

UC Berkeley

UC Berkeley Previously Published Works

Title

Evolution of chemosensory and detoxification gene families across herbivorous Drosophilidae

Permalink

<https://escholarship.org/uc/item/12b2j78f>

Journal

G3: Genes, Genomes, Genetics, 13(8)

ISSN

2160-1836

Authors

Peláez, Julianne N
Gloss, Andrew D
Goldman-Huertas, Benjamin
et al.

Publication Date

2023-08-09

DOI


10.1093/g3journal/jkad133

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

Evolution of chemosensory and detoxification gene families across herbivorous *Drosophilidae*

Julianne N. Peláez,^{1,2,*} Andrew D. Gloss,^{3,4} Benjamin Goldman-Huertas,^{1,3} Bernard Kim,⁵ Richard T. Lapoint,^{3,6} Giovanni Pimentel-Solorio,¹ Kirsten I. Verster,^{1,5} Jessica M. Aguilar,¹ Anna C. Nelson Dittrich,^{3,7} Malvika Singhal,^{1,8} Hiromu C. Suzuki,¹ Teruyuki Matsunaga,¹ Ellie E. Armstrong,⁵ Joseph L. M. Charboneau,³ Simon C. Groen,^{1,3,4,9,10,11} David H. Hembry,^{3,12} Christopher J. Ochoa,^{1,13} Timothy K. O'Connor,¹ Stefan Prost,^{1,5} Sophie Zaijier,^{3,14,15} Paul D. Nabity,^{3,10} Jiarui Wang,^{1,16} Esteban Rodas,¹ Irene Liang,¹ Noah K. Whiteman ^{1,17,*}

¹Department of Integrative Biology, University of California Berkeley, Berkeley, CA 94720, USA

²Department of Biology, Brandeis University, Waltham, MA 02453, USA

³Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ 85721, USA

⁴Department of Biology and Center for Genomics and Systems Biology, New York University, New York, NY 10003, USA

⁵Department of Biology, Stanford University, Palo Alto, CA 94305, USA

⁶National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, USA

⁷Boyce Thompson Institute, Cornell University, Ithaca, NY 14853, USA

⁸Department of Chemistry & Biochemistry, University of Oregon, Eugene, OR 97403, USA

⁹Department of Nematology, University of California Riverside, Riverside, CA 92521, USA

¹⁰Department of Botany and Plant Sciences, University of California Riverside, Riverside, CA 92521, USA

¹¹Center for Plant Cell Biology and Institute for Integrative Genome Biology, University of California Riverside, Riverside, CA 92521, USA

¹²Department of Biology, University of Texas Permian Basin, Odessa, TX 79762, USA

¹³Molecular Biology Institute, University of California Los Angeles, Los Angeles, CA 90095, USA

¹⁴Jacobs Institute, Cornell Tech, New York, NY 10044, USA

¹⁵FIND Genomics, New York, NY 10044, USA

¹⁶Department of Biomedical Engineering, Viterbi School of Engineering, University of Southern California, Los Angeles, CA 90007, USA

¹⁷Department of Molecular and Cell Biology, University of California Berkeley, Berkeley, CA 94720, USA

*Corresponding author: Carl J. Shapiro Science Center, Brandeis University, 2-16, Waltham, MA 02453, USA. Email: jpelaez@brandeis.edu; *Corresponding author: University of California Berkeley, 142 Weill Hall #3200, Berkeley, CA 94720, USA. Email: whiteman@berkeley.edu

Abstract

Herbivorous insects are exceptionally diverse, accounting for a quarter of all known eukaryotic species, but the genomic basis of adaptations that enabled this dietary transition remains poorly understood. Many studies have suggested that expansions and contractions of chemosensory and detoxification gene families—genes directly mediating interactions with plant chemical defenses—underlie successful plant colonization. However, this hypothesis has been challenging to test because the origins of herbivory in many insect lineages are ancient (>150 million years ago (mya)), obscuring genomic evolutionary patterns. Here, we characterized chemosensory and detoxification gene family evolution across *Scaptomyza*, a genus nested within *Drosophila* that includes a recently derived (<15 mya) herbivore lineage of mustard (Brassicales) specialists and carnation (Caryophyllaceae) specialists, and several nonherbivorous species. Comparative genomic analyses revealed that herbivorous *Scaptomyza* has among the smallest chemosensory and detoxification gene repertoires across 12 *drosophilid* species surveyed. Rates of gene turnover averaged across the herbivore clade were significantly higher than background rates in over half of the surveyed gene families. However, gene turnover was more limited along the ancestral herbivore branch, with only gustatory receptors and odorant-binding proteins experiencing strong losses. The genes most significantly impacted by gene loss, duplication, or changes in selective constraint were those involved in detecting compounds associated with feeding on living plants (bitter or electrophilic phytochemicals) or their ancestral diet (fermenting plant volatiles). These results provide insight into the molecular and evolutionary mechanisms of plant-feeding adaptations and highlight gene candidates that have also been linked to other dietary transitions in *Drosophila*.

