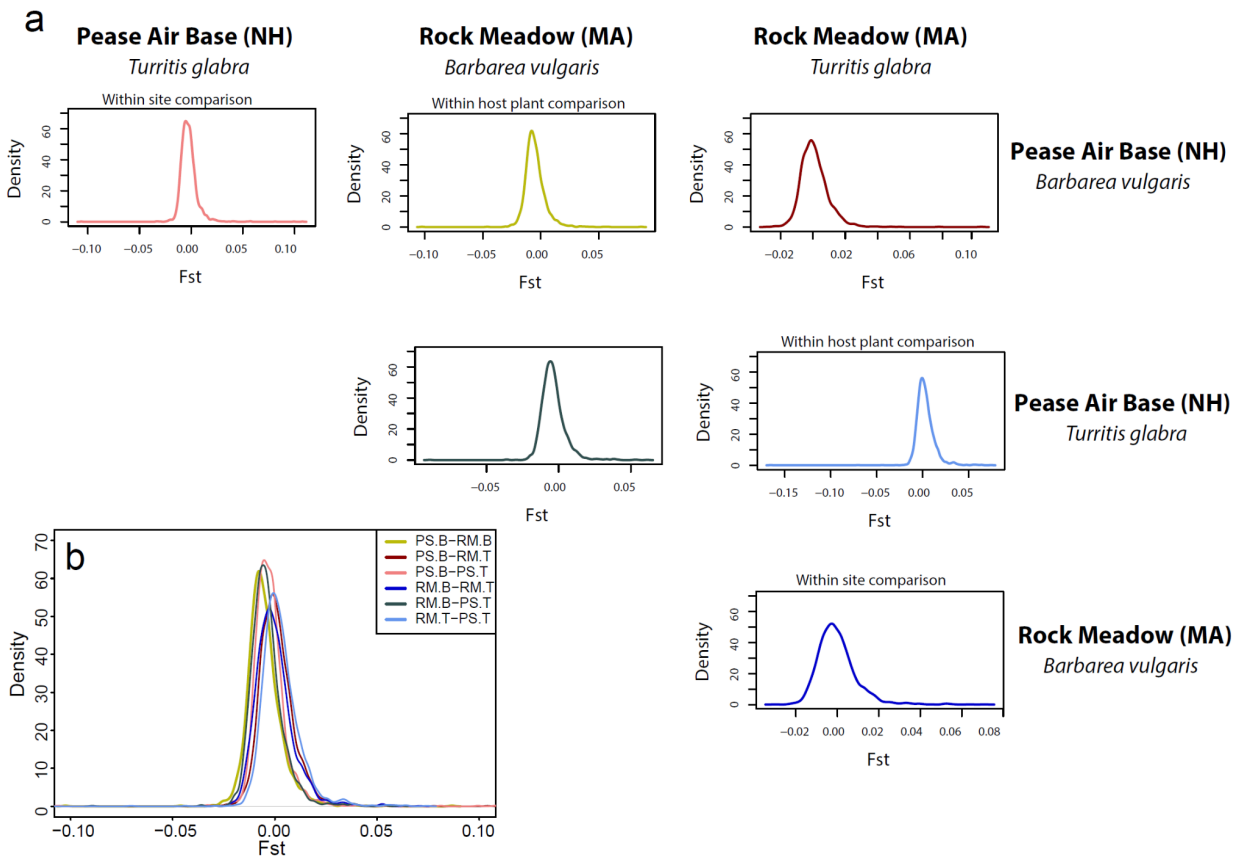


Figure S6. Low levels of population structure across *S. flava* populations between sites (Rock Meadow, MA and Pease Air Base, NH) and host plants (*Turritis glabra* and *Barbarea vulgaris*). Pairwise Fst distributions between sampled populations at different sites and different host plants, shown as paired comparisons in individual plots (a) and in aggregate (b).

Appendix D: Supplemental methods from Chapter 2.

S. flava genome sequencing and assembly (sfla_v2).

Dovetail HiC library preparation and sequencing. *S. flava* flies used for the genome assembly were collected from a partially inbred laboratory colony. The colony was founded from >150 larvae collected near Dover, NH, USA, and subsequently maintained for several years in the laboratory. 300 male flies were flash frozen and stored at -80°C. A Dovetail HiC library was prepared in a similar manner as described previously (Erez Lieberman-Aiden et al., 2009). Briefly, for each library, chromatin was fixed in place with formaldehyde in the nucleus and then extracted. Fixed chromatin was digested with DpnII, the 5' overhangs filled in with biotinylated nucleotides, and free blunt ends were ligated. After ligation, crosslinks were reversed, and the DNA purified from protein. Purified DNA was treated to remove biotin that was not internal to ligated fragments. DNA was then sheared to ~350 bp mean fragment size, and sequencing libraries were generated using NEBNext Ultra enzymes and Illumina-compatible adapters. Biotin-containing fragments were isolated using streptavidin beads before PCR enrichment of each library. The libraries were sequenced on an Illumina HiSeqX to produce 380 million 2x150 bp paired end reads, approximately 30x sequence coverage.

PacBio library and sequencing. DNA samples were quantified using Qubit 2.0 Fluorometer (Life Technologies, Carlsbad, CA, USA). SMRTbell libraries (~20kb) for PacBio Sequel were constructed using SMRTbell Template Prep Kit 1.0 (PacBio, Menlo Park, CA, USA) using the manufacturer's recommended protocol. The pooled library was bound to polymerase using the Sequel Binding Kit 2.0 (PacBio) and loaded onto PacBio Sequel using the MagBead Kit V2 (PacBio). Sequencing was performed on two PacBio Sequel SMRT cells, using Instrument Control Software Version 5.0.0.6235, Primary analysis software Version 5.0.0.6236 and SMRT Link Version 5.0.0.6792

Illumina library and sequencing. Illumina sequence data was generated from the "OHI-9 line" of *S. flava*, derived from a laboratory population initially collected in Belmont, MA, USA in 2008 that was inbred through 10 generations of single pair mating on *A. thaliana* Col-0 plants. Paired-end 180 bp and 300 bp insert libraries and 3 kbp and 5 kbp mate pair libraries from OHI-9 female flies were sequenced with 100 bp read length on an Illumina Hiseq 2000 at the University of Arizona. Reads were quality filtered and Illumina TruSeq3 adapters were removed using Trimmomatic v0.35 (Bolger et al., 2014) with the following parameters: "LEADING:10 TRAILING:10 SLIDINGWINDOW:4:15 MINLEN:99". Falcon and HiRise initial assembly. An initial draft assembly of sfla_v2 was generated by Dovetail using the following methods but was not used because of a high rate of errors detected and an inflated genome size, which we suspected was caused by high heterozygosity in the assembly. We report these methods, however, because this assembly was used for scaffolding our final assembly (details below). The genome assembly was performed using the FALCON 1.8.8 pipeline from Pacific Bioscience. First, 70-fold whole-genome, single-molecule, real-time sequencing (SMRT) data of *S. flava* was used as input to the traditional FALCON pipeline using a length cut-off that correspond to 50x coverage of data during the initial error-correcting stage. This resulted

in 0.511 million error corrected reads with an N50 read length equal to 19.6 kb. Second, the error-corrected reads were processed by the overlap portion of the FALCON pipeline. The aligned reads were assembled in the third stage of FALCON into 3,561 primary contigs containing 443.7 Mbp with an NG50 contig length of 543.7 kbp. Finally, the assembly was polished through PacBio's Arrow algorithm from SMRT Link 5.0.1, using the original raw-reads. The input de novo assembly, and Dovetail HiC library reads were used as input data for HiRise, a software pipeline designed specifically for using proximity ligation data to scaffold genome assemblies (Putnam et al., 2016). Dovetail HiC library sequences were aligned to the draft input assembly using bwa (http://github.com/lh3/bwa). The separations of Dovetail HiC read pairs mapped within draft scaffolds were analyzed by HiRise to produce a likelihood model for genomic distance between read pairs, and the model was used to identify and break putative misjoins, to score prospective joins, and make joins above a threshold.

Reassembly. As mentioned above, while this initial de novo assembly exhibited high scaffold contiguity, we discovered that it unfortunately contained a significant number of frameshift mutations in coding sequences that limited our ability to correctly annotate the genome. A BUSCO (v5.2.2, (Manni et al., 2021)) analysis only showed 83.4% complete BUSCOs, notably lower than typical long-read drosophilid genome assemblies (e.g., Kim et al. 2021). A manual examination of preliminary annotation records showed that approximately 15% of randomly checked genes contained one or more deletions resulting in fragmented exons. We attempted to further polish the scaffolds with Illumina data but that resulted in only 97.4% complete BUSCOs and many of the manually checked gene annotation records contained indels.

Re-assembly. We thus decided to reassemble the genome, using the same data described here, with the long-read hybrid assembly pipeline described in Kim et. al (2021), which has been shown to produce highly complete genome assemblies for drosophilid flies with very few coding sequence indels (B. Y. Kim et al., 2021). Briefly, we generated an initial draft assembly with Flye 2.9 (Kolmogorov et al., 2019), identified and removed duplicated haplotypes (haplotigs) with purge_dups (Guan et al., 2020), polished the draft assembly using the PacBio reads with one round of Racon (Vaser et al., 2017), then polished further using Illumina reads with one round of Pilon (Walker et al., 2014), only fixing base-level errors. The fully polished assembly was scanned for contaminant sequences using NCBI BLAST (M. Johnson et al., 2008) and BlobTools (Laetsch & Blaxter, n.d.). Repetitive sequences in the assembly were identified with RepeatModeler2 (Flynn et al., 2020). This produced a fairly contiguous 323.9 Mbp assembly contained in 1,430 contigs, with an N50 of 575,920 bp and L50 of 148. BUSCO completeness was significantly improved (98.8%) and manual spot checks of another preliminary annotation showed no indel issues.

Scaffolding. To fully integrate the long-range assembly information we had previously generated with HiC, we scaffolded the less contiguous but more accurate new assembly with the error-prone Dovetail version, reasoning that base-level errors in the previous version were unlikely to impact reference-based scaffolding. Specifically, we soft-masked both genomes using the repeat library generated in the previous step, using

RepeatMasker (Smit et al., 2013). Then, we created a whole-genome alignment with Progressive Cactus (Armstrong et al., 2020) and used the RagOut reference-based scaffolder (Kolmogorov et al., 2014) to scaffold the new genome, using the old assembly as the reference. The final genome assembly of *S. flava* (sfla_v2) was scaffolded into 1,252 scaffolds covering 315.4 Mbp (N50 = 32.966 Mb) with a maximum length of 85.98 Mbp and with 460 gaps.

Comparative annotation of *Scaptomyza* genomes. Gene annotations previously created for an Illumina-only *S. flava* assembly (sfla_v1) (described below) were transferred to the assemblies of *S. graminum, S. hsui, S. pallida, S. montana*, and this newer *S. flava* assembly, using whole-genome Progressive Cactus alignments and the Comparative Annotation Toolkit (CAT, (Fiddes et al., 2018)). A *D. grimshawi* genome (B. Y. Kim et al., 2021) was used as an outgroup. All genomes were repeat-masked using repeat libraries generated with RepeatModeler2 and soft-masked using RepeatMasker. A published phylogeny for these species (Suvorov et al., 2022) was used as a guide tree for the alignment. RNA-seq data generated from female *S. flava* gut, proboscis, and maxillary palp organs was downsampled to 100X coverage using BBTools (Bushnell, 2014) and aligned to the *S. flava* genome using STAR (Dobin et al., 2013) to assist CAT annotation. RNA-seq data are currently not available for any of the other species. Briefly, CAT uses the transMap mode to project annotations from a reference (i.e, the original *S. flava* genome) onto the target genomes, evaluating the projections and correcting them with Augustus (Stanke et al., 2008) and utilizing RNA-seq evidence if it exists.

S. flava Illumina-only assembly (sfla_v1).

Prior to the generation of our main assembly sfla_v2, an *S. flava* assembly generated from Illumina sequencing was utilized by various projects on this species (e.g. (Peláez, Gloss et al., 2020), (Gloss et al., 2022), Verster et al. 2019). The sequence data was also used in polishing steps in the long read assembly, for generating automated annotations that were then also copied over to the newer assembly, sfla_v2, and for estimating repeat coverage.

Assembly and annotation. Read pairs that survived quality filtering were subsampled to an estimated ~90x coverage of the *S. flava* genome and assembled using ALLPATHS-LG (Gnerre et al., 2011) on the XSEDE high performance computing system. Contigs were extended and ambiguous regions were resolved iteratively using GapCloser (Luo et al., 2012) to produce the primary assembly used for all analyses (unless otherwise noted). Prior to annotation, repeat regions were masked using RepeatMasker (Smit, AFA, Hubley, R & Green, P., 1996-2010) with the *Drosophila* repeat library. Protein-coding genes were annotated using MAKER2 (Holt & Yandell, 2011), with the *S. flava* transcriptome (Whiteman et al., 2012) and predicted gene sequences from 12 *Drosophila* species (FlyBase release 2013_06) provided to inform gene models. We recovered 17,997 genes in our liberal gene set (annotations predicted by Augustus). The proportion of core dipteran genes recovered in our assembly was determined with BUSCO v5.4.2 (Simão et al., 2015).

Repetitive element content. Repeat identification was carried out using both homology-based and ab-initio approaches. We used the Drosophila RepBase repeat database for the homology-based annotation (http://www.girinst.org/repbase; update 20150807) within RepeatMasker version open-4.0.6 (Smit, AFA, Hubley, R & Green, P., 1996-2010). The RepeatMasker option -gccalc was used to infer GC content for each contig separately to improve the repeat annotation. Ab-initio repeat finding was carried out using RepeatModeler version 1.73 (http://repeatmasker.org/RepeatModeler.html).

Gene model validation.

To validate lost genes GOIs, we performed an additional TBLASTN search for each GOI in *S. flava*, as well as genes assumed to be proximal to GOIs (PGOIs) based on their location in *D. grimshawi* or *D. virilis*. These searches used the predicted orthologs in *D. grimshawi* as queries. Genes of interest were considered truly absent if TBLASTN searches yielded no hits. In some cases, homology between conserved protein domains resulted in weakly supported hits to the *S. flava* genome and/or transcriptome, in which case the aligned region was extracted, translated, and BLASTed against the NCBI nr database. The gene was considered lost if the output showed stronger homology to genes other than the GOIs. To avoid confirmation bias, we also did this with the *S. flava* PGOIs to ensure their identity matched the expected ortholog in *D. grimshawi*. The absence of the GOIs, coupled with the presence of 95% of the PGOIs strongly support that the GOIs are truly lost and are not an artifact of missing scaffolds in the genome assembly. We implemented a similar BLAST-based search of both genomes to validate genes that underwent lineage-specific expansions in *S. flava*.

Finally, to guard against errors in our curation of the published *Drosophila* genomes, we compared our gene curations to those from published studies (Almeida et al., 2014; Good et al., 2014; Low et al., 2007) and those inferred in OrthoDB (Zdobnov et al., 2017). For any discrepancies, we performed comprehensive TBLASTN searches against the relevant genome assemblies to search for the full complement of orthologous genes, re-curated gene models if necessary, and manually inspected aligned gene models. In a few cases, we discarded genes that had high similarity (>99% identity) and perfect synteny to another scaffold in the assembly because these are likely artifactual duplicates.

Population genomics.

DNA pooled from 45 wild-collected *S. flava* larvae was sequenced to yield 100 bp paired-end reads on an Illumina HiSeq 2000. Methods on sample collection, sequencing, and read mapping can be found in Pelaez, Gloss et al. 2022. Nucleotide diversity (π) was calculated across four-fold degenerate sites using Popoolation v.1.2.2 (Kofler et al., 2011) for repeat-masked autosomal scaffolds greater than 20 kb in length. Variance-sliding.pl for repeat-masked autosomal scaffolds (N = 819) greater than 20 kb in length. Parameters were: a ploidy level of 90, minimum allele count of two, minimum quality score of 20, minimum coverage of four, and a maximum coverage of 100. Scaffolds were considered autosomal if more than half the predicted proteins had a best blastp score to an autosome in *D. melanogaster*.

Horizontal gene transfer.

We expanded our previous search for HGT events in *S. flava* (Verster et al., 2019), which evaluates the evidence for HGT using a phylogenetic approach, with less conservative search criteria to minimize the risk of false negative results. Specifically, we relaxed the criteria initially used to filter out scaffolds as likely contaminants on the basis of gene number and orthology: we no longer required a minimum number of protein-coding genes per scaffold, and we retained scaffolds if even a single gene showed stronger homology to arthropods than non-arthropods (alien index < 0). Applying this approach to the gene annotation set from *S. flava* assembly v1.0 uncovered 11 new HGT candidates that were subjected to phylogenetic analysis. Aside from the contaminant filtering described above and the initial homology search (which used e-values from BLASTP v. 2.3.30+ instead of phmmer bitscores), all steps were conducted as described in (Verster et al., 2019). Manual inspection of the resulting phylogenies revealed no new support for HGT.

Appendix E: Supplemental tables from Chapter 2.

Table S1. Species included in analyses and their assembly versions.

Species	Assembly version	CAFE	PAML
S. flava	sfla_v1		
S. flava	sfla_v2	✓	✓
S. montana	iso-CA-L1	✓	✓
S. graminum	TMU-2019	✓	✓
S. pallida	iso-CA-L1	✓	✓
S. hsui	iso-CA-L1	✓	✓
D. grimshawi	dgri_caf1	✓	✓
D. mojavensis	dmoj_caf1	✓	✓
D. virilis	dvir_caf1	✓	✓
D. melanogaster	Release_6_plus_ISO1_MT	✓	✓
D. ananassae	dana_caf1	✓	
D. erecta	dere_caf1	✓	
D. pseudoobscura	Pse_3.0	✓	

Table S2. *Scaptomyza* genome assembly statistics.

Assembly	Sequencing	Number of scaffolds	Longest scaffold (Mbp)	Assembly length (Mbp)	N50 (Mbp)	L50	Automated annotations
S. flava (sfla_v2)	PacBio + HiC + Illumina	781	92.65	331.7	31.83	3	12,365
S. flava (sfla_v1)	Illumina	6619	2.85	216.1	0.112	404	17,997
S. montana	Nanopore	735	22.67	229.1	2.44	18	11,924
S. graminum	Nanopore	352	21.82	137.8	17.41	4	11,747
S. pallida	Nanopore	273	10.12	201.7	4.28	15	11,701
S. hsui	Nanopore	313	20.51	223.5	5.29	10	11,680

Table S3. Chemosensory and detoxification gene family sizes across drosophilid genomes.

family	***Dmel***	***Dpse***	***Dere***	***Dana***	***Dvir***	***Dmoj***	***Dgri***	***Spal***	***Shsu***	***Sgra***	***Smon***	***Sfla***
CYP450	89	79	88	93	81	78	85	72	69	68	67	73
GST	41	36	42	46	34	33	32	29	30	31	34	35
UGT	35	29	30	37	29	25	28	26	25	22	22	22
GR	65	58	56	67	62	61	75	78	72	64	57	61
IR	58	58	63	62	54	54	59	55	63	52	51	47
OBP	52	45	50	50	41	42	50	41	41	33	34	34
OR	62	65	59	66	56	62	63	68	65	65	60	65
PPK	31	30	31	31	29	32	29	31	31	31	31	30
TRP	13	16	13	13	13	13	13	13	13	12	11	13
Detoxification (all)	165	144	160	176	144	136	145	127	124	121	123	130
Chemosensation (all)	281	272	272	289	255	264	289	286	285	257	244	250
Random Gene Set	186	183	184	206	212	201	238	196	195	189	193	198

Table S4. Orthologous genes identified as significantly expanding or contracting across the drosophilid tree. Family-wide *P*-value indicates whether significant shifts in turnover rate were detected within a given orthology group. Viterbi *P*-values indicate whether turnover rate was significantly different at individual branches within the phylogeny. Branch IDs (0-22) correspond to the labels shown in Fig. S10. Viterbi *P*-values <0.05 are highlighted in bold black text/pale green cells for all branches, except for the ancestral herbivore branch in which these values are highlighted in white bold text/dark green cells. Random chosen orthology groups are given by cluster ID with Viterbi *P*-values <0.05 highlighted in bold black text/pink cells.

		Viterbi *P*-values for individual branches																					
Gene	**Family-wide *P*-value**	**0**	**3**	**2**	**4**	**1**	**6**	**5**	**19**	**8**	**17**	**10**	**12**	**11**	**14**	**13**	**16**	**15**	**18**	**9**	**21**	**20**	**22**
G36abc	0.0329	**0.019**	0.588	**0.019**	0.071	0.051	0.606	0.547	**0.018**	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
Gr28b	0.0003	0.572	0.547	0.528	**0**	0.539	0.606	**0.036**	**0.001**	0.802	0.618	0.587	0.587	0.234	0.732	0.785	0.744	0.534	0.762	0.133	0.59	0.776	0.776
Gr39aA	0	0.572	0.547	**0.009**	0.528	**0.028**	0.259	0.057	**0.018**	0.111	**0.01**	**0.008**	0.602	0.265	0.156	0.807	0.096	0.539	0.225	0.601	0.59	0.776	0.413
Gr59cd	0.028	0.63	0.588	0.554	0.554	0.574	0.682	0.588	0.123	0.22	0.133	0.571	0.571	0.629	0.232	0.186	**0.024**	0.534	0.386	**0.004**	0.557	0.703	0.291
Gr85a	0	0.096	0.547	0.528	0.528	0.539	0.606	**0.036**	**0.009**	**0**	0.601	0.111	**0.029**	0.654	**0.003**	0.759	0.31	0.528	0.161	0.109	0.574	0.744	0.357
Gr92a93bcd	0.0001	**0.001**	0.624	**0**	0.103	0.605	0.737	0.117	0.592	0.652	0.544	0.555	0.069	0.602	0.642	**0.046**	0.609	0.512	0.622	0.544	0.539	**0.028**	0.083
Gr98a	0.04965	0.572	0.547	0.528	0.528	0.539	0.053	0.547	0.065	0.652	**0.002**	0.571	0.091	0.629	0.676	0.726	0.687	0.523	0.705	0.544	0.539	0.652	0.212
Ir47abc94abc	0.00545	**0.023**	0.611	**0.039**	0.57	0.594	0.334	0.611	0.616	0.608	0.53	0.519	0.519	0.537	0.554	0.172	**0.043**	0.508	**0.049**	**0.001**	0.551	0.683	0.124
IR52	0.0185	0.659	0.611	0.106	0.57	0.111	0.274	0.587	0.591	0.649	0.544	0.554	0.068	0.601	0.18	0.685	0.064	0.512	0.62	0.544	0.539	0.649	**0.007**
IR56a	0.0001	0.549	0.532	0.519	0.519	0.527	0.574	0.532	0.534	0.559	**0**	0.519	0.519	**0.006**	**0.002**	0.685	**0.01**	0.512	0.62	0.515	0.514	0.559	0.559
IR60bcdfo	0.04565	0.078	0.587	0.082	**0.029**	0.574	0.677	0.092	**0.035**	0.559	0.515	0.519	0.519	0.537	0.554	0.576	0.062	0.504	0.546	0.515	0.514	0.559	0.559
IR67a	0.0693	0.549	0.532	0.519	**0.01**	0.527	0.574	0.532	0.534	0.559	0.515	0.537	0.537	0.571	**0.03**	0.017	0.541	0.504	0.546	0.515	0.514	0.559	0.559
IR76a	0.0083	0.549	0.532	0.519	0.519	0.527	0.574	0.532	0.534	0.559	**0.023**	**0.01**	**0.046**	0.571	0.601	0.637	0.117	0.508	**0.049**	0.515	0.514	0.559	0.559
IR7d	0.01955	0.549	0.532	0.519	0.519	0.527	0.574	0.532	0.534	**0.002**	0.515	0.519	0.519	0.537	0.554	0.576	0.541	0.504	**0**	0.515	0.514	0.559	0.559
IR7e	0.0017	0.549	0.532	0.519	0.519	0.527	0.574	0.532	0.534	**0**	0.515	0.519	0.519	0.537	0.554	0.576	0.062	0.504	0.069	0.515	**0.007**	0.5	0.5
IR94d	0.0478	0.549	0.532	0.519	0.519	0.527	0.574	0.532	0.534	0.089	0.515	0.519	**0.023**	0.537	0.066	0.576	0.541	0.504	0.546	0.515	0.514	0.089	0.559
Obp18a	0.0144	0.543	0.528	**0.025**	0.517	0.523	0.096	0.528	0.529	0.552	0.513	0.5	0.5	0.5	0.5	0.032	0.536	0.503	0.54	0.513	0.512	0.552	0.077
Obp22a	0.0204	0.064	0.528	**0.008**	0.517	0.523	0.096	0.528	**0.044**	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
Obp51a56fi	0.0086	**0.044**	0.553	0.533	**0.017**	**0.012**	0.096	0.528	**0.044**	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
Obp57ab	0.0886	0.543	**0.014**	0.533	0.533	0.523	0.096	0.528	**0.044**	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
Obp58b	0.0192	0.543	0.528	0.517	0.517	0.523	0.565	0.528	0.529	**0**	0.513	0.5	0.5	0.5	0.5	0.032	0.536	0.503	0.54	0.513	0.512	0.552	0.552
Obp58c	0.00555	0.543	0.528	0.517	0.517	0.523	0.565	0.528	0.529	**0**	0.513	0.5	0.5	0.5	0.5	0.032	0.536	0.503	0.54	0.513	0.512	0.552	0.552
Obp99b	0.04745	**0.001**	0.528	**0.025**	0.517	0.523	0.565	0.528	0.529	0.552	0.513	0.505	0.505	0.51	0.515	0.522	0.536	0.503	0.54	0.513	0.512	0.552	0.552
Or22ab	0.01125	**0.003**	0.579	**0.018**	0.549	0.567	0.666	0.056	0.544	0.107	0.52	0.5	0.5	0.5	0.5	0.043	0.072	0.505	0.56	0.52	0.518	0.577	0.577
Or42b	0.02145	0.565	0.542	0.525	0.525	0.535	0.595	0.542	0.544	**0**	**0.007**	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.52	0.518	0.577	0.577

Or59a	0.0316	0.565	0.542	0.525	0.525	0.535	0.595	0.542	0.06	0.687	0.558	0.521	0.521	**0.003**	0.651	0.696	0.192	0.515	0.652	0.053	0.535	0.074	0.195
Or59bORN2	0.00065	0.565	0.542	0.525	0.525	0.535	0.595	**0.034**	0.113	0.687	**0.003**	**0.045**	0.594	0.663	0.322	0.765	**0.045**	0.525	**0.018**	0.558	0.551	**0.012**	0.687
Or65abc	0.01435	0.661	0.612	0.571	0.571	0.595	0.06	0.107	**0.036**	**0.035**	0.52	0.5	0.5	0.5	0.5	**0.043**	0.554	0.505	**0.025**	0.52	0.518	0.577	**0.035**
Or98a85a	0.012	0.661	0.612	0.571	**0.028**	0.595	0.719	0.612	0.618	0.114	0.077	0.61	0.156	**0.017**	0.686	0.734	0.671	0.52	**0.039**	0.558	0.551	**0.04**	0.687
Or98b47a	0.08915	0.618	0.579	0.549	0.549	0.567	0.666	0.056	0.544	**0.035**	0.52	0.541	0.541	0.058	0.56	0.583	0.072	0.505	0.56	0.52	0.518	**0.035**	0.577
ppk10	0.0844	0.511	0.507	0.504	0.504	0.506	0.517	0.507	0.507	0.513	0.503	0.509	0.509	0.518	0.527	0.538	**0.002**	0.501	0.51	0.503	0.503	0.513	0.513
ppk29	0.0563	0.511	0.507	0.504	0.504	0.506	**0.021**	0.507	0.507	0.513	0.503	0.509	0.509	0.518	0.527	0.538	0.509	0.501	0.51	0.503	0.503	0.513	**0.016**
ppk8	0.00815	0.511	0.507	0.504	0.504	0.506	0.517	0.507	0.507	**0.016**	0.503	0.5	0.5	0.5	0.5	0.056	0.509	0.501	0.51	0.503	0.503	0.513	**0.016**
trp01	0.01	0.517	0.511	0.506	0.506	0.509	0.526	0.511	0.511	0.521	0.505	**0.003**	0.502	0.504	**0.009**	0.508	0.514	0.501	0.516	0.505	0.504	0.521	0.521
trp02	0.0746	0.517	0.511	0.506	0.506	0.509	0.526	0.511	0.511	0.521	0.505	**0.003**	0.502	0.504	0.506	0.508	0.514	0.501	0.516	0.505	0.504	0.521	0.521
trp06	0.07137 5	0.517	0.511	0.506	0.506	0.509	**0**	0.511	0.511	0.521	0.505	0.502	0.502	0.504	0.506	0.508	0.514	0.501	0.516	0.505	0.504	0.521	0.521
cluster7	0.00015	0.758	0.697	0.635	0.083	0.095	**0.008**	0.674	0.681	0.204	**0.003**	0.525	0.525	0.548	**0.001**	0.595	0.645	0.517	0.66	0.597	**0.049**	0.784	**0.018**
cluster8	0.0354	0.705	0.649	0.597	0.597	0.628	0.086	0.649	0.237	**0**	0.562	0.525	0.525	0.548	0.569	0.595	0.645	0.517	0.66	0.562	0.555	0.296	0.694
cluster16	0.01185	0.669	0.619	0.576	0.576	0.601	**0.023**	0.619	0.624	**0**	0.094	0.517	0.517	0.533	0.548	0.567	0.605	0.511	0.617	0.562	0.555	0.694	0.694
cluster30	0.0169	0.569	0.545	0.527	0.527	0.538	0.602	0.545	0.548	0.583	**0**	0.533	**0.017**	0.562	0.589	0.621	0.679	0.522	0.695	0.522	0.519	0.583	0.583
cluster32	0	**0**	0.545	0.527	0.527	0.538	0.602	0.545	0.548	0.583	0.522	0.508	0.508	0.517	0.525	0.535	0.558	0.506	0.565	0.522	0.519	0.583	0.583
cluster34	0.07575	0.569	0.545	0.527	0.527	0.538	0.602	0.545	0.548	0.583	**0**	0.533	0.533	0.562	0.589	0.621	0.679	0.522	0.695	0.522	0.519	0.583	0.583
cluster36	0.07575	0.569	0.545	0.527	0.527	0.538	0.602	0.545	0.548	0.583	**0**	0.533	0.533	0.562	0.589	0.621	0.679	0.522	0.695	0.522	0.519	0.583	0.583
cluster58	0.07435	0.569	0.545	0.527	0.527	0.538	0.602	0.545	0.548	0.583	**0**	0.525	**0.013**	0.548	0.569	0.595	0.645	0.517	0.66	0.522	0.519	0.583	0.583
cluster87	0.06	0.072	0.585	0.553	**0.028**	0.571	0.261	0.585	0.589	0.086	0.543	0.508	0.508	**0.049**	0.548	0.567	0.605	0.511	0.617	0.543	0.057	0.583	0.583
cluster104	0.0063	0.187	0.585	0.553	**0.028**	0.571	0.261	0.585	0.589	**0.017**	0.064	0.508	0.508	0.517	0.525	0.535	0.558	0.506	0.565	0.543	0.538	0.219	**0.017**
cluster105	0.06555	0.569	0.545	0.527	0.527	0.107	0.674	0.585	0.589	0.219	0.543	0.517	**0.009**	0.533	0.072	0.567	0.059	0.511	0.617	0.543	0.538	0.646	0.646
cluster122	0.0771	0.625	0.585	0.079	0.553	0.571	0.107	0.585	0.589	**0.017**	0.064	0.508	0.508	0.517	0.525	0.535	0.558	0.506	0.565	0.543	0.538	0.219	0.646
cluster140	0.0201	0.569	0.545	0.527	0.527	0.538	0.602	0.545	0.548	0.583	**0**	**0.025**	0.517	0.072	0.569	0.595	0.645	0.517	0.66	0.522	0.519	0.583	0.583
cluster161	0.03065	0.569	0.545	0.527	**0**	0.538	0.602	0.127	0.589	**0.017**	0.064	0.508	0.508	0.517	0.525	0.535	0.558	0.506	0.565	0.543	0.538	0.646	0.646
cluster171	0.06885	0.569	0.545	0.527	0.527	0.538	**0**	0.545	0.548	0.583	0.522	0.508	0.508	0.517	0.525	0.535	0.558	0.506	0.565	0.522	0.519	0.583	0.583

Table S5. Gene counts of all genes that have been duplicated or lost in all herbivorous *Scaptomyza*.

	Gene	*Dmel*	*Dere*	*Dana*	*Dpse*	*Dvir*	*Dmoj*	*Dgri*	*Spal*	*Shsu*	*Sgra*	*Smon*	*Sfla*	
Gr	*Gr39aA*	1	0	1	3	6	5	4	9	9	8	4	6	loss
	Gr39aE	1	1	1	2	0	1	1	1	1	0	0	0	loss
	Gr59ab	2	3	4	4	2	1	4	4	3	2	2	2	loss
	Gr59cd	2	2	2	2	4	3	7	9	7	3	4	4	loss
	Gr68a	1	1	1	1	1	1	1	1	1	0	0	0	loss
Gst	*GstS1*	1	1	1	1	1	1	1	1	1	2	2	2	**gain**
Ir	*Ir47abc94abc*	4	3	6	3	5	4	2	1	3	1	1	1	loss
	Ir51be	1	2	2	2	1	1	1	1	1	0	0	0	loss
	Ir56e	0	0	0	0	1	1	1	2	2	1	1	1	loss
	Ir60e	1	1	1	1	0	1	1	1	1	0	0	0	loss
	Ir67a	0	1	1	1	1	1	1	1	1	2	2	3	**gain**
	Ir7f	1	1	1	1	1	1	1	1	1	0	0	0	loss
	Ir94f	1	1	1	0	1	1	1	1	1	0	0	0	loss
Obp	*Obp18a*	1	0	1	0	0	1	1	1	1	0	0	0	loss
	Obp46a	1	1	1	1	1	1	1	1	1	0	0	0	loss
	Obp50cd	2	2	2	2	1	1	1	1	1	0	0	0	loss
	Obp56b	1	1	1	1	1	1	1	1	1	0	0	0	loss
	Obp58b	1	1	1	1	1	1	4	1	1	0	0	0	loss
	Obp58c	1	1	1	1	1	1	5	1	1	0	0	0	loss
	Obp58d	1	1	1	1	1	1	1	1	1	0	0	0	loss
	Obp93a	1	1	1	1	1	1	1	1	1	0	0	0	loss
Or	*Or22a*	2	1	5	2	1	1	2	2	1	0	0	0	loss
P450	*Cyp4ad1*	1	1	1	1	1	1	1	1	1	0	0	0	loss
	Cyp4d1	1	1	2	1	1	2	1	1	1	0	0	0	loss
Ppk	*ppk8*	1	1	1	1	0	1	0	1	1	0	0	0	loss
Ugt	*Ugt302e1*	1	1	0	1	1	1	2	1	1	0	0	0	loss

Table S6. PAML analyses – full results of significant models.