Keywords: *Scaptomyza*, gene family evolution, genomics, trophic shift, *Drosophila*, plant-herbivore interactions, adaptation

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 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 (Strong et al. 1984; Bernays 1998; Farrell 1998). It has long been hypothesized that the diversity of herbivorous insects emerged as a result of co-diversification processes with their host plants (Mitter et al. 1988; Farrell 1998; Marvaldi et al. 2002).

Received: March 19, 2023. Accepted: May 31, 2023

© The Author(s) 2023. Published by Oxford University Press on behalf of The Genetics Society of America.

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

As proposed by Ehrlich and Raven (1964), herbivores can diversify by specializing in plants bearing the same chemical defenses through adaptations to chemical defenses (in one or a few plant families), that is, until new plant defenses evolve, allowing plants to escape and diversify under this release from herbivore pressure in this “co-evolutionary arms-race” (Ehrlich and 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 (Heidel-Fischer and Vogel 2015; Simon et al. 2015; Vertacnik and Linnen 2017). As a result, we have learned that genes that mediate interactions with plant secondary compounds, such as chemosensory and detoxification genes, likely underlie adaptive mechanisms for plant colonization (Ehrlich and Raven 1964; Berenbaum 1983; Li, Schuler et al. 2003; Futuyama and Agrawal 2009; Edger et al. 2015). Despite significant progress toward understanding the genetic basis of herbivore-plant interactions, the evolutionary processes shaping these large, complex, and rapidly evolving gene families are still not fully understood.

A principal source of the functional genetic variation that underlies dietary novelty in herbivorous arthropods arises from extensive gene family evolution. For example, the spider mite (*Tetranychus urticae*) and diamondback moth (*Plutella xylostella*) genomes harbor expanded gene families encoding enzymes involved in the detoxification of plant secondary compounds they encounter (Grbić et al. 2011; Dermauw et al. 2013; You et al. 2013). Similarly, genes encoding gustatory receptors (GRs) involved in host finding have experienced extensive lineage-specific duplications in butterflies (Briscoe et al. 2013). Despite the identification of gene family expansions and contractions in herbivorous insects (McBride 2007; Edger et al. 2015; Johnson et al. 2018), isolating herbivory as the cause remains controversial, as these changes may have resulted from subsequent specialization occurring over a hundred million years. For example, the most diverse extant herbivore lineages—butterflies and moths (Lepidoptera), as well as leaf beetles, weevils, and close relatives (Phytophaga)—arose in the late Paleozoic and early Mesozoic, respectively (Wiens et al. 2015; Kawahara et al. 2019). Parsing herbivore-specific effects from those resulting from specialization is particularly challenging given the prominent role of specialization in driving herbivore diversification rates. This is strongly supported by associations between host shifts and speciation events and by the higher species richness found among specialist herbivores compared to generalist herbivores (Futuyama and Agrawal 2009; Forister et al. 2015).

While it is unclear whether specialization on specific plant taxa evolves during or after the evolution of herbivory (Bernays 1998), many phytophagous insects nonetheless exhibit phylogenetic conservatism: associating with the same plant taxa for many millions of years (Futuyama and Agrawal 2009). Comparative genomic studies examining younger herbivore lineages would thus allow for a more refined analysis to identify herbivore-associated changes from those arising in response to specialization or other evolutionary forces (Gardiner et al. 2008; Yassin et al. 2016). Furthermore, most herbivore lineages lack the functional genetic tools necessary to examine the implications of these copy number changes. Here, we addressed these limitations by studying the evolution of herbivory within Drosophilidae using a comparative genomics approach.

Although larvae of most Drosophilidae species retain the ancestral habit of feeding on decaying plant tissue and associated microbes, herbivory has evolved several times in the lineage (Okada and Sasakawa 1956; Jones 1998; Whiteman et al. 2011;

Maunsell et al. 2017; Durkin et al. 2021). A major clade of herbivorous species in the family are members of the genus *Scaptomyza*, a monophyletic lineage of ~272 species and 21 subgenera (O’Grady and Desalle 2008). *Scaptomyza* spp. are nested within the paraphyletic subgenus *Drosophila*, which also includes Hawaiian *Drosophila* and the virilis-repleta radiation (Fig. 1a) (O’Grady and Desalle 2008; Lapoint et al. 2013; Katoh et al. 2017; O’Grady and DeSalle 2018; Church and Extavour 2022). Herbivorous *Scaptomyza* is found across the Holarctic, and *S. flava*, in particular, has been introduced into New Zealand, where it is a pest (Martin 2004). DNA barcoding revealed that the clade of herbivores may be cryptic radiation, with the divergence of 8 species in North America within the last ~10 million years (Fig. 1b) (Katoh et al. 2017; Peláez et al. 2022). This is similar in species richness to the *D. melanogaster* subgroup worldwide (9 species within ~12 million years) (David et al. 2007).

The herbivorous *Scaptomyza* species have become models for the evolution of herbivory and are an ideal group to test hypotheses about gene family evolution for several reasons. Most specialize in plants in the Brassicales (mustards and their relatives) and can be reared on the genetic model plant, *Arabidopsis thaliana* (Whiteman et al. 2011, 2012). Genetic dissection of adaptive traits in this lineage (Gloss et al. 2014; Goldman-Huertas et al. 2015; Peláez et al. 2022) 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 (O’Grady and DeSalle 2018; Finet et al. 2021), and the ability to test hypotheses using genetic tools from both *Drosophila* and *Arabidopsis* (Groen and Whiteman 2016).