Gene	Model	κ	tree length	dN/dS				lnL	LRT	*q*-value (FDR 5%)
Csp2	M0 (one ratio)	1.4	5.3	ω= 0.07				-1897.9		
	Branch (two-ratios)	1.4	5.5	ω0= 0.06	ω1= 0.24			-1894.3	7.14	**0.05**
Gr59e	M0 (one ratio)	2.1	6.7	ω= 0.26				-6683.9		
	Branch (two-ratios)	2.1	7	ω0= 0.23	ω1= 999			-6674.7	18.26	**<0.001**
Gr63a	Branch-site Model A	1.8	3.6	ω0= 0.03 (81%)	ω1= 1 (7%)	ω2a= 1 (11%)	ω2b= 1 (1%)	-6110.5		
	Branch-site Model A (ω = 1)	1.8	3.6	ω0= 0.03 (0%)	ω1= 1 (0%)	ω2a= 1 (92%)	ω2b= 1 (8%)	-6118.7	16.52	**<0.01**
Gr98a	M0 (one ratio)	1.9	16.3	ω= 0.29				-19802.4		
(SmonGr98a1, SflaGr98a1, SgraGr98a2)	Branch (two-ratios)	1.9	16.3	ω0= 0.29	ω1= 0.87			-19798.5	7.74	**0.04**
	Branch-site Model A	2	17	ω0= 0.22 (71%)	ω1= 1 (26%)	ω2a= 32.44 (2%)	ω2b= 32.44 (1%)	-19593		
	Branch-site Model A (ω = 1)	2	16.9	ω0= 0.22 (65%)	ω1= 1 (24%)	ω2a= 1 (8%)	ω2b= 1 (3%)	-19598.7	11.4	**0.03**
Gr98bcd	Branch-site Model A	2.1	23.6	ω0= 0.16 (73%)	ω1= 1 (27%)	ω2a= 179.55 (0.004%)	ω2b= 179.55 (0.001%)	-12238.2		
	Branch-site Model A (ω = 1)	2.1	23.5	ω0= 0.16 (73%)	ω1= 1 (27%)	ω2a= 1 (0%)	ω2b= 1 (0%)	-12243.3	10.2	**0.05**
GstE9	M0 (one ratio)	1.6	4.1	ω= 0.14				-3640.7		
	Branch (two-ratios)	1.6	4.1	ω0= 0.15	ω1= 0.01			-3636	9.35	**0.02**
GstO2	M0 (one ratio)	1.8	5.5	ω= 0.11				-4112.2		
	Branch (two-ratios)	1.8	5.4	ω0= 0.12	ω1= 0			-4107.1	10.3	**0.01**
GstS1	M0 (one ratio)	2	1.9	ω= 0.11				-2833		
(SflaGstS1b, SmonGstS1b, SgraGstS1b)	Branch (two-ratios)	2	1.9	ω0= 0.09	ω1= 3.27			-2807.8	50.43	<0.0001
	Branch-site Model A	2.1	1.9	ω0= 0.08 (84%)	ω1= 1 (2%)	ω2a= 38.47 (13%)	ω2b= 38.47 (0.004%)	-2800.91		
	Branch-site Model A (ω = 1)	2	1.9	ω0= 0.07 (0%)	ω1= 1 (0%)	ω2a= 1 (97%)	ω2b= 1 (3%)	-2807.93	14.04	**0.01**
Ir21a	M0 (one ratio)	1.5	4.7	ω= 0.11				-13380.2		
	Branch (two-ratios)	1.5	4.8	ω0= 0.11	ω1= 0.26			-13375.6	9.2	**0.02**
Ir48d	M0 (one ratio)	1.9	3.5	ω= 0.19				-8419.1		
	Branch (two-ratios)	1.9	3.5	ω0= 0.18	ω1= 0.38			-8415.2	7.96	**0.04**
Ir56a	M0 (one ratio)	1.8	15	ω= 0.34				-23071.9		
(SflaIr56a, SmonIr56a)	Branch (two-ratios)	1.8	15.1	ω0= 0.33	ω1= 0.98			-23062.5	18.8	<0.001
(SgraIr56a1, SgraIr56a2, SgraIr56a4, SgraIr56a5)	Branch (two-ratios)	1.8	15	ω0= 0.33	ω1= 3.66			-23067.9	7.98	**0.03**
	Branch (two-ratios, fixed ω1 = 1)	1.8	15	ω0= 0.34	ω1= 1			-23068.7	1.56	0.31
Ir60a	Branch-site Model A	1.8	7.1	ω0= 0.06 (83%)	ω1= 1 (16%)	ω2a= 999 (0.005%)	ω2b= 999 (0.001%)	-10580.9		
	Branch-site Model A (ω = 1)	1.8	5.4	ω0= 0.06 (82%)	ω1= 1 (16%)	ω2a= 1 (2%)	ω2b= 1 (0.003%)	-10585.9	10	**0.05**
Ir67a	M0 (one ratio)	1.8	4	ω= 0.31				-10318.5		

	Branch (two-ratios)	1.8	4	ω0= 0.3	ω1= 0.6					-10313.2	10.6	**0.01**
Obp57cL1	Branch-site Model A	2.1	9.3	ω0= 0.21 (59%)	ω1= 1 (36%)	ω2a= 148.51 (3%)	ω2b= 148.51 (2%)			-2539.7		
	Branch-site Model A (ω = 1)	2	7.3	ω0= 0.21 (55%)	ω1= 1 (30%)	ω2a= 1 (10%)	ω2b= 1 (5%)			-2547.5	15.68	**<0.01**
Or19a	M0 (one ratio)	1.7	7.6	ω= 0.27						-7966.9		
	Branch (two-ratios)	1.7	7.5	ω0= 0.3	ω1= 0.09					-7960.1	13.7	**<0.01**
Or22c	M0 (one ratio)	1.6	3.7	ω= 0.17						-5823.7		
	Branch (two-ratios)	1.6	3.7	ω0= 0.18	ω1= 0.05					-5820.1	7.24	**0.05**
Or42a	M0 (one ratio)	1.7	7.5	ω= 0.14						-10294.1		
(SflaOr42a1, SmonOr42a1, SgraOr42a1)	Branch (two-ratios)	1.7	7.6	ω0= 0.13	ω1= 0.45					-10288.4	11.46	**<0.01**
Or56a	M0 (one ratio)	1.5	4.3	ω= 0.1						-6120.1		
	Branch (two-ratios)	1.5	4.3	ω0= 0.11	ω1= 0.03					-6115.6	8.98	**0.02**
Or63a	M0 (one ratio)	1.6	4.7	ω= 0.22						-7544.9		
	Branch (two-ratios)	1.6	4.8	ω0= 0.2	ω1= 0.76					-7530.2	29.5	**<0.0001**
Or67d	M0 (one ratio)	1.9	4.7	ω= 0.16						-4995.1		
	Branch (two-ratios)	1.9	4.5	ω0= 0.18	ω1= 0.03					-4987.5	15.18	**<0.01**
Or85aLike	M0 (one ratio)	1.5	3.9	ω= 0.14						-5773.6		
	Branch (two-ratios)	1.5	4	ω0= 0.13	ω1= 0.29					-5770.1	6.98	**0.05**
Or85f	Branch-site Model A	2.1	5.5	ω0= 0.13 (88%)	ω1= 1 (11%)	ω2a= 46.61 (1%)	ω2b= 46.61 (0.001%)			-5725.5		
	Branch-site Model A (ω = 1)	2.1	5.4	ω0= 0.13 (85%)	ω1= 1 (10%)	ω2a= 1 (5%)	ω2b= 1 (1%)			-5731.8	12.54	**0.03**
Or98aLike1	M0 (one ratio)	1.6	2.9	ω= 0.18						-5564.4		
	Branch (two-ratios)	1.6	2.9	ω0= 0.19	ω1= 0.05					-5560.6	7.58	**0.04**
Or98aLike2	M0 (one ratio)	1.9	4.3	ω= 0.18						-5970.6		
	Branch (two-ratios)	1.9	4.2	ω0= 0.2	ω1= 0					-5952.7	35.92	**<0.0001**
OrN2.3prime	Branch-site Model A	1.9	9.2	ω0= 0.16 (73%)	ω1= 1 (20%)	ω2a= 15.3 (5%)	ω2b= 15.3 (1%)			-11146.2		
	Branch-site Model A (ω = 1)	1.9	9.2	ω0= 0.16 (60%)	ω1= 1 (17%)	ω2a= 1 (18%)	ω2b= 1 (5%)			-11154	15.6	**<0.01**
Cyp28a5	M0 (one ratio)	1.6	4.6	ω= 0.12						-8110		
	Branch (two-ratios)	1.6	4.5	ω0= 0.13	ω1= 0.05					-8105.1	9.95	**0.02**
Cyp310a1	M0 (one ratio)	1.9	4.4	ω= 0.21						-8063.8		
	Branch (two-ratios)	1.9	4.4	ω0= 0.19	ω1= 0.57					-8054.6	18.43	**<0.001**
Cyp4d14	Branch-site Model A	1.8	7.2	ω0= 0.06 (84%)	ω1= 1 (15%)	ω2a= 809.22 (1%)	ω2b= 809.22 (0.001%)			-7537.6		
	Branch-site Model A (ω = 1)	1.8	5.3	ω0= 0.06 (81%)	ω1= 1 (15%)	ω2a= 1 (4%)	ω2b= 1 (1%)			-7543.3	11.44	**0.03**
Cyp6a16	M0 (one ratio)	1.7	3.5	ω= 0.2						-8081		
	Branch (two-ratios)	1.7	3.5	ω0= 0.21	ω1= 0.1					-8077.2	7.57	**0.04**

Cyp6a22	M0 (one ratio)	1.6	3.4	ω= 0.09					-5558.1		
	Branch (two-ratios)	1.6	3.5	ω0= 0.07	ω1= 0.26				-5547.5	21.17	**<0.001**
Cyp6u1	M0 (one ratio)	1.5	5.2	ω= 0.15					-8378.5		
	Branch (two-ratios)	1.6	5.3	ω0= 0.14	ω1= 0.61				-8368.5	20.06	**<0.001**
Ppk10	M0 (one ratio)	1.5	5.4	ω= 0.1					-7191.5		
	Branch (two-ratios)	1.5	5.3	ω0= 0.1	ω1= 0.01				-7185.4	12.12	**<0.01**
Ppk12	M0 (one ratio)	1.5	5.6	ω= 0.19					-10244.3		
	Branch (two-ratios)	1.5	5.6	ω0= 0.18	ω1= 0.38				-10239.3	10.07	**0.02**
Ppk13	M0 (one ratio)	1.5	3.2	ω= 0.06					-5756.3		
	Branch (two-ratios)	1.5	3.3	ω0= 0.05	ω1= 0.19				-5748.3	16.12	**<0.01**
Ppk19	Branch-site Model A	1.7	6.1	ω0= 0.11 (75%)	ω1= 1 (23%)	ω2a= 15.75 (2%)	ω2b= 15.75 (1%)		-8909.7		
	Branch-site Model A (ω = 1)	1.7	6	ω0= 0.11 (72%)	ω1= 1 (23%)	ω2a= 1 (4%)	ω2b= 1 (1%)		-8916.7	13.98	**0.01**
Ppk22	M0 (one ratio)	1.7	4.5	ω= 0.16					-9151.1		
	Branch (two-ratios)	1.7	4.5	ω0= 0.16	ω1= 0.37				-9147.3	7.7	**0.04**
Ppk25	M0 (one ratio)	1.5	6.1	ω= 0.2					-8564.4		
	Branch (two-ratios)	1.5	6	ω0= 0.21	ω1= 0.04				-8556	16.82	**<0.01**
Ppk28	M0 (one ratio)	1.5	3.4	ω= 0.1					-7869.2		
	Branch (two-ratios)	1.5	3.3	ω0= 0.1	ω1= 0.03				-7863.4	11.72	**<0.01**
Ppk30	M0 (one ratio)	1.5	6.1	ω= 0.22					-8365		
	Branch (two-ratios)	1.5	6.3	ω0= 0.2	ω1= 0.44				-8358.9	12.2	**<0.01**
Ppk31	M0 (one ratio)	1.1	4	ω= 0.13					-6663		
	Branch (two-ratios)	1.1	4	ω0= 0.12	ω1= 0.39				-6654.4	17.23	**<0.01**
Ppk5	M0 (one ratio)	1.5	5.4	ω= 0.12					-8397.9		
	Branch (two-ratios)	1.5	5.5	ω0= 0.11	ω1= 0.36				-8390.3	15.16	**<0.01**
Ppk6	M0 (one ratio)	1.9	3.2	ω= 0.1					-6550		
	Branch (two-ratios)	1.9	3.2	ω0= 0.09	ω1= 0.25				-6544.9	10.35	**0.01**
Ppk7	M0 (one ratio)	1.5	3.5	ω= 0.16					-8885.9		
	Branch (two-ratios)	1.5	3.5	ω0= 0.15	ω1= 0.27				-8882	7.75	**0.04**
Trp04	M0 (one ratio)	1.6	2.7	ω= 0.03					-8768.9		
	Branch (two-ratios)	1.6	2.7	ω0= 0.03	ω1= 0.09				-8763.6	10.66	**0.01**
	Branch-site Model A	1.8	3.6	ω0= 0.01 (94%)	ω1= 1 (6%)	ω2a= 999 (0.004%)	ω2b= 999 (0.0002%)		-8618		
	Branch-site Model A (ω = 1)	1.8	2.9	ω0= 0.01 (92%)	ω1= 1 (6%)	ω2a= 1 (3%)	ω2b= 1 (0.002%)		-8624.1	12.2	**0.03**
Trp12	Branch-site Model A	1.9	2.1	ω0= 0.01 (97%)	ω1= 1 (3%)	ω2a= 999 (0.001%)	ω2b= 999 (0.00003%)		-16143.7		
	Branch-site Model A (ω = 1)	1.9	1.9	ω0= 0.01 (96%)	ω1= 1 (3%)	ω2a= 1 (0.004%)	ω2b= 1 (0.0001%)		-16149.6	11.8	**0.03**
Ugt301D1	M0 (one ratio)	1.8	4.2	ω= 0.08					-7343.2		
	Branch (two-ratios)	1.8	4.1	ω0= 0.09	ω1= 0.03				-7339.6	7.22	**0.05**

Ugt302C1	M0 (one ratio)	1.8	3.9	ω= 0.1			-7695.6		
	Branch (two-ratios)	1.8	3.9	ω0= 0.11	ω1= 0.04		-7691.9	7.42	**0.05**

Table S7. Convergent evolution: loss of Gr39aA paralogs in specialist drosophild species.

		Gr39a Exons							
Group	**Species**	**A**	**B**	**C**	**D**	**E**	**Specialist or Generalist**	**Diet**	**Diet Ref.**
Hawaiian *Drosophila*	*D. grimshawi*	4	1	0	1	1	generalist	bark/stem/rachis oligophagous/polyphagous	a
	D. murphyi	5	1	0	1	1	generalist	bark/stem/rachis	a
	D. sproati	2	1	0	1	0	specialist	monophagous on Araliaceae, bark/stem/rachis	a
Lordiphosa	*L. mommai*	3	1	1	1	1	generalist	decaying leaves	b
	L. magnipectinata	1	1	0	1	1	specialist	leaf-mining	c
	L. clarofinis	2	1	0	1	2	specialist	leaf-mining	d
	L. collinella	2	1	0	1	1	specialist	leaf-mining	b
repleta	*D. repleta*	6	1	0	1	1	generalist	cactus, fruit	e
	D. hydei	5	1	0	1	1	generalist	necrotic fruit and feces	f
	D. arizonae	7	1	0	1	1	specialist	columnar & Opuntia cacti	g
	D. navojoa	6	1	0	1	1	specialist	Opuntia specialist	g
	D. mojavensis	5	1	0	1	1	specialist	cactus	h
Scaptomyza	*S. pallida*	8	1	0	1	1	generalist	decaying leaves and fruits	i
	S. hsui	9	1	0	1	1	generalist	decaying leaves	j
	S. montana	4	1	0	1	0	specialist	living leaves, Brassicaceae	
	S. flava	6	1	0	1	0	specialist	living leaves, Brassicaceae	
	S. graminum	7	1	0	1	0	specialist	living leaves, Caryophyllaceae	
melanogaster	*D. mauritiana*	1	1	1	1	0	generalist	fruit, fungus	k
	D. melanogaster	1	1	1	1	0	generalist	fruits, flowers	k
	D. simulans	1	1	1	1	0	generalist	fruit, fungus	k
	D. teissieri - 273.3	1	1	1	1	0	generalist	fruit, fungus	k
	D. yakuba	1	1	1	1	0	generalist	fruit, fungus	k
	D. sechellia	0	1	1	1	0	specialist	Morinda citrifolia	k
	D. erecta	0	1	1	1	0	specialist	Pandanus spp. fruit	k

[a] Magnacca et al. 2008
[b] Kimura 1979 thesis
[c] Attallah et al. 2012 Evolution & Development
[d] Gao et al. 2022 biorxiv
[e] Goni et al. 2012 Zoologia
[f] O'Grady & Markow 2011
[g] Ruiz et al. 1990 Journal of Heredity
[h] Koerte et al. 2020
[i] Kimura 1976
[j] Lapoint et al. 2013
[k] Markow & O'Grady 200

Table S8. Reported host plants of herbivorous *Scaptomyza* species

Species	Plant Family	Species	References
Included in Fig. 1b			
S. montana (California, USA)	Brassicaceae	*Nasturtium officinale, Raphanus raphanistrum*	this study
S. flava (California, USA)	Brassicaceae	*Brassica oleracea, Raphanus raphanistrum, Sisymbrium officinale*	this study
S. sp. (California, USA)	Brassicaceae	*Barbarea orthoceras, Raphanus raphanistrum, Sisymbrium officinale*	this study
S. nr. *nigrita* (California, USA)	Brassicaceae	*Barbarea orthoceras, Brassica oleracea, Cardamine californica, Cardamine oligosperma, Nasturtium officinale, Sisymbrium officinale, Turritis glabra*	this study
Other reported species/populations			
S. montana (Arizona, USA)	Brassicaceae	strongly attracted by *Arabidopsis thaliana* bearing indole glucosinolates	Gloss et al. 2017
S. flava (Massachusetts, USA)	Brassicaceae	indifferent towards *Arabidopsis thaliana* bearing indole glucosinolates	Gloss et al. 2017
S. nigrita (Colorado, USA)	Brassicaceae	*Cardamine cordifolia (monophagous)*	Collinge 1989
S. *flava* (Canton Jura, Switzerland)	Brassicaceae	*Brassica napa*	Haye et al. 2015
S. flava (Europe)	Brassicaceae	*Brassica oleracea, Brassica napus, Brassica rapa, Raphanus sativus*	Maca 1972
S. flava (Europe)	Asteraceae		Maca 1972
S. flava (Europe)	Capparidaceae		Maca 1972
S. flava (Europe)	Resedaceae		Maca 1972
S. flava (Europe)	Tropaeolaceae		Maca 1972
S. flava (Europe)	Viciaceae		Maca 1972
S. flava (Europe)	Fabaceae	*Pisum sativum*	Maca 1972
S. flava (Europe)	Brassicaceae	*Capsella bursapastoris, Raphanus raphanistrum*	Godfray 1984
S. flava (Europe)	Asteraceae	*Sonchus asper*	Godfray 1984
S. flava (New Zealand)	Brassicaceae	*Brassica rapa, Brassica campestris spp. pekinensis*	Seraj & Fenemore 1993
S. flava (New Zealand)	Brassicaceae	*Brassica juncea, Eruca vesicaria*	MAF PPIN database
S. flava (New Zealand)	Fabaceae	*Phaseolus coccineus*	MAF PPIN database
S. flava (New Zealand)	Caryophyllaceae	*Gypsophila* spp.	MAF PPIN database
S. *flava* (New Zealand)	Brassicaceae	*Cardamine debile agg., Cardamine flexuosa, Cardamine hirsuta, Rorippa nasturtium-aquaticum, Rorripa divaricata, Cakile edulata, Coronopus didymus, Hesperis matronalis, Raphanus raphanistrum, Raphanus maritumus, Sisymbrium officinale*	Marin 2004
S. *flava* (New Zealand)	Fabaceae	*Pisum sativum*	Marin 2004
S. *flava* (New Zealand)	Tropaeolaceae	*Tropaeolum majus*	Marin 2004
S. *flava* (New Zealand)	Solanaceae	*Solanum tuberosum*	Marin 2004
S. *flava* (New Zealand)	Liliaceae	*Allium cepa*	Martin 2004

S. flava (Victoria, Australia)	Brassicaceae	*Brassica rapa var. pekinensis, Raphanus raphanistrum*	Bjorksten et al. 2005
S. flava (Saskatchewan, Canada)	Brassicaceae	*Brassica napus*	Andreassen et al. 2015

* MAF PPIN: New Zealand Ministry of Agriculture and Forestry Plant Pest Information Network

Appendix F: Supplemental figures from Chapter 2.

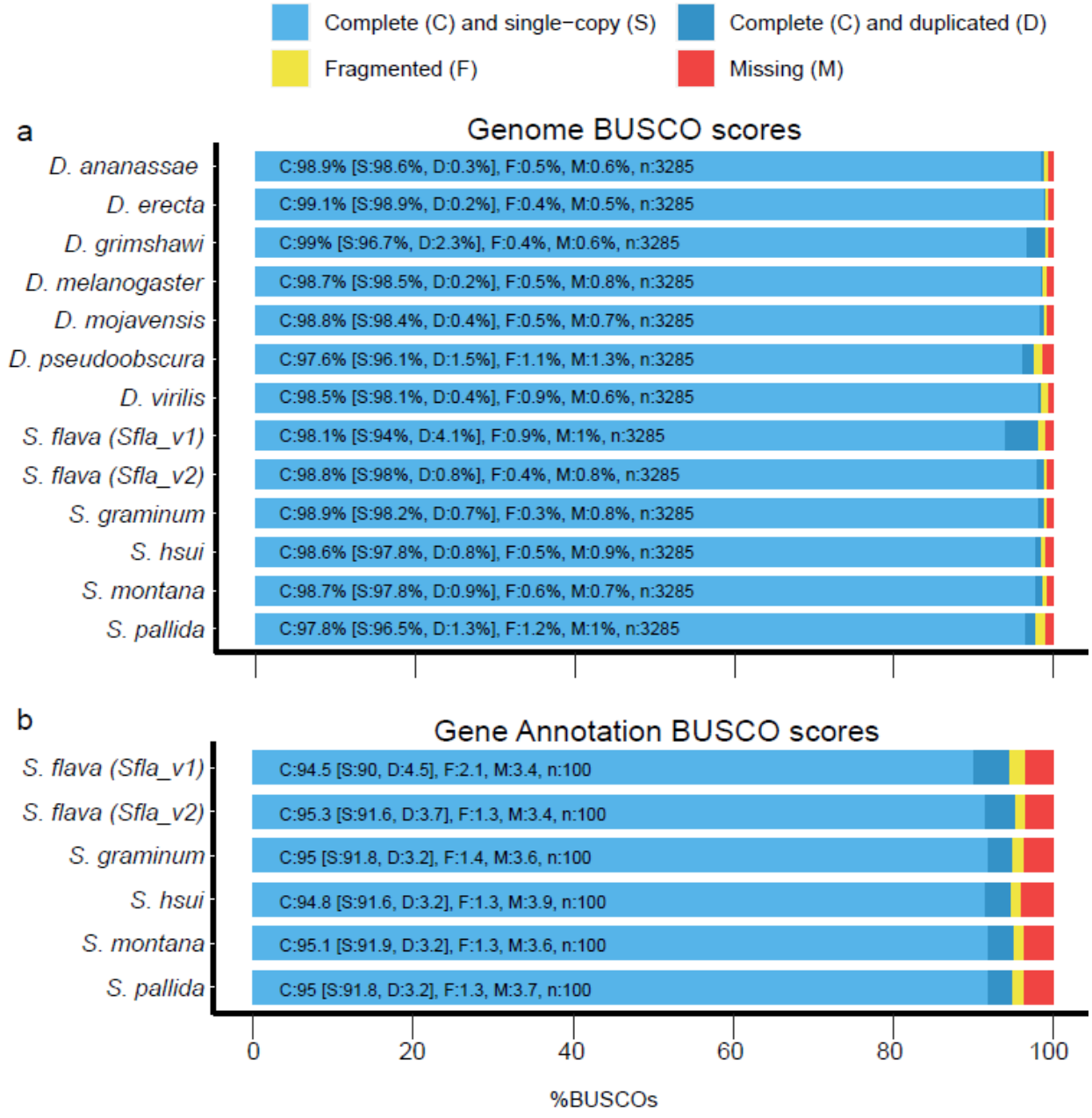


Figure S1. Benchmarking with core dipteran genes
High levels of completeness are found in (a) the genome assemblies and (b) gene annotations used to curate the chemosensory and detoxification gene families, estimated using BUSCO v.5.4.2 (diptera_odb10, n=3285 genes).

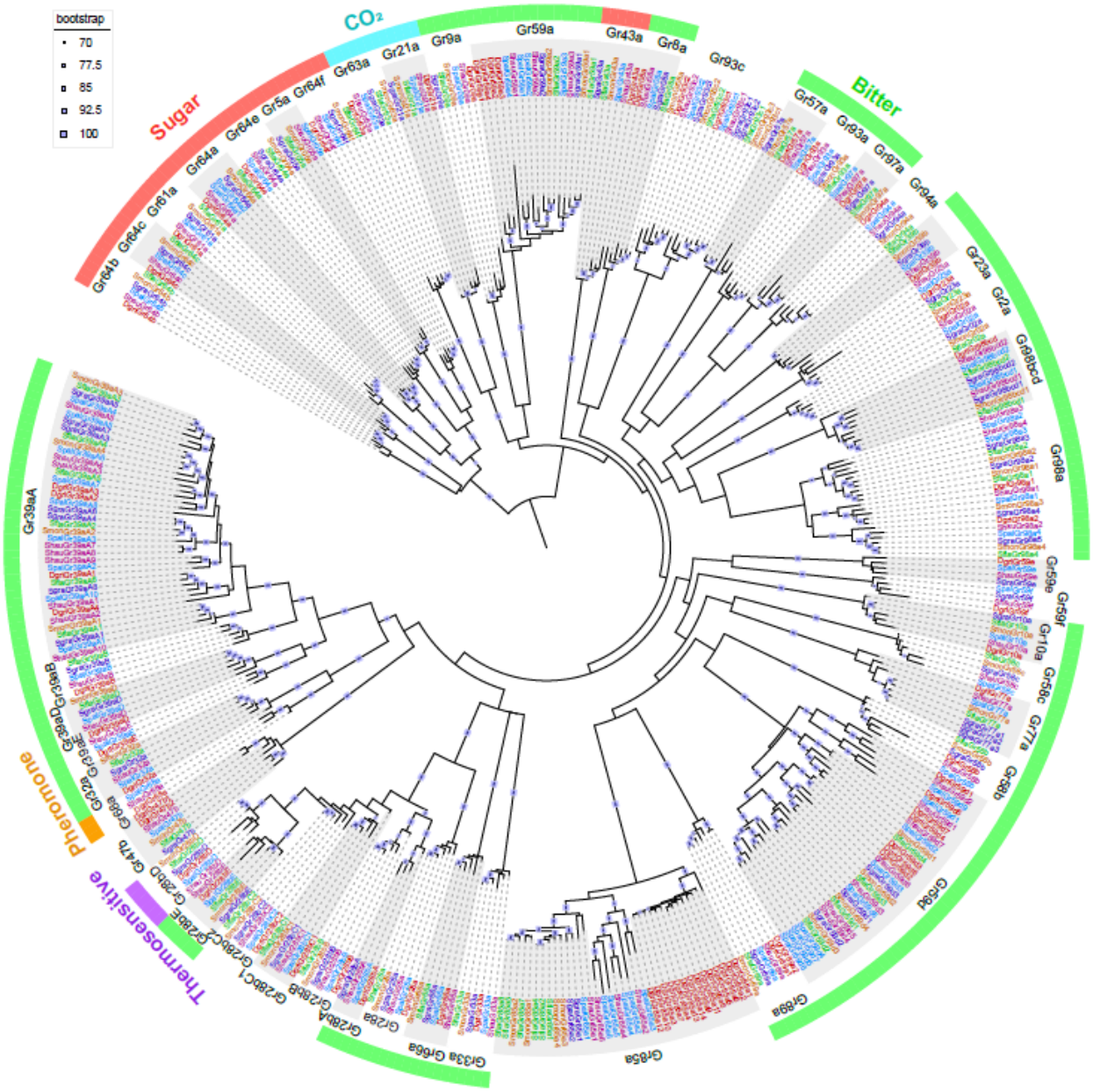


Figure S2. Phylogeny of gustatory receptors.

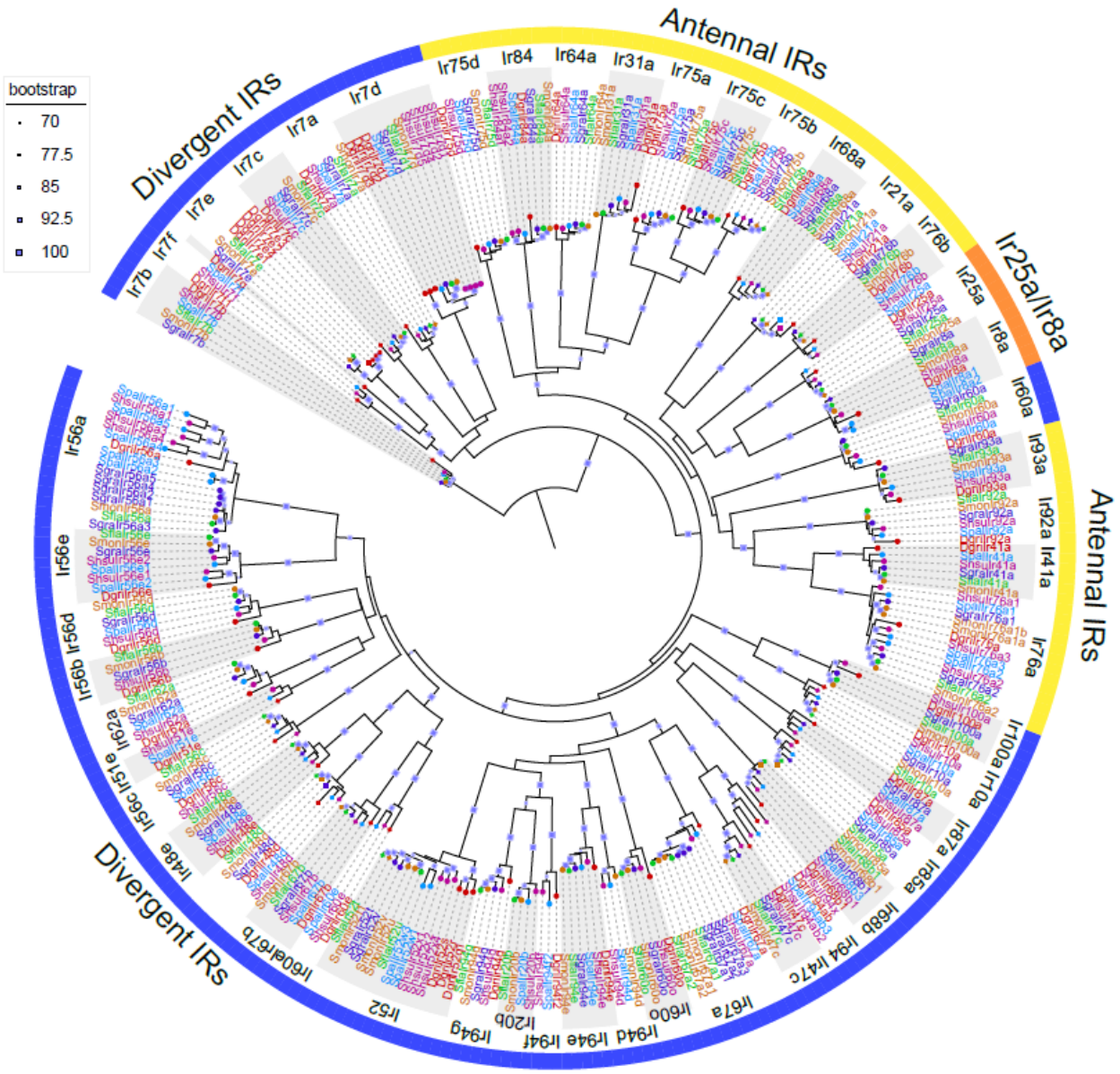


Figure S3. Phylogeny of ionotropic receptors.

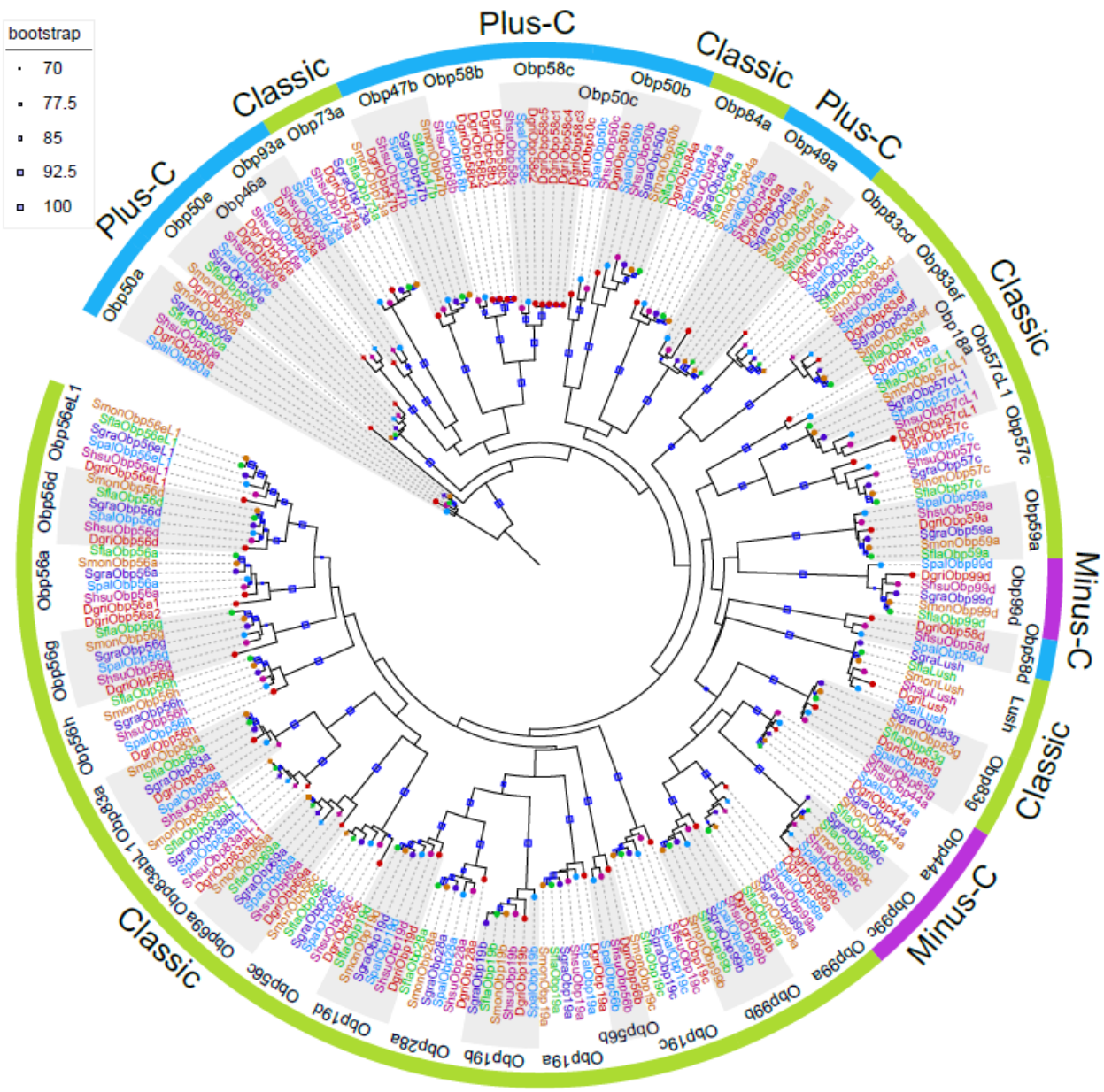


Figure S4. Phylogeny of odorant binding receptors.

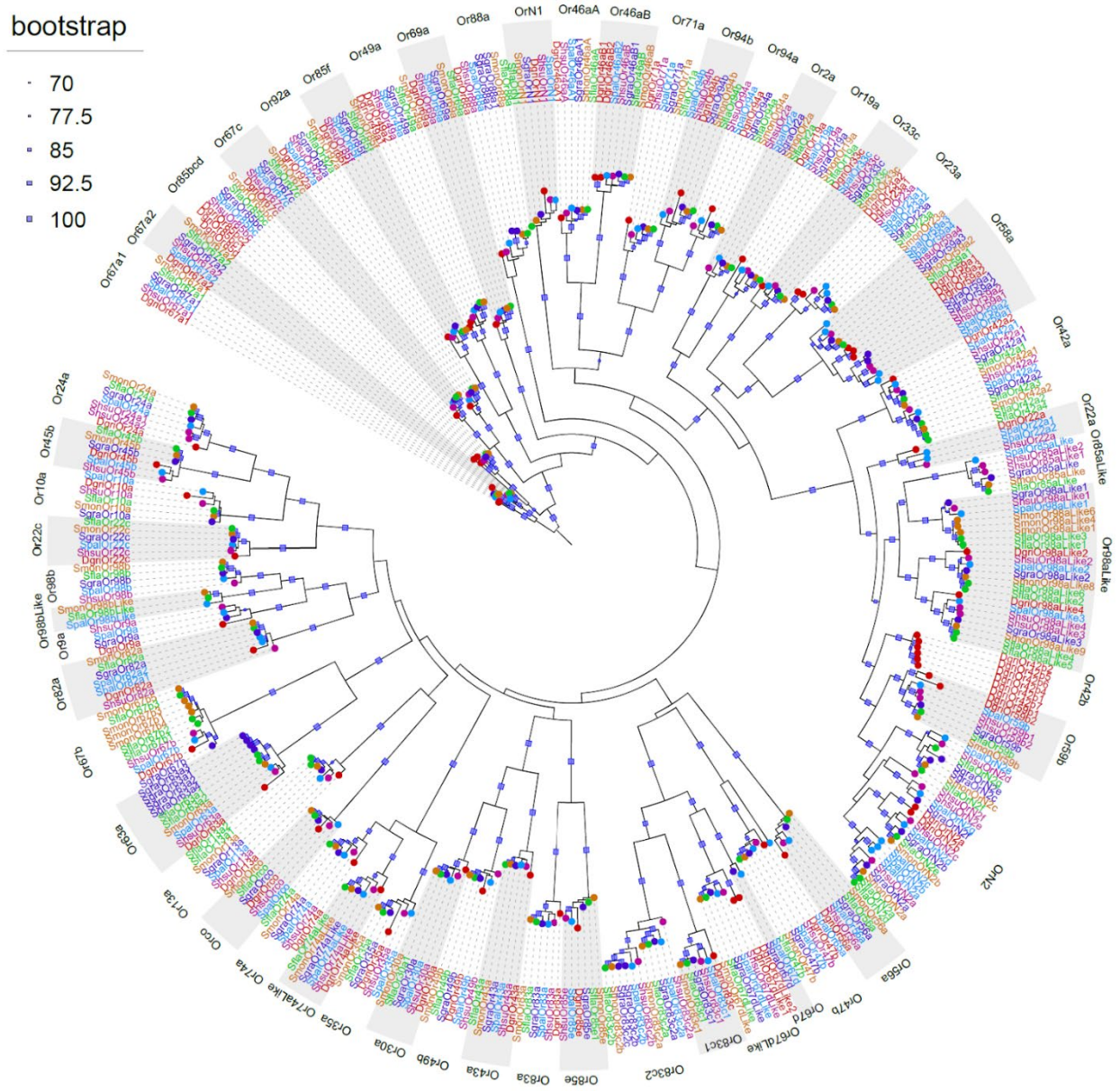


Figure S5. Phylogeny of olfactory receptors.