A critical advantage of studying *Scaptomyza* is that the genus encompasses species exhibiting varying degrees of specialization on different plant families: *S. graminum*, which specializes in plants of the Caryophyllaceae family (pinks or carnations), and *S. flava* as well as *S. montana*, which are largely specialists on Brassicales. The Brassicales harbor nontoxic glucosinolates that are the precursors to toxic mustard oils. While *S. flava* feeds on numerous genera of Brassicales, *S. montana* is more specialized, showing a strong preference for plants with indole glucosinolates (Gloss et al. 2017). *S. flava* has also been found to attack common pea plants (*Pisum sativum*, Fabaceae) and has further expanded its host range to some Caryophyllaceae in New Zealand (Martin 2004, 2012). Genetic changes shared by these 3 species may be more likely to be associated with the initial transition to herbivory, rather than their respective subsequent specializations. We cannot discount the possibility that these changes may be driven by other processes unrelated to herbivory. However, comparisons with closely related nonherbivores, such as *S. pallida* and *S. hsui*, each from different subgenera (*Parascaptomyza* and *Hemiscaptomyza*, respectively), allow us to differentiate herbivore-specific changes from those that arose earlier along with the *Scaptomyza* lineage. These 2 species are the closest relatives of the herbivorous species (~20 million years diverged (Lapoint et al. 2013; Katoh et al. 2017; Peláez et al. 2022)) with genomic sequences available (Kim et al. 2021). Finally, nonherbivore specialists outside *Scaptomyza* can be used to further polarize patterns of specialization from those linked to herbivory. For instance, genome assemblies are available for *D. erecta*, a specialist on screw pine fruit (*Pandanus* spp.) and *D. mojavensis* that has populations specializing in rotting prickly pear cactus (*Opuntia littoralis*) (Rio et al. 1983; Pfeiler et al. 2005).

Given the salience of gene expansions and contractions associated with herbivorous insects, we hypothesized that genes involved in sensing or detoxifying plant chemical defenses would

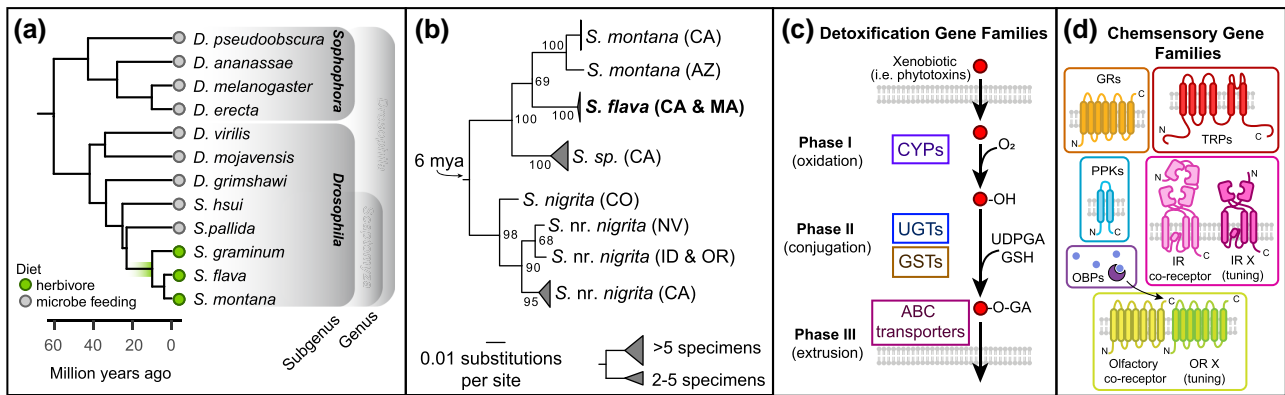


Fig. 1. Gene families at the interface of plant-herbivore interactions. a) Phylogenetic placement of herbivorous *Scaptomyza* within the paraphyletic genus *Drosophila* (Matsunaga et al. 2022). b) ML nucleotide phylogeny built using COI sequences from North American *Scaptomyza* collected feeding on mustard plants (Brassicales spp.). Individuals with <1% pairwise nucleotide divergence collapsed into clades. Collection localities indicated by 2-letter state abbreviations. For this study, the genome of an *S. flava* line from MA, USA (bolded) was sequenced and assembled. Sequences and complete phylogeny are included in Supplementary Dataset 4 in Supplementary File 1. Divergence time estimated by Whiteman et al. 2012. c) Detoxification of xenobiotics/plant secondary metabolites (phytotoxins) involves 3 phases: oxidation/reduction by CYPs, enzymatic conjugation by UGTs or GSTs, and excretion/transport out of the cells by ATP-binding cassette transporters. d) Insects detect environmental compounds mostly with 6 major chemosensory gene families: GRs, OBPs, ORs, degenerin/epithelial sodium channels (PPKs), IRs, and TRPs.

experience the greatest copy number changes. Specifically, in accordance with the neural limitation theory, we expected herbivores to lose numerous chemosensory genes, to streamline neural processing in the face of choosing between many host plants (Bernays 2001) and also their ancestral diet. We also expected to find duplicated genes with signatures of rapid protein evolution, which could have enabled early herbivores to gain new chemosensory or detoxification functions through neofunctionalization or subfunctionalization (Ohno 1970; Lynch and Conery 2000).

To address these hypotheses and predictions, we sequenced and assembled a high-quality genome sequence of *S. flava* to analyze alongside publicly available genome sequences of 2 herbivorous species (*S. montana* and *S. graminum*), 2 nonherbivores (*S. pallida* and *S. hsui*), and 7 nonherbivorous *Drosophila*, across a phylogenetic gradient from the herbivores (Fig. 1a). Following the curation of the major chemosensory and detoxification gene families (those in Fig. 1c and d), we then used maximum likelihood (ML) methods to test whether rates of gene gain and loss in herbivore branches differed from background branches. We next generated codon-based, ML models to identify genes with signatures of changing selective regimes along the branch at the base of the herbivore clade. We also evaluated whether our results may be explained by demographic events in recent history. We found that the most dramatic changes occurred in chemosensory genes involved in sensing yeast fermentation products/fruit volatiles and bitter/toxic plant chemical defenses.

Materials and methods

S. flava genome sequencing, assembly, and annotation

Full details of these methods can be found in the Supplementary Methods in Supplementary File 1.

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.

Three hundred male flies were flash frozen and stored at -80°C . SMRTbell libraries (~20 kb) 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 2 PacBio Sequel SMRT cells. A Dovetail HiC library was prepared in a similar manner as described previously (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 2×150 bp paired-end reads, approximately 30x sequence coverage.

Illumina library and sequencing

Additional Illumina sequence data was used to polish the PacBio 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 matings. Paired-end 180 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 was assembled using the long-read hybrid assembly pipeline described in Kim et al. (2021) (https://github.com/flyseq/drosophila_assembly_pipelines), which has been shown to produce highly complete genome assemblies for drosophilid flies resulting in very few coding sequence indels (Kim et al. 2021). Following these methods, we generated an initial draft assembly with Flye v2.9 with default settings (Kolmogorov et al. 2019), and identified and removed duplicated haplotypes (haplotigs) with purge_haplotigs v1.1.1 (cutoffs: 3,33,195) (Roach et al. 2018). We polished the draft assembly using the PacBio reads with one round of Racon v1.4.3 (options: -m8 -x 6 g 8 w 500) (Vaser et al. 2017), then polished further using Illumina reads with one round of Pilon v1.23 (-fix snps, indels) (Walker et al. 2014), only fixing base-level errors.

The fully polished assembly was scanned for contaminant sequences using NCBI BLAST v.2.10.0 (Johnson et al. 2008) and BlobTools (Laetsch and Blaxter 2017). 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 that was based on the PacBio and HiC libraries described above (full details included in the [Supplementary Methods](#) in [Supplementary File 1](#)). The Dovetail assembly exhibited high scaffold contiguity but contained a significant number of frameshift mutations in coding sequences that limited our ability to correctly annotate the assembly. We, therefore, scaffolded our less contiguous but more accurate assembly with the error-prone Dovetail version, reasoning that base-level errors in the previous version were unlikely to impact reference-based scaffolding. Briefly, the Dovetail assembly was performed using the FALCON 1.8.8 pipeline from PacBio with 70× SMRT data used as input. A cutoff length that corresponded to 50× coverage was used during the initial error-correcting stage, and error-corrected reads were aligned into 3,561 primary contigs. The assembly was then polished through PacBio's Arrow algorithm from SMRT Link 5.0.1, using the original raw-reads. This assembly and the HiC library were used as input for HiRise which uses proximity ligation data to scaffold genome assemblies (Putnam et al. 2016). Dovetail HiC library sequences were aligned to the input assembly using bwa (<http://github.com/lh3/bwa>). To scaffold our assembly with the Dovetail assembly, we soft-masked both genomes using the repeat library described above, 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 scaffold (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 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 the newest *S. flava* assembly (sfla_v2), using whole-genome Progressive Cactus alignments and the Comparative Annotation Toolkit (CAT (Fiddes et al. 2018)). Briefly, using the Illumina-only assembly (sfla_v1), repeat regions were masked using RepeatMasker (Smit et al. 2013) with the *Drosophila* repeat library. Protein-coding genes were annotated using MAKER2 (Holt and 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. The other 5 *Scaptomyza* assemblies 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 whole-genome alignment. CAT was used to project annotations from sfla_v1 onto the other genomes. Further details of these methods are in the [Supplementary Methods](#) in [Supplementary File 1](#).