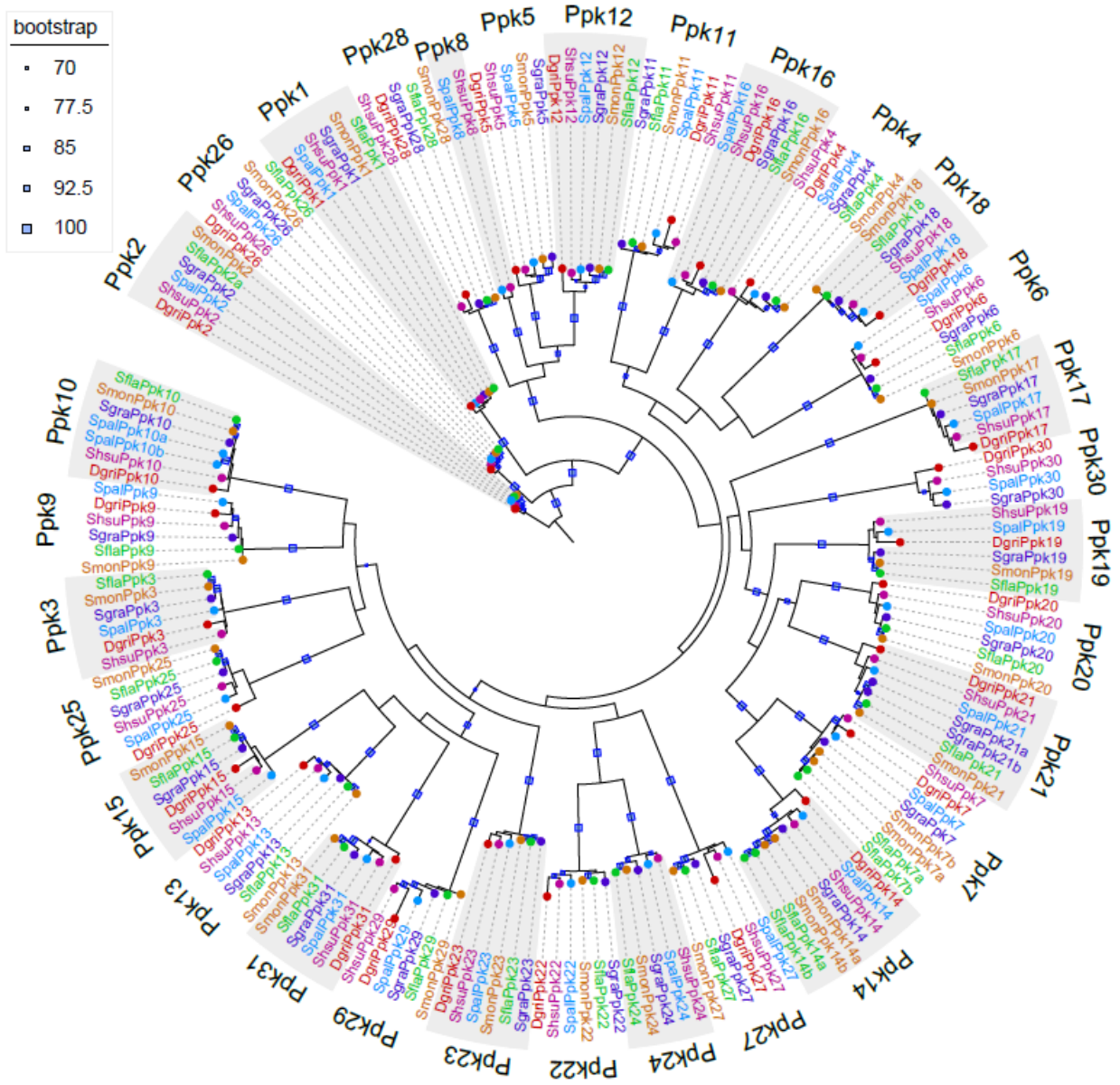


Figure S6. Phylogeny of degenerin/epithelial sodium channels.

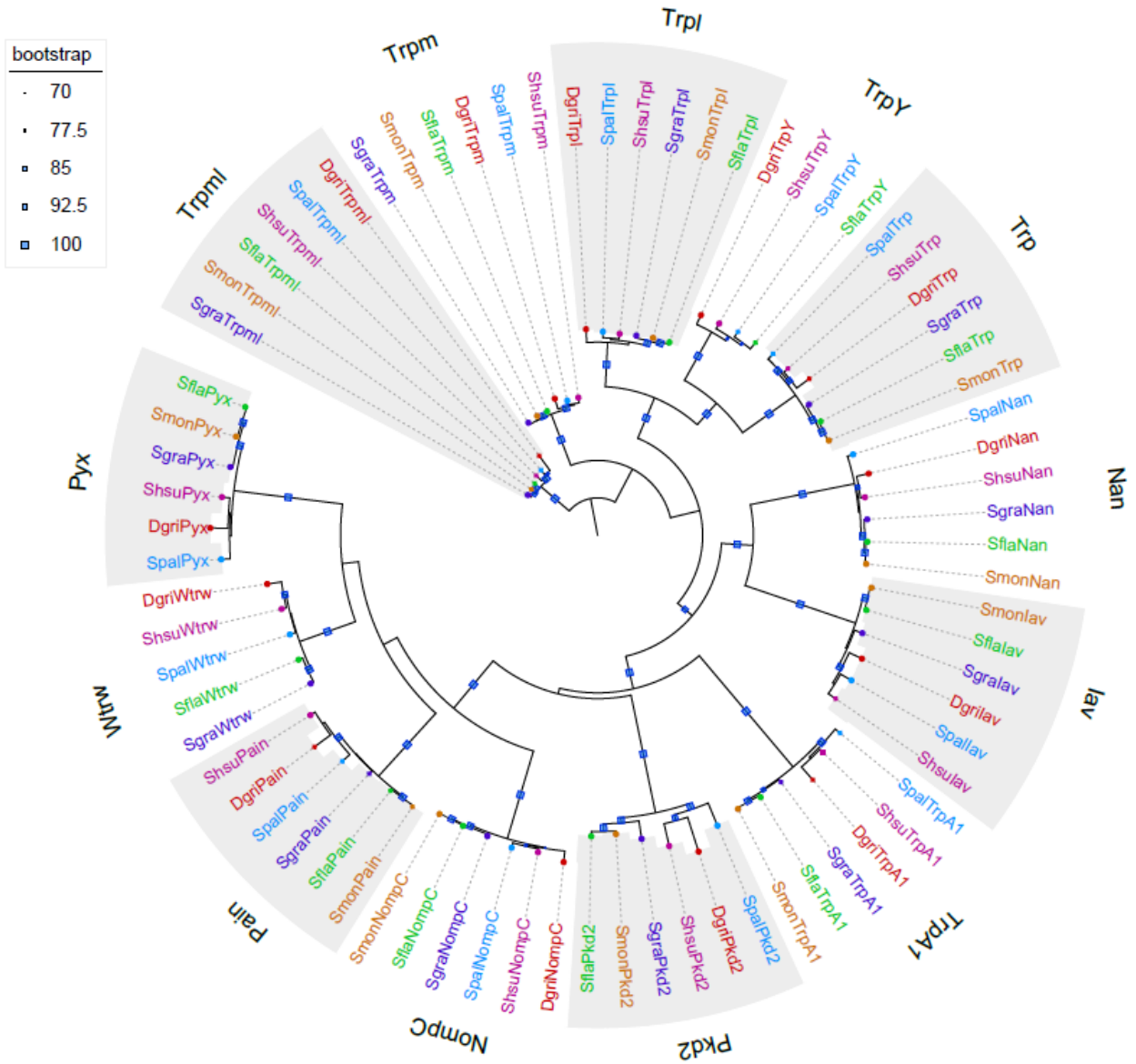


Figure S7. Phylogeny of transient receptor potential channels.

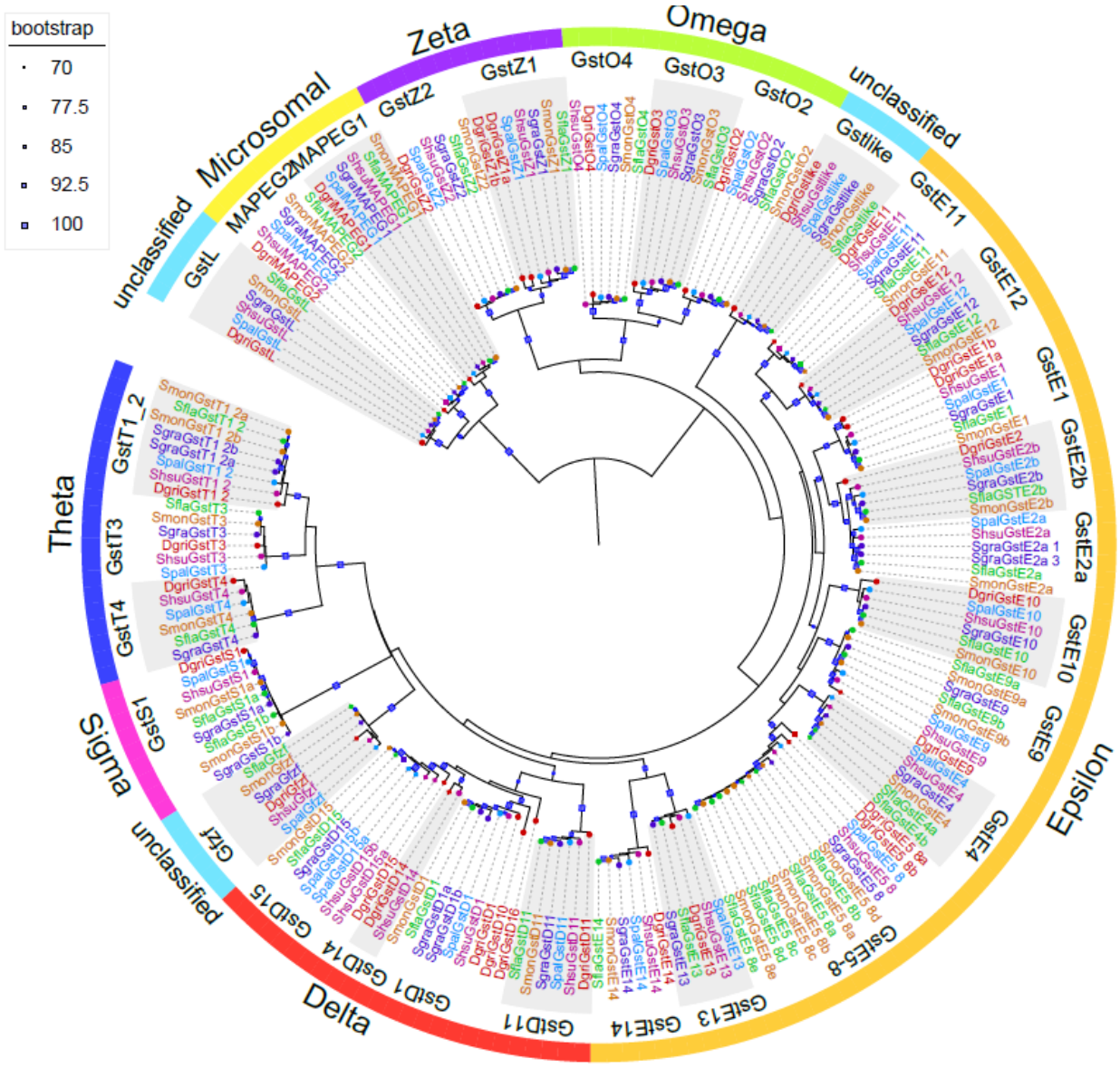


Figure S8. Phylogeny of glutathione S-transferases

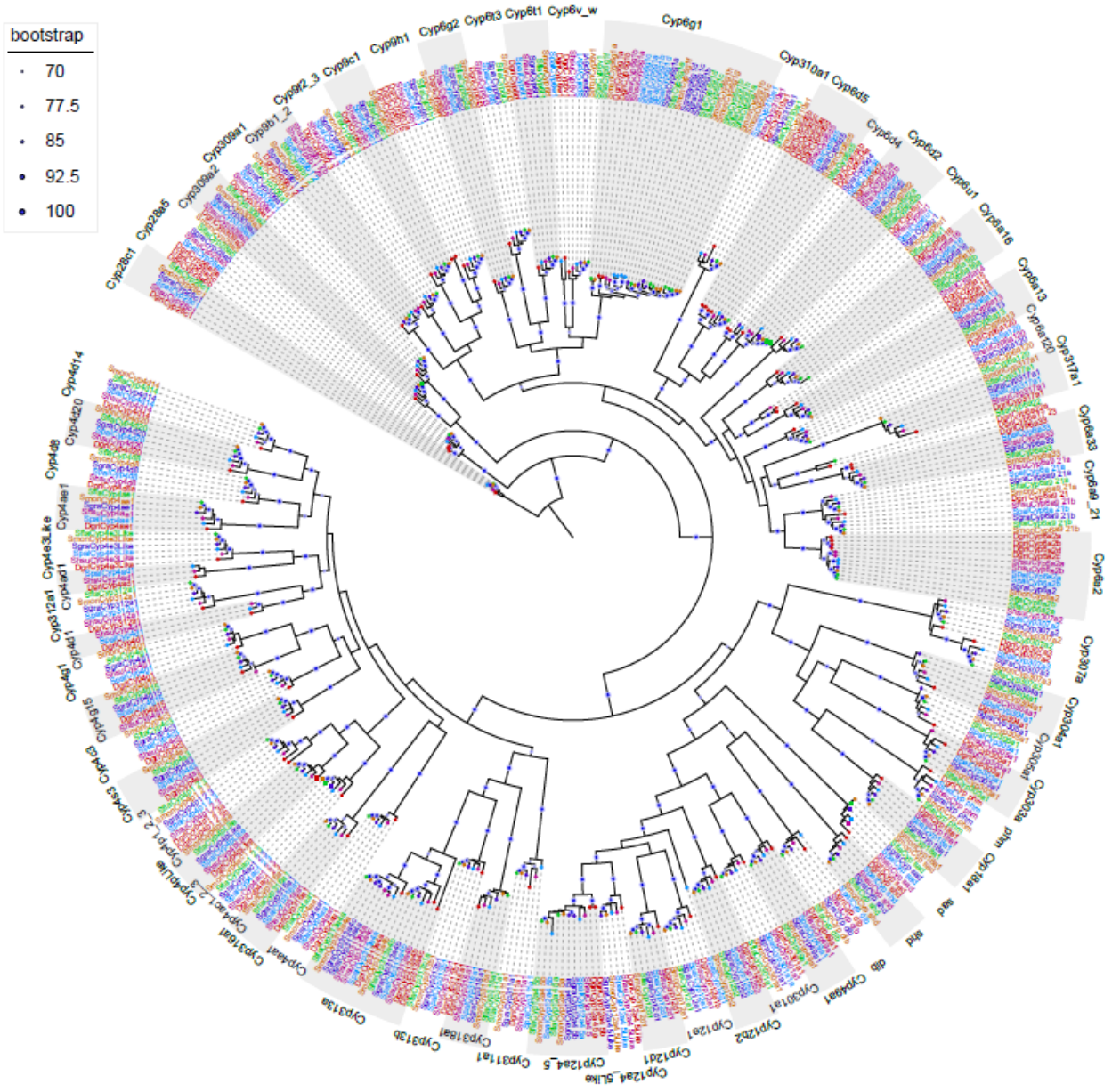


Figure S9. Phylogeny of cytochrome CYP450s

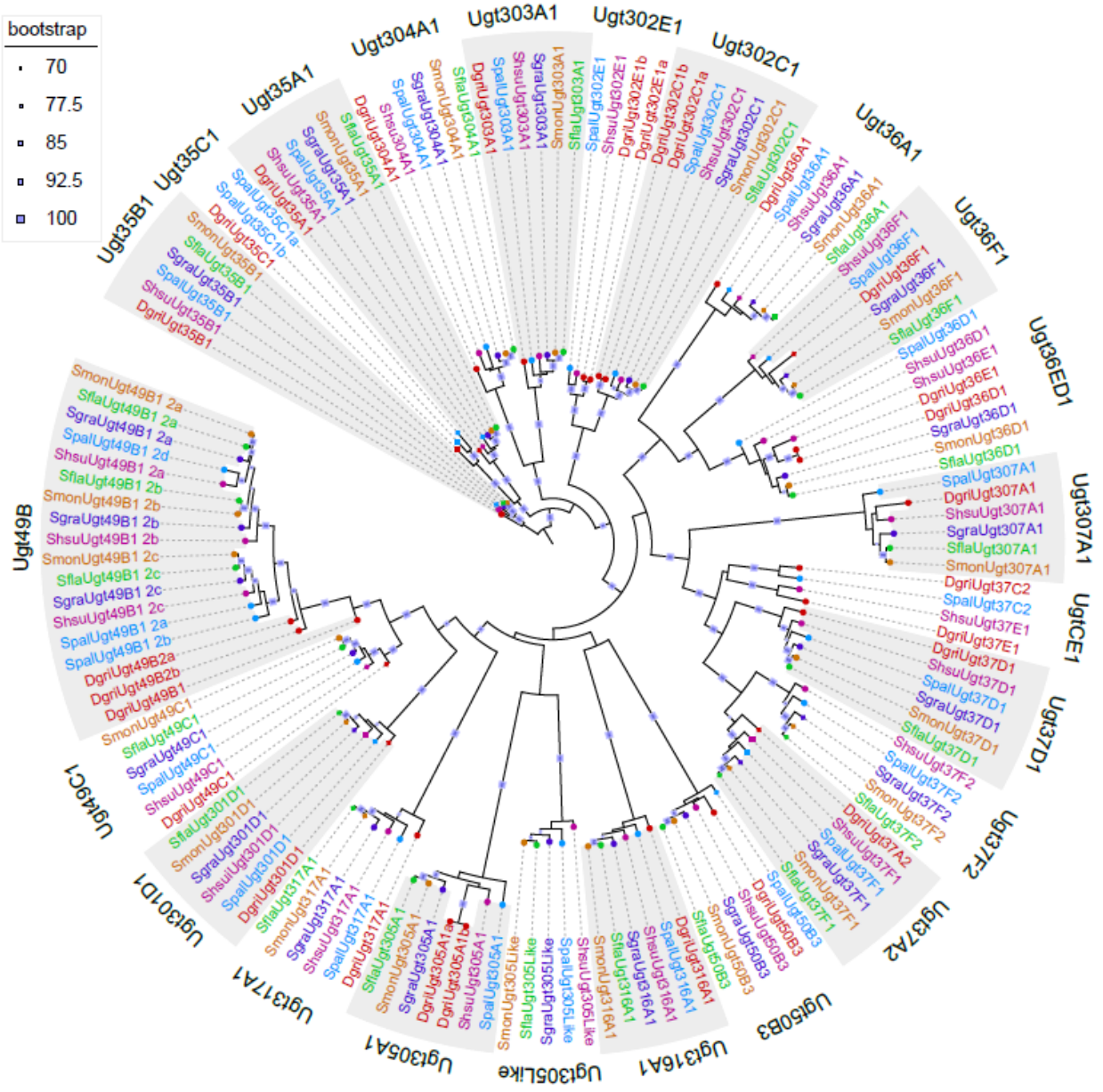


Figure S10. Phylogeny of UDP glucuronosyltransferases UGTs.