Gene model curation and orthology inference

An iterative curation strategy was used to identify the complement of chemosensory and detoxification genes in 4 published *Drosophila* genomes (*D. melanogaster*, *D. virilis*, *D. mojavensis*, and *D. grimshawi*) ([Drosophila 12 Genomes Consortium et al. 2007](#)), 4 published *Scaptomyza* genomes (*S. pallida*, *S. hsui*, *S. graminum*, and *S. montana*) (Kim et al. 2021), and in our *S. flava* assembly

(sfla_v2). Genome assembly versions are listed in [Supplementary Table 1](#) in [Supplementary File 2](#), along with which species were included in each analysis. Gene curation included 6 chemosensory gene families (GRs, ionotropic receptors (IRs), olfactory receptors (ORs), odorant binding proteins (OBPs), pickpocket proteins (PPKs), transient receptor potential channels (TRPs)) and 3 detoxification gene families (Cytochrome P450s (CYPs), Glutathione S-transferases (GSTs), UDP-glycosyltransferases (UGTs)). 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 putatively lost genes, we performed an additional TBLASTN search (e-value cutoff $<1e-4$). Genes were considered truly absent if this yielded no hits. Gene models significantly deviating in length from *D. melanogaster* orthologs were manually inspected, and corrected using aligned homologous sequences to the annotated genes in other species. Additional validation steps we implemented are described in the [Supplementary Methods](#) in [Supplementary File 1](#). Nucleotide sequences were aligned in Geneious v.10.2.6 using MUSCLE (Edgar 2004) and manually inspected. All gene coordinates are provided in [Supplementary Dataset 1](#) in [Supplementary File 1](#).

To verify correct orthology assignments, ML gene trees were constructed from sequences from *D. grimshawi* and all *Scaptomyza*, using RAXML with default settings (Stamatakis 2006). Genes were binned into orthologous groups if they formed a clade with >70 bootstrap support. For poorly supported clades (bootstrap support <70), orthology groups were assigned based on previous orthology assignments (Low et al. 2007; Almeida et al. 2014; Good et al. 2014). Gene trees were midpoint rooted and visualized in iTOL v6 (Letunic and Bork 2021).

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 9 species mentioned above, we included 3 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 (Low et al. 2007; Almeida et al. 2014; Good et al. 2014). As input, we provided a matrix of gene counts for each orthologous gene cluster ([Supplementary Dataset 2](#) in [Supplementary File 1](#)). Orthology groups were merged until there was at least 1 homologous gene reconstructed at the base of the phylogeny as the analysis assumes at least 1 ancestral gene per group. 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. 2022).

First, we generated models of the average turnover rate (λ) (the ML estimate of 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 2 rates: 1 rate estimated for a foreground branch, and another for the remaining background branches, where the foreground

branch was the internal (ancestral) branch leading to all herbivores, a terminal branch leading to 1 of the 12 species, or the entire clade of herbivores. All models were run in triplicate, and the iteration with the highest ML probability was retained. To test whether a two-rate model was significantly better than the single-rate model, we performed likelihood ratio tests (LRTs). P-values were adjusted at a false discovery rate (FDR) of 1% using the *q*-value package in R (Storey 2008). Using the same methods, we also generated models estimating gains and losses (λ , μ) separately, rather than as a single parameter.

To test whether expansion and contraction rates of chemosensory and detoxification gene families deviated from genome-wide rates, we performed the same analyses on 200 randomly chosen orthology groups. We identified these gene sets using the OrthoVenn2 server (Wang et al. 2015; Xu et al. 2019), a web-based orthology assignment tool based on OrthoMCL (Li, Stoeckert et al. 2003). We uploaded genome-wide proteomes for the 5 *Scaptomysza* species and used protein sequences for the remaining 7 species available through OrthoVenn2 (Ensembl database, release January 2019). Default parameters were used with an *e*-value cutoff of 0.05 and an inflation value of 1.5. Clusters were randomly chosen that summed up to ~200 genes per species.

Molecular evolutionary analysis

To identify genes that experienced changes in selective pressure along the ancestral branch at the base of the herbivorous species, we used codon-based models of evolution (*codeml* program) in the phylogenetic analysis by maximum likelihood (PAML) package (Yang 2007). These models estimate the nonsynonymous (dN) to synonymous (dS) substitution rate ratio ($\omega = dN/dS$), 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 $dN/dS \approx 1$, and those experiencing negative or purifying selection $dN/dS < 1$.

For each group of orthologous genes, nucleotide sequences from 9 taxa (*S. flava*, *S. montana*, *S. graminum*, *S. pallida*, *S. hsui*, *D. grimshawi*, *D. mojavensis*, *D. virilis*, and *D. melanogaster*) were aligned in Geneious v.10.2.6 using MAFFT v7.450 translation alignment with default settings (Katoh et al. 2002; Katoh and Standley 2013). A species tree was generated for these analyses by summarizing the 100 ML-generated gene trees published by (Kim et al. 2021), using ASTRAL v.5.5.9 (Zhang et al. 2018) with default settings. For genes with multiple paralogs, we estimated gene trees using FastTreeMP v2.1 with a general time reversible model with gamma rate distribution and 100 bootstrap replicates (Price et al. 2010).

We employed branch, branch-site, and clade models within PAML. We used the branch model to assess whether background and foreground dN/dS rates differ across branches and whether this could be attributed to positive selection, reduced purifying selection, or stronger purifying selection. (Yang 1998). 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 rate over all branches (model = 0, NSites = 0). To test whether the foreground branch experienced positive selection ($dN/dS > 1$), we compared the two-ratio model to a model in which ω is fixed to 1 (model = 2, NSites = 0, fix_omega = 1).

We also used branch-site models, which more realistically model protein evolution, allowing dN/dS to vary across both

codons and branches (Yang and Nielsen 2002; Zhang et al. 2005). We compared the “branch-site model A” (model = 2, NSites = 2, fix_omega = 0) against a null model in which ω_2 is fixed to 1 (model = 2, NSites = 2, fix_omega = 1). This comparison also offers a direct test for positive selection.

Finally, we used the “Clade model C” (CmC) to estimate changing selection pressures when genes exhibited numerous paralogs within species (Weadick and Chang 2012). Based on branch-site models, CmC model allows dN/dS to vary in a proportion of sites between 2 gene clades, and can detect more subtle divergent selection, particularly after duplication events (Bielawski and Yang 2004; Weadick and Chang 2012). We tested for divergent selection by comparing the CmC model (model = 3, NSites = 2, fixed = 0) against the null model “M2a_rel” (model = 0, NSites = 22, fixed = 0), which has the same number of site classes as the CmC model but does not vary across branches. To test for positive selection on the ancestral herbivore branch, we compared the CmC model against a null model in which the ω of the divergent site class is constrained to 1 (model = 3, NSites = 2, fixed = 1) (Van Nynatten et al. 2015). For clades with foreground divergent sites significantly deviating from the null expectation, we then performed branch and branch-site tests to test which set of paralogs experienced divergent selection.

For all PAML analyses, model comparisons were made using LRTs with a χ^2 distribution and 1 degree of freedom. P-values were adjusted at a FDR of 5% using the *q*-value package in R (Storey 2008).

Results

S. flava genome assembly and annotation

To complement existing genome assemblies of nonherbivorous and herbivorous drosophilids, we sequenced, assembled, and annotated the genome of *Scaptomysza flava*, a leaf-mining specialist on mustards. Our main assembly (sfla_v2) contained 781 scaffolds with an N50 of 31.83 Mb and an assembly size of 331.7 Mb (Supplementary Table 2 in Supplementary File 2). Using gene models from other *Drosophila* species and transcriptome sequences across *S. flava* life stages, we generated annotations for the *S. flava* assembly sfla_v1, which were carried over to our main assembly sfla_v2, amounting to 12,365 predicted genes. We performed the same carry-over procedures on 4 other *Scaptomysza* species (*S. pallida*, *S. hsui*, *S. montana*, and *S. graminum*) to facilitate the annotation of their chemosensory and detoxification gene families. The completeness of the assemblies was assessed by comparing the assembly against the Benchmarking Universal Single-Copy Orthologs (BUSCO) dipteran database. The genome assembly of *S. flava* is near complete, as are the assemblies of the 4 other *Scaptomysza* species with values comparable to the published assemblies of the 7 other *Drosophila* species included in subsequent analyses (*D. melanogaster*, *D. erecta*, *D. ananassae*, *D. pseudoobscura*, *D. mojavensis*, *D. virilis*, and *D. grimshawi*) (Supplementary Fig. 1a in Supplementary File 3). 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 *Scaptomysza* species (Supplementary Fig. 1b in Supplementary File 3). 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.

Contractions and expansions 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 nonherbivorous species and herbivorous *Scaptomyza* species. We used CAFE, a program that makes ML estimates of gene copy number evolutionary rates, (Han et al. 2013) and included 9 nonherbivorous and 3 herbivorous species (all species in Fig. 1a), all of which use a range of different feeding substrates with varying degrees of specialization. While we tested whether rates were significantly higher along terminal herbivore branches and across the entire herbivore clade, compared to the remainder of the tree, we were particularly interested in significantly higher copy number changes along the ancestral herbivore branch. We manually curated 6 chemosensory gene families (GRs, IRs, ORs, OBPs, pick-pocket proteins, PPKs, TRPs) and 3 detoxification gene families (CYPs, GSTs, UGTs) through iterative BLAST searches and additional validation steps (described in the Supplementary Methods in Supplementary File 1) to confirm the loss or duplication of genes (gene coordinates: Supplementary Dataset 1 in Supplementary File 1, RAXML gene trees: Supplementary Figs. 2–10 in Supplementary File 3).

Overall, herbivorous species had smaller repertoires of detoxification and chemosensory genes than nonherbivorous species. (Fig. 2b, Supplementary Table 3 in Supplementary File 2). The rate of gene turnover was significantly higher along the ancestral herbivore branch, than the background branches when all chemosensory genes were considered ($\lambda_{\text{anc_herb}} = 0.005$; $\lambda_{\text{bkgrd}} = 0.002$; $q\text{-value} = 0.005$) but not for all detoxification genes ($\lambda_{\text{anc_herb}} = 0.005$; $\lambda_{\text{bkgrd}} = 0.002$; $q\text{-value} > 0.05$) (Fig. 2c, Supplementary Table 4 in Supplementary File 2). When individual gene families were analyzed, the ancestral herbivore branch only experienced significantly higher rates of turnover among OBPs. We also estimated rates of gene duplication and loss separately, finding that along the ancestral herbivore branch the gene duplication rate was not significantly higher for any gene family compared to background rates (Fig. 2d, Supplementary Table 5 in Supplementary File 2).