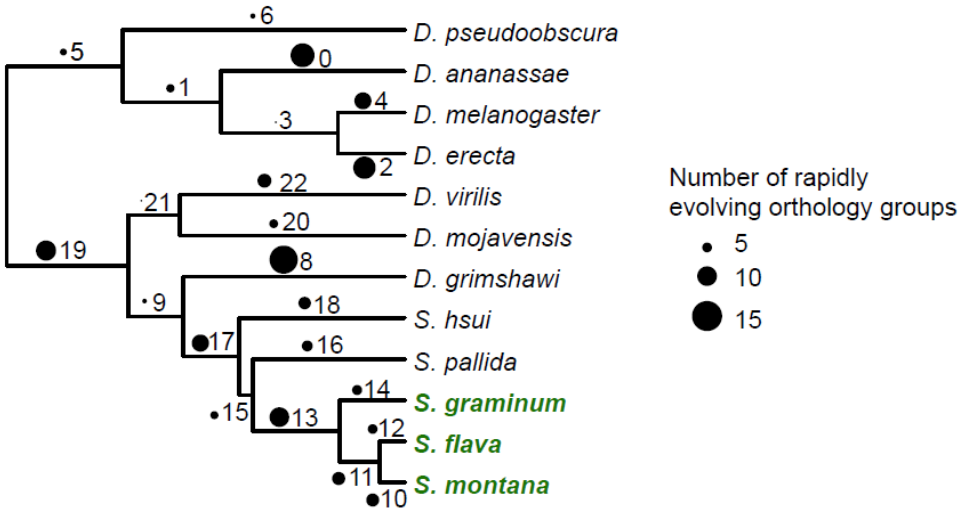


Figure S11. Elevated turnover among chemosensory and detoxification orthology groups The CAFE analysis identified specific orthology groups (i.e., specific genes) that are rapidly expanding or contracting (See Table S5 for significant orthology groups). Numerical values along each branch in the tree indicate the node ID, corresponding to those shown in Table S5. Circles represent the number of rapidly expanding or contracting groups along each branch.

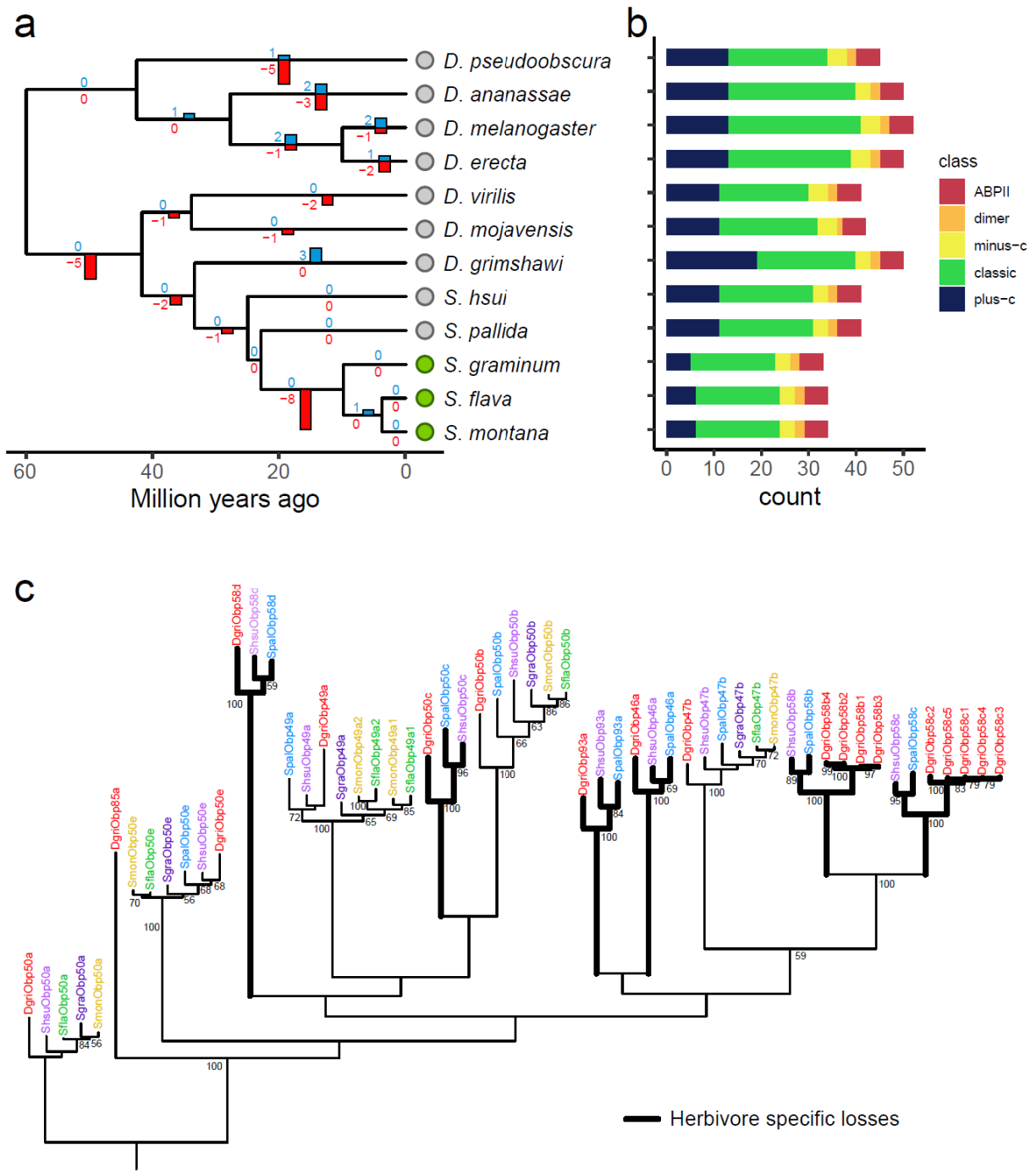


Figure S12. Dramatic loss of "Plus-C" odorant binding proteins at the base of herbivorous drosophilid clade. (a) CAFE gene family expansion/contraction analysis for all OBPs – gains are shown above branches in blue, losses are shown below in red. Circles indicate feeding ecology: green = herbivorous, gray = non-herbivorous. (b) Gene counts by OBP class show that the majority of herbivore-specific losses are among the Plus-C OBPs. (c) Phylogeny of Plus-C OBPs. Thick branches indicate lineages with herbivore-specific losses. Node values indicate bootstrap support.

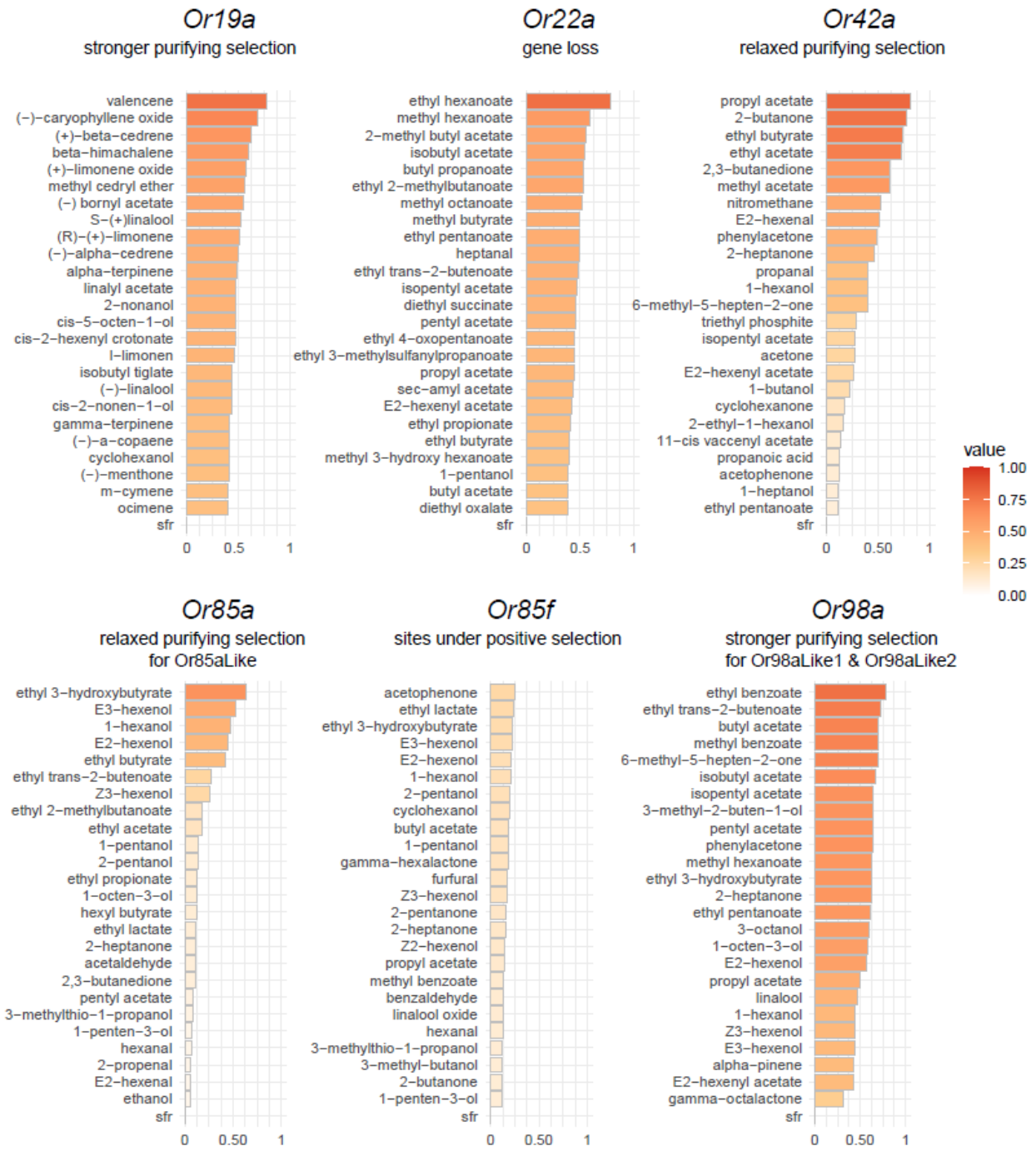


Figure S13. Responses towards odorants among ORs rapidly evolving in the ancestral lineage to all herbivores using data from DoOR.
ORs shown were either lost or experienced changes in selection pressures and have orthologs deorphanized in *D. melanogaster*. Response profiles show the relative sensitivity (along the x-axis) of each *DmelOR* towards individual odorants (y-axis). The top 25 odorants that elicit the highest spike response are shown. Responses are normalized against the spontaneous firing rate ("sfr"). Plots were generated through two DoOR packages (DoOR.data and DoOR.functions).

Appendix G: Supplemental methods from Chapter 3.

MultiCAFE feeding assays.

Because flies used for electrophysiology were reared under slightly different conditions (in Gif-Sur-Yvette, France) compared to those used in the main two-choice CAFE assays (performed in Berkeley, CA, USA), we performed CAFE assays alongside electrophysiological experiments to test that feeding preferences were comparable between the two conditions. This also tested whether feeding preferences were consistent under a different assay, the MultiCAFE assay (Sellier et al. 2011) with modifications. This assay employed a two-choice test, but rather than scoring based on abdomen color, preferences were indicated by the volume of each solution consumed. Newly eclosed adult female flies were kept on Nutrifly media prior to starvation. Roughly 20 mated female flies (5-8 days old) were transferred to a vial with moistened filter paper and starved at 25ºC, 70% humidity for 24 hours. They were then transferred to individual chambers (Colombi et al. unpublished) and allowed to feed from two capillaries containing 5mM sucrose alone versus 5mM sucrose with 10 mM sinigrin for 16 hours in a chamber lit with two LED arrays (25ºC, 70% humidity). Both capillaries contained erioglaucine blue dye (0.125mg/ml), which enabled video tracking of consumption volumes. Six boxes (with eight individual chambers) per species were run over two separate days. Images of the capillaries were taken every 2.5 minutes, and volume consumed was calculated based on the reduction in solution in the capillaries, with evaporation rates subtracted from control chambers without flies. Images were analyzed with the MultiCAFE program in Icy (de Chaumont et al. 2011), and subsequent data analysis in R.

Appendix H: Supplemental tables from Chapter 3.

Table S1. Summary statistics of feeding preferences towards caffeine and lobeline (Fig. 1c).

		Summary statistics						Kruskal-Wallis test			Post-hoc Dunn's test					
	Conc. (mM)	**Species**	**N**	**Preference Index**	**sd**	**se**	**ci**	**chi-squared**	**df**	**p-value**	**Compari-son**	**Z**	**P.unadj**	**P.adj**	**Group**	**Letter**
Caffeine	1	Dmel	12	0.541	0.144	0.041	0.091	23.026	3	**3.99E-05**	Dmel-Sfla	1.200	0.23	1.000	Dmel	a
		Shsu	9	0.186	0.19	0.063	0.146				Dmel-Shsu	3.583	0.0003	**0.002**	Sfla	ab
		Spal	11	0.138	0.108	0.033	0.073				Sfla-Shsu	1.860	0.063	0.378	Shsu	bc
		Sfla	6	0.396	0.102	0.042	0.107				Dmel-Spal	4.312	0.00002	**0.0001**	Spal	c
											Sfla-Spal	2.365	0.018	0.108		
											Shsu-Spal	0.489	0.624	1.000		
	10	Dmel	14	0.083	0.136	0.036	0.078	1.566	3	0.667	Dmel-Sfla	0.463	0.643	1.000	Dmel	a
		Shsu	12	0.142	0.195	0.056	0.124				Dmel-Shsu	-0.419	0.675	1.000	Sfla	a
		Spal	11	0.132	0.166	0.05	0.112				Sfla-Shsu	-0.763	0.445	1.000	Shsu	a
		Sfla	5	0.052	0.097	0.043	0.12				Dmel-Spal	-0.919	0.358	1.000	Spal	a
											Sfla-Spal	-1.134	0.257	1.000		
											Shsu-Spal	-0.492	0.623	1.000		
Lobeline	0.1	Dmel	7	0.182	0.057	0.022	0.053	13.071	3	**0.004**	Dmel-Sfla	2.596	0.009	0.057	Dmel	a
		Shsu	7	0.047	0.098	0.037	0.091				Dmel-Shsu	2.362	0.018	0.109	Sfla	b
		Spal	9	0.216	0.185	0.062	0.142				Sfla-Shsu	-0.440	0.66	1.000	Shsu	b
		Sfla	5	0.013	0.03	0.013	0.037				Dmel-Spal	-0.002	0.998	1.000	Spal	a
											Sfla-Spal	-2.727	0.006	**0.038**		
											Shsu-Spal	-2.507	0.012	0.073		
	1	Dmel	12	0.015	0.038	0.011	0.024	7.819	3	0.05	Dmel-Sfla	0.867	0.386	1.000	Dmel	ab
		Shsu	5	0	0	0	0				Dmel-Shsu	0.814	0.415	1.000	Sfla	a
		Spal	7	0.044	0.075	0.028	0.07				Sfla-Shsu	0.000	1	1.000	Shsu	a
		Sfla	6	0	0	0	0				Dmel-Spal	-1.921	0.055	0.329	Spal	b
											Sfla-Spal	-2.421	0.015	0.093		
											Shsu-Spal	-2.300	0.021	0.129		

Table S2. Summary statistics of feeding preferences towards glucosinolates (Fig. 1d).

		Summary statistics						Kruskal-Wallis test			Post-hoc Dunn's test				
Compound	**Conc. (mM)**	**Species**	**N**	**Preference Index**	**sd**	**se**	**ci**	**chi-squared**	**df**	**p-value**	**Comparison**	**Z**	**P.unadj**	**P.adj**	**Group**	**Letter**
Gluconasturtiin	1	Dmel	21	0.327	0.199	0.043	0.09	9.839	3	**0.02**	Dmel-Sfla	-2.89	0.004	**0.023**	Dmel	a
		Shsu	12	0.345	0.179	0.052	0.114				Dmel-Shsu	-0.289	0.773	1	Sfla	b
		Spal	18	0.312	0.202	0.048	0.1				Sfla-Shsu	2.506	0.012	0.073	Shsu	a
		Sfla	5	0.662	0.084	0.038	0.105				Dmel-Spal	0.301	0.764	1	Spal	a
											Sfla-Spal	3.036	0.002	**0.014**		
											Shsu-Spal	0.539	0.59	1		
	10	Dmel	10	0.297	0.187	0.059	0.134	17.405	3	**0.001**	Dmel-Sfla	0.819	0.413	1	Dmel	a
		Shsu	9	0.02	0.026	0.009	0.02				Dmel-Shsu	3.972	0.00007	**0.0004**	Sfla	a
		Spal	8	0.176	0.168	0.059	0.14				Sfla-Shsu	3.174	0.002	**0.009**	Shsu	b
		Sfla	10	0.189	0.126	0.04	0.09				Dmel-Spal	1.398	0.162	0.973	Spal	a
											Sfla-Spal	0.626	0.532	1		
											Shsu-Spal	-2.391	0.017	0.101		
Glucotropaeolin	1	Dmel	10	0.554	0.121	0.038	0.086	13.529	3	**0.004**	Dmel-Sfla	1.338	0.181	1	Dmel	a
		Shsu	7	0.368	0.132	0.05	0.122				Dmel-Shsu	2.021	0.043	0.26	Sfla	ab
		Spal	6	0.244	0.078	0.032	0.082				Sfla-Shsu	0.548	0.584	1	Shsu	bc
		Sfla	6	0.434	0.175	0.071	0.183				Dmel-Spal	3.613	0.0003	**0.002**	Spal	c
											Sfla-Spal	2.035	0.042	0.251		
											Shsu-Spal	1.564	0.118	0.708		
	10	Dmel	8	0.247	0.107	0.038	0.089	13.069	3	**0.004**	Dmel-Sfla	0.535	0.593	1	Dmel	a
		Shsu	5	0.104	0.133	0.059	0.165				Dmel-Shsu	1.926	0.054	0.324	Sfla	a
		Spal	7	0.029	0.038	0.014	0.035				Sfla-Shsu	1.457	0.145	0.87	Shsu	ab
		Sfla	8	0.22	0.185	0.066	0.155				Dmel-Spal	3.265	0.001	**0.007**	Spal	b
											Sfla-Spal	2.749	0.006	**0.036**		
											Shsu-Spal	1.011	0.312	1		
Neoglucobrassicin	1	Dmel	9	0.584	0.189	0.063	0.145	6.972	3	0.073	Dmel-Sfla	0.756	0.45	1	Dmel	a
		Shsu	7	0.371	0.199	0.075	0.184				Dmel-Shsu	1.947	0.051	0.309	Sfla	ab
		Spal	4	0.3	0.127	0.063	0.201				Sfla-Shsu	0.172	0.863	1	Shsu	ab
		Sfla	1	0.4	NA	NA	NaN				Dmel-Spal	2.366	0.018	0.108	Spal	b
											Sfla-Spal	0.559	0.576	1		
											Shsu-Spal	0.703	0.482	1		
	10	Dmel	8	0.218	0.179	0.063	0.15	11.862	3	**0.008**	Dmel-Sfla	-1.329	0.184	1	Dmel	a
		Shsu	7	0.118	0.093	0.035	0.086				Dmel-Shsu	0.843	0.399	1	Sfla	a
		Spal	6	0.022	0.036	0.015	0.038				Sfla-Shsu	1.937	0.053	0.317	Shsu	ab
		Sfla	3	0.398	0.037	0.021	0.092				Dmel-Spal	2.522	0.012	0.07	Spal	b
											Sfla-Spal	3.199	0.001	**0.008**		
											Shsu-Spal	1.663	0.096	0.578		
Glucobrassicin	1	Dmel	10	0.462	0.207	0.065	0.148	12.388	3	**0.006**	Dmel-Sfla	0.174	0.862	1	Dmel	a

		Shsu	6	0.809	0.11	0.045	0.115				Dmel-Shsu	-2.948	0.003	**0.019**	Sfla	a
		Spal	6	0.414	0.196	0.08	0.206				Sfla-Shsu	-2.971	0.003	**0.018**	Shsu	b
		Sfla	8	0.403	0.265	0.094	0.222				Dmel-Spal	0.37	0.711	1	Spal	a
											Sfla-Spal	0.202	0.84	1		
											Shsu-Spal	2.968	0.003	**0.018**		
	10	Dmel	11	0.158	0.16	0.048	0.108	9.493	3	**0.023**	Dmel-Sfla	-2.052	0.04	0.241	Dmel	a
		Shsu	8	0.312	0.315	0.111	0.263				Dmel-Shsu	-0.544	0.586	1	Sfla	b
		Spal	11	0.075	0.095	0.029	0.064				Sfla-Shsu	1.429	0.153	0.918	Shsu	ab
		Sfla	7	0.484	0.318	0.12	0.294				Dmel-Spal	1.111	0.266	1	Spal	a
											Sfla-Spal	3.032	0.002	**0.015**		
											Shsu-Spal	1.564	0.118	0.707		
Sinigrin	1	Dmel	15	0.459	0.135	0.035	0.075	4.756	3	0.191	Dmel-Sfla	-0.584	0.56	1	Dmel	a
		Shsu	11	0.381	0.218	0.066	0.147				Dmel-Shsu	1.209	0.227	1	Sfla	a
		Spal	13	0.373	0.164	0.045	0.099				Sfla-Shsu	1.501	0.133	0.8	Shsu	a
		Sfla	6	0.513	0.193	0.079	0.202				Dmel-Spal	1.568	0.117	0.701	Spal	a
											Sfla-Spal	1.775	0.076	0.455		
											Shsu-Spal	0.279	0.78	1		
	10	Dmel	28	0.455	0.134	0.025	0.052	37.447	3	**0**	Dmel-Sfla	4.09	0.00004	**0.0003**	Dmel	a
													0.000000			
		Shsu	20	0.114	0.155	0.035	0.073				Dmel-Shsu	5.39	07	**0.0000004**	Sfla	b
		Spal	24	0.247	0.19	0.039	0.08				Sfla-Shsu	-0.815	0.415	1	Shsu	b
		Sfla	5	0.044	0.099	0.044	0.123				Dmel-Spal	3.376	0.0007	**0.004**	Spal	c
											Sfla-Spal	-2.129	0.033	0.2		
											Shsu-Spal	-2.11	0.035	0.209		
Glucoraphanin	1	Dmel	12	0.541	0.216	0.062	0.137	4.844	3	0.184	Dmel-Sfla	1.718	0.086	0.515	Dmel	a
		Shsu	9	0.425	0.239	0.08	0.184				Dmel-Shsu	1.41	0.159	0.951	Sfla	ab
		Spal	12	0.349	0.147	0.042	0.093				Sfla-Shsu	-0.288	0.773	1	Shsu	ab
		Sfla	9	0.375	0.22	0.073	0.169				Dmel-Spal	1.981	0.048	0.286	Spal	b
											Sfla-Spal	0.116	0.908	1		
											Shsu-Spal	0.424	0.672	1		
	10	Dmel	8	0.442	0.211	0.075	0.177	6.89	3	0.075	Dmel-Sfla	2.217	0.027	0.16	Dmel	a
		Shsu	8	0.225	0.132	0.047	0.11				Dmel-Shsu	2.234	0.026	0.153	Sfla	b
		Spal	8	0.292	0.172	0.061	0.144				Sfla-Shsu	-0.258	0.797	1	Shsu	b
		Sfla	5	0.2	0.158	0.071	0.197				Dmel-Spal	1.514	0.13	0.781	Spal	ab
											Sfla-Spal	-0.889	0.374	1		
											Shsu-Spal	-0.72	0.471	1		

Table S3. Summary statistics of electrophysiological responses to sucrose, caffeine, and lobeline using generalized linear models with gamma distributions (Fig. 3c). Spikes per second from labellar sensilla was regressed against species, concentration, and their interaction term.

Compound	Sensilla		Estimate	Std. Error	t value	Pr(>|t|)		Sensilla		Estimate	Std. Error	t value	Pr(>|t|)	
caffeine	L	(Intercept)	0.619	0.095	6.549	**<0.001**	***	S-a	(Intercept)	0.064	0.011	5.655	**<0.001**	***
		insectDmel	0.246	0.157	1.571	0.123			insectDmel	0.03	0.021	1.421	0.16	
		insectSpal	0.281	0.173	1.624	0.112			insectSpal	0.029	0.019	1.522	0.133	
		conc1	0.001	0.002	0.564	0.575			conc1	-0.0004	<0.001	-3.237	**0.002**	**
		insectDmel:conc1	-0.001	0.003	-0.434	0.666			insectDmel:conc1	0.0001	<0.001	0.32	0.75	
		insectSpal:conc1	<0.001	0.003	0.036	0.971			insectSpal:conc1	-0.0002	<0.001	-0.937	0.352	
	l-a	(Intercept)	0.148	0.031	4.691	**<0.001**	***	S-b	(Intercept)	0.046	0.008	5.582	**<0.001**	***
		insectDmel	0.268	0.111	2.419	**0.019**	*		insectDmel	0.005	0.013	0.38	0.705	
		insectSpal	-0.095	0.033	-2.879	**0.006**	**		insectSpal	0.101	0.031	3.279	**0.002**	**
		conc1	-0.001	<0.001	-3.301	**0.002**	**		conc1	-0.0002	<0.001	-1.32	0.192	
		insectDmel:conc1	0.004	0.003	1.549	0.127			insectDmel:conc1	0.000003	<0.001	0.018	0.986	
		insectSpal:conc1	0.001	<0.001	2.567	**0.013**	*		insectSpal:conc1	0.002	0.001	1.458	0.15	
	l-b	(Intercept)	0.11	0.023	4.834	**<0.001**	***	S-c	(Intercept)	0.064	0.022	2.885	**0.006**	**
		insectDmel	-0.065	0.024	-2.687	**0.01**	**		insectDmel	0.701	0.216	3.249	**0.002**	**
		insectSpal	-0.057	0.025	-2.312	**0.025**	*		insectSpal	0.071	0.044	1.608	0.114	
		conc1	-0.001	<0.001	-2.341	**0.024**	*		conc1	0.001	0.001	1.34	0.186	
		insectDmel:conc1	0.001	0	2.047	**0.046**	*		insectDmel:conc1	0.001	0.005	0.242	0.81	
		insectSpal:conc1	<0.001	<0.001	1.429	0.16			insectSpal:conc1	0.001	0.002	0.752	0.456	
lobeline	L	(Intercept)	0.826	0.11	7.525	**<0.001**	***	S-a	(Intercept)	0.03	0.004	8.392	**<0.001**	***
		insectDmel	0.174	0.158	1.107	0.274			insectDmel	0.01	0.006	1.573	0.121	
		insectSpal	-0.034	0.148	-0.232	0.818			insectSpal	0.029	0.007	3.959	**<0.001**	***
		conc1	0.002	0.019	0.109	0.913			conc1	-0.001	0.001	-2.467	**0.017**	*
		insectDmel:conc1	-0.002	0.027	-0.077	0.939			insectDmel:conc1	0.002	0.001	2.357	**0.022**	*
		insectSpal:conc1	0.005	0.025	0.18	0.858			insectSpal:conc1	-0.001	0.001	-1.428	0.158	
	l-a	(Intercept)	0.056	0.008	6.948	**<0.001**	***	S-b	(Intercept)	0.021	0.005	4.245	**<0.001**	***
		insectDmel	0.022	0.015	1.488	0.142			insectDmel	0.016	0.009	1.703	0.094	.
		insectSpal	0.002	0.012	0.171	0.865			insectSpal	0.106	0.029	3.728	**<0.001**	***
		conc1	-0.002	0.001	-1.775	0.081	.		conc1	-0.0005	0.001	-0.645	0.521	
		insectDmel:conc1	0.006	0.003	2.364	**0.021**	*		insectDmel:conc1	0.0001	0.002	0.065	0.949	
		insectSpal:conc1	0.001	0.002	0.375	0.709			insectSpal:conc1	-0.001	0.005	-0.19	0.85	
	l-b	(Intercept)	0.044	0.009	4.99	**<0.001**	***	S-c	(Intercept)	0.087	0.03	2.933	**0.005**	**
		insectDmel	0.433	0.09	4.796	**<0.001**	***		insectDmel	0.529	0.197	2.683	**0.01**	**
		insectSpal	0.011	0.013	0.794	0.431			insectSpal	0.099	0.071	1.399	0.168	
		conc1	-0.001	0.001	-0.509	0.613			conc1	0.019	0.011	1.722	0.091	.
		insectDmel:conc1	0.047	0.026	1.796	0.078	.		insectDmel:conc1	-0.01	0.034	-0.291	0.772	
		insectSpal:conc1	-0.001	0.002	-0.739	0.463			insectSpal:conc1	0.055	0.042	1.299	0.2	
sucrose	L	(Intercept)	0.0365	0.007	5.252	**<0.001**	***	S-a	(Intercept)	0.06	0.011	5.624	**<0.001**	***
		insectDmel	-0.008	0.009	-0.853	0.396			insectDmel	0.64	0.194	3.3	**0.002**	**
		insectSpal	-0.0008	0.009	-0.087	0.931			insectSpal	0.008	0.016	0.49	0.626	
		conc1	-0.0002	<0.001	-2.209	**0.03**	*		conc1	-0.0004	<0.001	-3.314	**0.002**	**
		insectDmel:conc1	0.00002	<0.001	0.188	0.851			insectDmel:conc1	-0.066	0.0194	-3.414	**0.001**	**
		insectSpal:conc1	0.00001	<0.001	0.12	0.905			insectSpal:conc1	0.00006	<0.001	0.326	0.745	

I-a	(Intercept)	0.106	0.016	6.599	**<0.001**	***
	insectDmel	-0.032	0.023	-1.396	0.168	
	insectSpal	-0.015	0.023	-0.635	0.528	
	conc1	-0.001	<0.001	-3.666	**0.001**	***
	insectDmel:conc1	0.0003	<0.001	1.155	0.253	
	insectSpal:conc1	0.0003	<0.001	0.89	0.377	
I-b	(Intercept)	0.094	0.018	5.349	**<0.001**	***
	insectDmel	-0.009	0.025	-0.356	0.723	
	insectSpal	-0.049	0.019	-2.608	**0.012**	*
	conc1	-0.0007	<0.001	-4	**<0.001**	***
	insectDmel:conc1	0.0001	<0.001	0.366	0.716	
	insectSpal:conc1	0.0006	<0.001	2.772	**0.008**	**

S-b	(Intercept)	0.037	0.005	7.034	**<0.001**	***
	insectDmel	0.03	0.0143	2.192	**0.032**	*
	insectSpal	0.048	0.0157	3.063	**0.003**	**
	conc1	-0.0002	<0.001	-3.323	**0.002**	**
	insectDmel:conc1	-0.0003	<0.001	-2.021	**0.048**	*
	insectSpal:conc1	-0.0002	<0.001	-1.135	0.261	
S-c	(Intercept)	0.033	0.005	5.993	**<0.001**	***
	insectDmel	0.012	0.011	1.108	0.273	
	insectSpal	0.021	0.0111	1.861	0.069	.
	conc1	-0.0002	<0.001	-2.246	**0.029**	*
	insectDmel:conc1	-0.0002	<0.001	-1.334	0.188	
	insectSpal:conc1	-0.0002	<0.001	-1.547	0.128	

Table S4. Summary statistics of electrophysiological responses to glucosinolates (sinigrin, gluconasturtiin, and neoglucobrassicin), using generalized linear models with gamma distributions (Fig. 4a). Spikes per second from labellar sensilla was regressed against species, concentration, and their interaction term.

Compound	Sensilla		Estimate	Std. Error	t value	Pr(>|t|)		Sensilla		Estimate	Std. Error	t value	Pr(>|t|)	
sinigrin	L	(Intercept)	0.897	0.159	5.628	**<0.001**	***	S-a	(Intercept)	0.117	0.023	5.092	**<0.001**	***
		insectDmel	-0.346	0.202	-1.711	0.091	.		insectDmel	0.492	0.177	2.782	**0.007**	**
		insectSpal	0.123	0.255	0.48	0.632			insectSpal	0.511	0.138	3.694	**<0.001**	***
		conc1	-0.026	0.005	-4.729	**<0.001**	***		conc1	-0.002	0.001	-1.756	0.083	.
		insectDmel:conc1	0.022	0.009	2.464	**0.016**	*		insectDmel:conc1	0.002	0.01	0.155	0.877	
		insectSpal:conc1	0.017	0.013	1.319	0.19			insectSpal:conc1	-0.011	0.006	-2.025	**0.046**	*
	I-a	(Intercept)	0.185	0.037	4.959	**<0.001**	***	S-b	(Intercept)	0.043	0.005	8.529	**<0.001**	***
		insectDmel	0.428	0.182	2.348	**0.021**	*		insectDmel	0.82	0.147	5.59	**<0.001**	***
		insectSpal	-0.018	0.051	-0.358	0.721			insectSpal	0.433	0.077	5.637	**<0.001**	***
		conc1	-0.002	0.002	-1.391	0.168			conc1	-0.001	<0.001	-2.771	**0.007**	**
		insectDmel:conc1	0.005	0.012	0.471	0.638			insectDmel:conc1	-0.019	0.006	-3.364	**0.001**	**
		insectSpal:conc1	0.002	0.003	0.659	0.512			insectSpal:conc1	-0.006	0.003	-1.738	0.086	.
	I-b	(Intercept)	0.284	0.051	5.581	**<0.001**	***	S-c	(Intercept)	0.086	0.02	4.205	**<0.001**	***
		insectDmel	0.515	0.187	2.762	**0.007**	**		insectDmel	0.93	0.335	2.772	**0.007**	**
		insectSpal	0.049	0.077	0.631	0.53			insectSpal	0.217	0.08	2.716	**0.008**	**
		conc1	-0.007	0.002	-3.676	**<0.001**	***		conc1	-0.002	0.001	-2.571	**0.012**	*
		insectDmel:conc1	0.015	0.014	1.115	0.269			insectDmel:conc1	-0.003	0.018	-0.166	0.869	
		insectSpal:conc1	0.002	0.003	0.454	0.651			insectSpal:conc1	-0.005	0.003	-1.54	0.127	
gluconasturtiin	L	(Intercept)	0.829	0.15	5.514	**<0.001**	***	S-a	(Intercept)	0.057	0.016	3.603	**0.001**	***
		insectDmel	-0.058	0.196	-0.295	0.769			insectDmel	0.151	0.066	2.298	**0.024**	*
		insectSpal	0.163	0.23	0.71	0.48			insectSpal	0.142	0.052	2.732	**0.008**	**
		conc1	0.002	0.01	0.232	0.817			conc1	-0.001	0.001	-0.897	0.372	
		insectDmel:conc1	-0.014	0.012	-1.219	0.227			insectDmel:conc1	0.008	0.007	1.127	0.263	
		insectSpal:conc1	-0.011	0.014	-0.774	0.442			insectSpal:conc1	-0.005	0.002	-2.587	**0.011**	*
	I-a	(Intercept)	0.153	0.028	5.549	**<0.001**	***	S-b	(Intercept)	0.051	0.011	4.733	**<0.001**	***
		insectDmel	0.316	0.11	2.868	**0.005**	**		insectDmel	0.228	0.071	3.222	**0.002**	**
		insectSpal	-0.038	0.035	-1.09	0.279			insectSpal	0.223	0.056	3.983	**<0.001**	***
		conc1	-0.003	0.001	-2.818	**0.006**	**		conc1	0.001	0.001	1.416	0.161	
		insectDmel:conc1	0.018	0.012	1.585	0.116			insectDmel:conc1	0.014	0.009	1.576	0.119	
		insectSpal:conc1	0.0003	0.001	0.228	0.82			insectSpal:conc1	-0.004	0.003	-1.277	0.205	
	I-b	(Intercept)	0.104	0.015	6.757	**<0.001**	***	S-c	(Intercept)	0.054	0.015	3.497	**0.001**	***
		insectDmel	0.896	0.178	5.046	**<0.001**	***		insectDmel	0.767	0.235	3.263	**0.002**	**
		insectSpal	0.0004	0.023	0.016	0.987			insectSpal	0.153	0.059	2.6	**0.011**	*
		conc1	-0.002	0.001	-2.343	**0.023**	*		conc1	0.004	0.002	2.009	**0.048**	*
		insectDmel:conc1	0.002	0.011	0.154	0.878			insectDmel:conc1	-0.011	0.012	-0.934	0.353	
		insectSpal:conc1	-0.0004	0.001	-0.382	0.704			insectSpal:conc1	0.013	0.009	1.426	0.158	
neoglucobrassicin	L	(Intercept)	0.929	0.070	13.311	**<0.001**	***	S-a	(Intercept)	0.094	0.025	3.827	**<0.001**	***
		insectDmel	0.071	0.100	0.718	0.476			insectDmel	0.141	0.057	2.465	**0.016**	*
		insectSpal	-0.073	0.094	-0.775	0.441			insectSpal	0.024	0.036	0.663	0.509	
		conc1	0.003	0.005	0.58	0.564			conc1	-0.002	0.001	-2.133	**0.036**	*
		insectDmel:conc1	-0.003	0.007	-0.422	0.674			insectDmel:conc1	-0.003	0.002	-1.494	0.139	
		insectSpal:conc1	0.00008	0.006	0.013	0.989			insectSpal:conc1	-0.001	0.001	-0.728	0.469	

I-a	(Intercept)	0.159	0.036	4.417	**<0.001**	***	S-b	(Intercept)	0.075	0.024	3.073	**0.003**	**
	insectDmel	0.515	0.149	3.453	**0.001**	***		insectDmel	0.295	0.102	2.891	**0.005**	**
	insectSpal	-0.08	0.039	-2.037	**0.046**	*		insectSpal	0.07	0.054	1.306	0.195	
	conc1	-0.003	0.001	-2.237	**0.029**	*		conc1	0.001	0.001	0.407	0.685	
	insectDmel:conc1	0.017	0.013	1.321	0.191			insectDmel:conc1	-0.005	0.005	-1.176	0.243	
	insectSpal:conc1	0.001	0.001	0.749	0.456			insectSpal:conc1	0.01	0.006	1.793	0.077	.
I-b	(Intercept)	0.147	0.039	3.728	**<0.001**	***	S-c	(Intercept)	0.063	0.01	6.616	**<0.001**	***
	insectDmel	0.742	0.235	3.163	**0.002**	**		insectDmel	0.665	0.114	5.819	**<0.001**	***
	insectSpal	-0.052	0.046	-1.127	0.264			insectSpal	0.02	0.019	1.052	0.296	
	conc1	-0.003	0.002	-1.862	0.068	.		conc1	-0.001	<0.001	-1.039	0.302	
	insectDmel:conc1	0.003	0.015	0.165	0.87			insectDmel:conc1	0.011	0.009	1.274	0.207	
	insectSpal:conc1	0.001	0.002	0.395	0.694			insectSpal:conc1	0.041	0.006	6.602	**<0.001**	***

Table S5. Summary statistics of electrophysiological responses to six glucosinolates (gluconasturtiin, glucotropaeolin, neoglucobrassicin, glucobrassicin, sinigrin, glucoraphanin) at 10 mM, using generalized linear models with gamma distributions (Fig. 4b). Spikes per second from labellar sensilla was regressed against species, concentration, and their interaction term.