The gene loss rate, however, was significantly higher for all chemosensory genes ($\lambda_{\text{anc_herb}} = 0.008$; $\lambda_{\text{bkgrd}} = 0.002$; $q\text{-value} < 0.001$) and detoxification genes ($\lambda_{\text{anc_herb}} = 0.007$; $\lambda_{\text{bkgrd}} = 0.002$; $q\text{-value} = 0.003$) (Fig. 2e, Supplementary Table 5 in Supplementary File 2). Specifically, higher rates of gene loss along the ancestral herbivore branch were found among GRs, OBPs, and UGTs. The CAFE analysis also allowed us to identify specific genes that experienced expansions and contractions (Supplementary Table 6 in Supplementary File 2; Supplementary Fig. 11 in Supplementary File 3). Significant contractions along the ancestral herbivore branch included OBPs (*Obp18a*, *Obp58b*, and *Obp58c*) and *Or22a*, an ester-sensitive, yeast-volatile receptor. The only expansion was of *Ir67a*, which was duplicated in the ancestral herbivorous lineage.

While we found limited copy number changes along the ancestral herbivore branch, rates of gene turnover, duplication, and loss estimated across the entire herbivore clade were significantly higher than background rates for numerous gene families (gene turnover: GRs, IRs, ORs, PPKs, and GSTs (Fig. 2c); duplication: IRs, ORs, PPKs, TRPs, and GSTs (Fig. 2d); loss: GRs, IRs, OBPs, ORs, PPKs, TRPs, and GSTs (Fig. 2e)). We confirmed that it is unlikely that these elevated rates among the herbivorous species could be attributed to the longer branch length of the ancestral branch

preceding the herbivore clade. We performed another set of CAFE analyses to compare rates between 2 clades of similar age: the clade of *D. melanogaster* and *D. erecta* (28 million years old) vs the clade of *S. flava* and *S. graminum* (23 million years old). This confirmed that the former pair showed no gene families evolving at higher rates, whereas the clade of *S. flava* and *S. graminum* still showed elevated rates of turnover, loss, and gain (Supplementary Fig. 12 in Supplementary File 3).

Based on the RAXML trees for each gene family (Supplementary Figs. 2–10 in Supplementary File 3), there were many herbivore-specific losses, in addition to a few gains (summarized in Fig. 3; Supplementary Table 7 in Supplementary File 2 for gene counts). Most copy number changes shared by all herbivores were concentrated among chemosensory gene families (22 out of 27). Herbivores lost several GRs that are required for bitter reception or are expressed in bitter gustatory neurons in *D. melanogaster* (multiple paralogs of isoform A of *Gr39a* [*Gr39aA*], *Gr59a*, *Gr59d*) (Kwon et al. 2011; Dweck and Carlson 2020). Herbivorous lineages also lost *Gr68a*, which is involved in the detection of an antiaphrodisiac (Bray and 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 6 cysteine residues, and only 2 losses (*Obp18a* and *Obp56b*) were among the “classic OBPs” with 6 cysteines (Supplementary Fig. 13 in Supplementary File 3). Strikingly, nonherbivorous *Scaptomyza* has 11 Plus-C OBPs, while the herbivores possess only five. Almost all IRs 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 in all herbivores was *Ir60a*, and the only duplication was the divergent IR *Ir67a*. A single loss was found among PPKs (*ppk8*), whose function is unknown. Expression localization of *D. melanogaster* chemosensory gene orthologs that were lost or duplicated in all herbivores is presented in Supplementary Fig. 14 in Supplementary File 3.

Only 5 herbivore-specific changes in copy number were found 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 3 CYPs (*Cyp4ad1*, *Cyp4d1*, and *Cyp12g1*). While there is no ortholog of *Cyp12g1* in *D. melanogaster*, both *Cyp4ad1* and *Cyp4d1* are upregulated in response to the pyrethroid deltamethrin (Liu et al. 2020).

Genes under divergent selection in the herbivore lineage

We next used an ML approach through PAML to identify chemosensory and detoxification genes with signatures of changes in the selection regime 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 or coincident with 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. Clade models offer more sensitivity in detecting divergent selection among a clade of genes, particularly those with multiple recent paralogs undergoing complex evolution (Weadick and Chang 2012). For genes that showed evidence of divergent selection through the use of clade models ($P < 0.05$; FDR < 5%), we