			Kruskal-Wallis test			Post-hoc Dunn's test			
Compound	**Sensilla**	**Conc. (mM)**	**chi-squared**	**df**	**p-value**	**Comparison**	**Z**	**P.unadj**	**P.adj**
gluconasturtiin	I-a	10	11.228	2	**0.004**	Dmel-Sfla	-2.128	0.033	0.1
						Dmel-Spal	-3.346	0.001	**0.002**
						Sfla-Spal	-1.462	0.144	0.431
	I-b	10	9.514	2	**0.009**	Dmel-Sfla	-2.534	0.011	**0.034**
						Dmel-Spal	-2.869	0.004	**0.012**
						Sfla-Spal	-0.649	0.516	1
	S-a	10	12.77	2	**0.002**	Dmel-Sfla	-3.411	0.001	**0.002**
						Dmel-Spal	-1.185	0.236	0.708
						Sfla-Spal	2.594	0.009	**0.028**
glucotropaeolin	I-a	10	14.251	2	**0.0008**	Dmel-Sfla	-1.803	0.071	0.214
						Dmel-Spal	-3.744	0	**0.001**
						Sfla-Spal	-2.206	0.027	0.082
	I-b	10	7.9917	2	**0.018**	Dmel-Sfla	-1.078	0.281	0.843
						Dmel-Spal	-2.802	0.005	**0.015**
						Sfla-Spal	-1.724	0.085	0.254
	S-a	10	5.606	2	0.06				
neoglucobrassicin	I-a	10	13.345	2	**0.001**	Dmel-Sfla	-2.109	0.035	0.105
						Dmel-Spal	-3.627	0	**0.001**
						Sfla-Spal	-1.35	0.177	0.531
	I-b	10	12.946	2	**0.002**	Dmel-Sfla	-1.923	0.054	0.163
						Dmel-Spal	-3.595	0	**0.001**
						Sfla-Spal	-1.672	0.094	0.283
	S-a	10	7.779	2	**0.02**	Dmel-Sfla	-2.391	0.017	0.05
						Dmel-Spal	-2.464	0.014	**0.041**
						Sfla-Spal	0.005	0.996	1
glucobrassicin	I-a	10	9.508	2	**0.009**	Dmel-Sfla	-2.764	0.006	**0.017**
						Dmel-Spal	-2.695	0.007	**0.021**
						Sfla-Spal	-0.084	0.933	1
	I-b	10	11.421	2	**0.003**	Dmel-Sfla	-1.974	0.048	0.145
						Dmel-Spal	-3.369	0.001	**0.002**
						Sfla-Spal	-1.522	0.128	0.384
	S-a	10	9.502	2	**0.009**	Dmel-Sfla	-2.917	0.004	**0.011**
						Dmel-Spal	-2.175	0.03	0.089
						Sfla-Spal	0.384	0.701	1
sinigrin	I-a	10	0.35	2	0.84				
	I-b	10	4.264	2	0.119				
	S-a	10	9.39	2	**0.009**	Dmel-Sfla	-2.181	0.029	0.087
						Dmel-Spal	-0.651	0.515	1
						Sfla-Spal	2.625	0.009	**0.026**
glucoraphanin	I-a	10	8.245	2	**0.016**	Dmel-Sfla	-2.447	0.014	**0.043**
						Dmel-Spal	-2.584	0.01	**0.029**
						Sfla-Spal	-0.143	0.886	1
	I-b	10	8.818	2	**0.012**	Dmel-Sfla	-2.911	0.004	**0.011**
						Dmel-Spal	-0.87	0.384	1
						Sfla-Spal	2.081	0.037	0.112
	S-a	10	12.518	2	**0.002**	Dmel-Sfla	-3.23	0.001	**0.004**
						Dmel-Spal	-0.344	0.731	1
						Sfla-Spal	3.014	0.003	**0.008**

Table S6. Summary statistics of electrophysiological responses towards mixed solutions of sinigrin (1 and 10 mM) and sucrose (5 mM), using Welch two sample t-test (Fig. 4c)

Species	mean (1 mM sinigrin)	mean (10 mM sinigrin)	t	df	p-value	95% confidence interval
Dmel	19.733	7.571	4.148	19	**<0.001**	[6.025, 18.298]
Spal	11.286	2.667	4.471	10	**0.001**	[4.325, 12.913]
Sfla	18.333	11	2.107	15	0.053	[-0.099, 14.766]

Appendix I: Supplemental figures from Chapter 3.

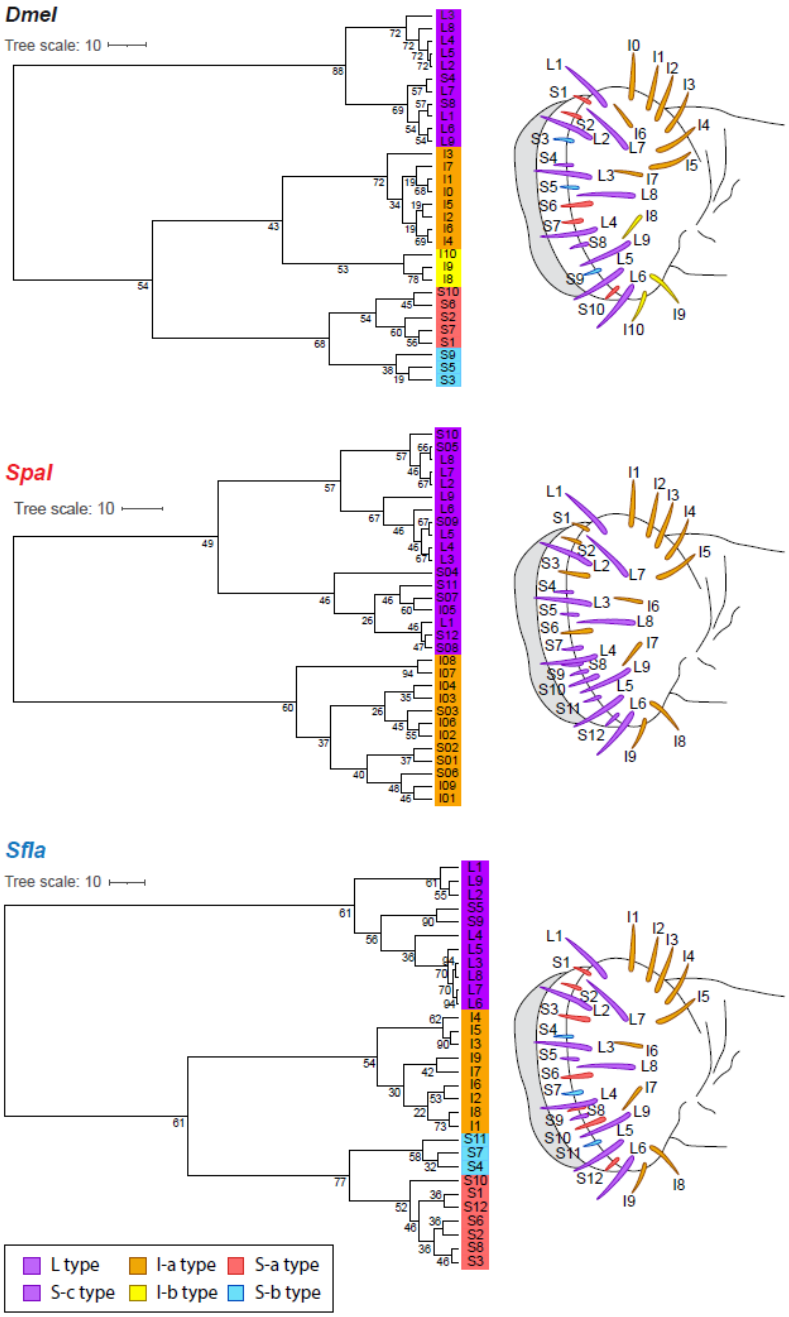


Figure S1. Hierarchical cluster analysis to identify sensilla types on the labellum. Based on Ward's classification method in the PAST program (Hammer et al., 2001) and using electrophysiological responses to caffeine, lobeline, and sucrose. Support values at the nodes were estimated with 100 bootstrap replicates.

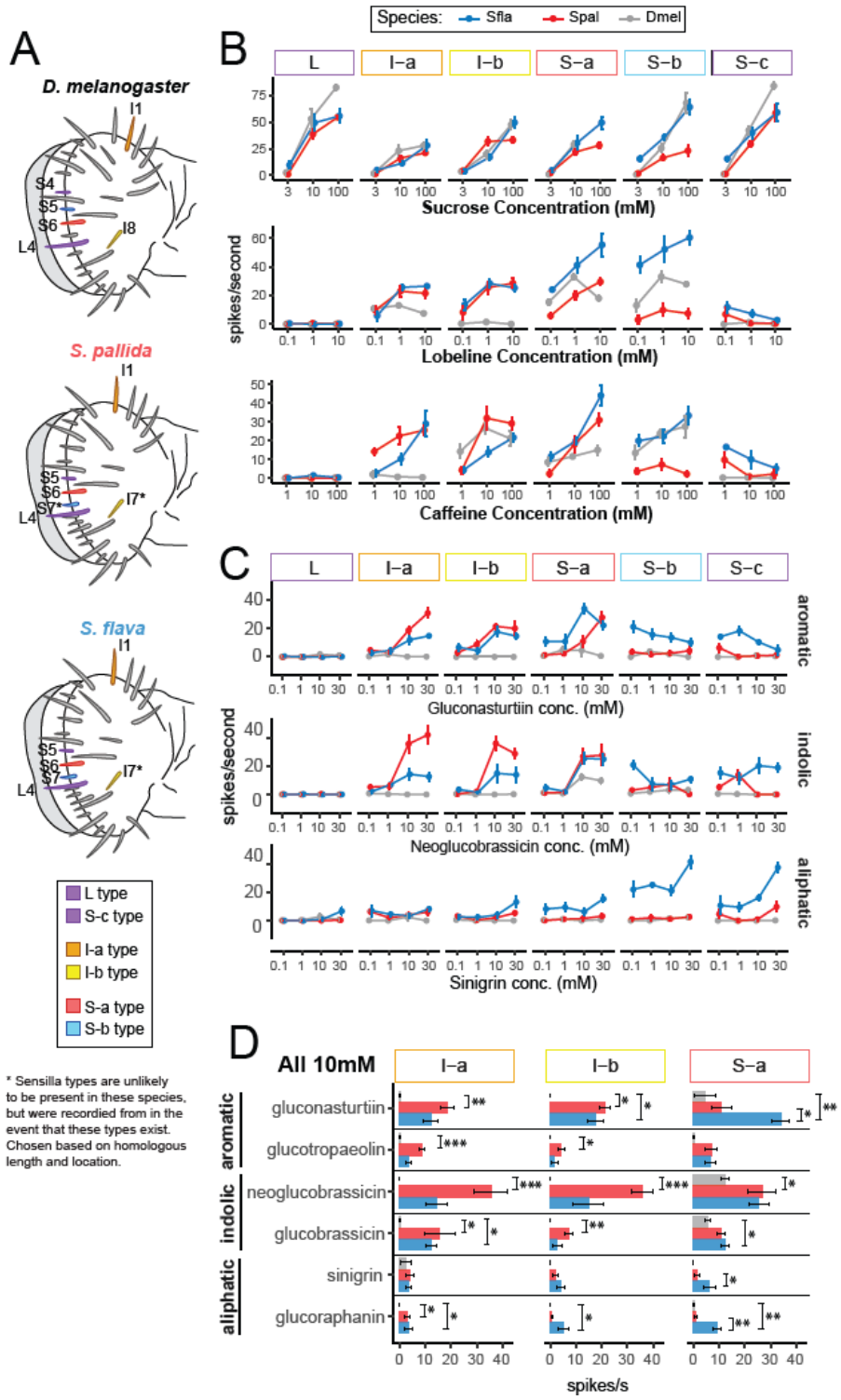


Figure S2. Spike responses from all tested labellar sensilla. (a) Illustration of sensilla representing each potential sensilla class that were recorded from in b, c, and d. Dose-dependent responses towards: (b) sucrose, lobeline, and caffeine; and (c) glucosinolates gluconasturtiin, neoglucobrassicin, and sinigrin. (d) Responses towards 6 glucosinolates (gluconasturtiin, glucotropaeolin, neoglucobrassicin, glucobrassicin, sinigrin, and glucoraphanin), all presented at a 10mM concentration.

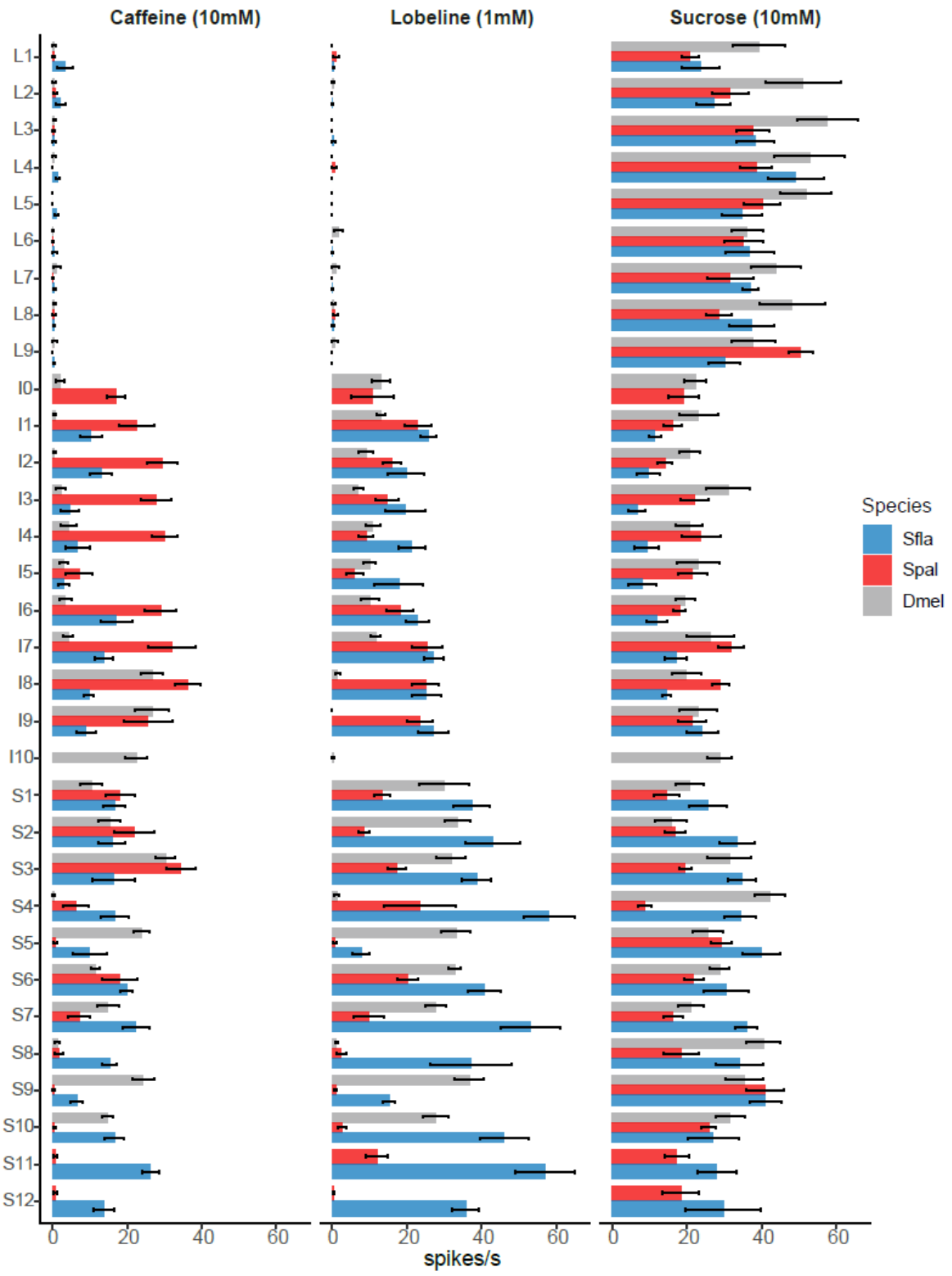


Figure S3. Distributions of spike responses to caffeine (10mM), lobeline (1mM), and sucrose (10mM) across all labellar sensilla (shown as a heatplot in Fig. 3a)

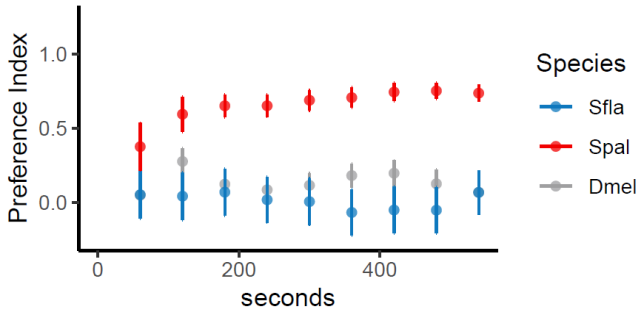


Figure S4. Feeding preferences of herbivorous *S. flava* compared to non-herbivorous species, *S. pallida* and *D. melanogaster* towards the aliphatic glucosinolate sinigrin. Data were collected using a two-choice CAFE assay (see Supplemental Methods).