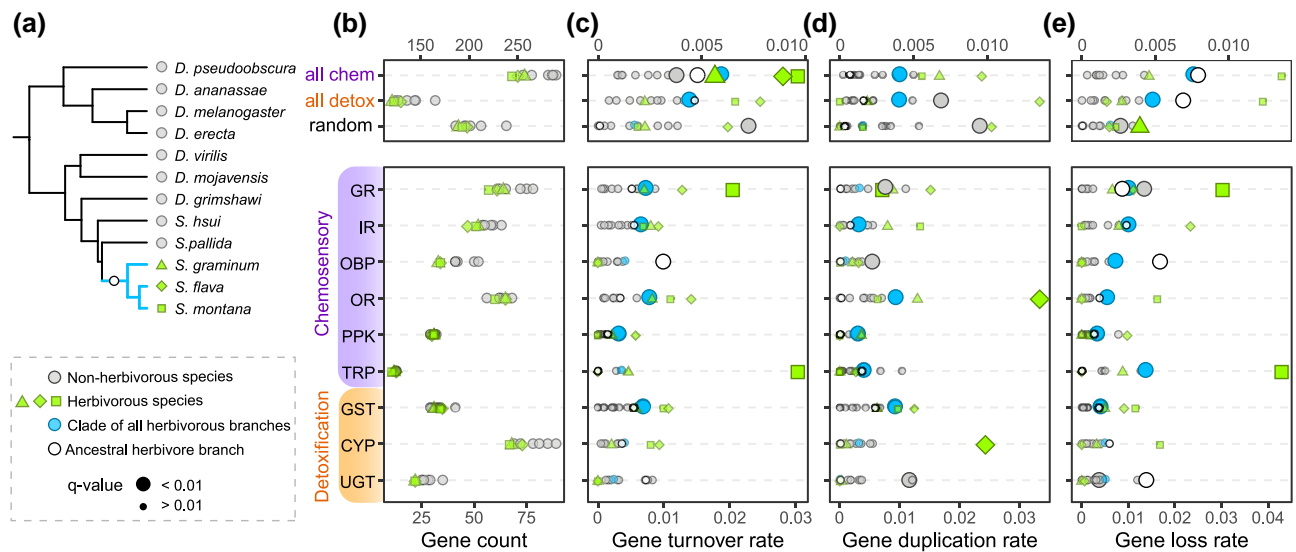


Fig. 2. The clade of *Scaptomyza* herbivores (but not their ancestral lineage) exhibit elevated gene turnover within many chemosensory and detoxification gene families. a) Phylogeny of drosophilids included in CAFE analyses: 3 herbivorous and 9 nonherbivorous *Drosophila*. b) Gene counts per gene family. Evolution of gene family size was estimated by ML in CAFE, where the rates of (c) gene turnover, λ (cumulative gains and losses), d) gene duplication, λ , and (e) gene loss, μ , were estimated for each species. Points represent the rate of a foreground branch or clade from models in which it was allowed to evolve at a rate separate from background branches. Each model was compared to a null model (single-rate estimated for the entire phylogeny) using a likelihood ratio test at an analysis-wide FDR of 1% ($q < 0.01$). Foreground branches with significantly higher rates than the rest of the phylogeny are indicated by larger shapes. Full details can be found in [Supplementary Tables 3–5](#). All chem = all chemosensory genes. All detox = all detoxification genes. Random = random set of 200 orthology groups.

additionally tested branch and branch-site models on branches leading to different paralogs to identify which paralog experienced significant changing selection pressure. We report significant results from the initial branch and branch-site tests and significant results from clade model follow-up tests. In total, 44 genes were found to have experienced a significant change in selection pressure along the basal herbivore branch ([Tables 1](#) and [Supplementary Table 8](#)). Results on all genes can be found in [Supporting Supplementary Dataset 3](#) in [Supplementary File 1](#).

Eight genes were identified as having experienced positive selection ($dN/dS > 1$) in the basal herbivore lineage: *Gr98a*, *Gr98bcd*, *GstS1b*, *Obp57cL1*, *Or85f*, *OrN2.3prime*, *Cyp4d14*, and *Ppk19* ([Table 1](#)). The proportion of positively selected sites among these genes was generally low (mean \pm SD, $4\% \pm 4.17$), with a few exceptions. One such exception was *GstS1b*, which exhibited 13% positively selected sites ($dN/dS = 38.47$), almost all of which had exceptionally strong support (i.e. high posterior probabilities, $P > 0.95$). As noted earlier, *GstS1* was duplicated in the ancestor of herbivorous *Scaptomyza*. The single ortholog in *D. melanogaster* is involved in oxidative stress response, conjugation of the lipid peroxidation product 4-hydroxynonenal, and modulating methylmercury toxicity ([Singh et al. 2001](#); [Whitworth et al. 2005](#); [Vorojeikina et al. 2017](#)). Despite the expression of both *GstS1a* and *GstS1b* in the *S. flava* larval gut ([Gloss et al. 2019](#)), positive selection on only *GstS1b* indicates potential neofunctionalization of this copy. Notably, mustard-derived toxins (isothiocyanates) induce the formation of peroxidized lipids ([Gago-Dominguez et al. 2007](#)).

Two genes (*Gr98a* and *Gr98bcd*) that experienced positive selection along the ancestral herbivore lineage are involved in the detection of noxious, “bitter” compounds. The *D. melanogaster* ortholog of *Gr98a* is involved in detecting histamine, which is high in fermented foods ([Aryal and Lee 2022](#)) and *Gr98b* detects L-canavanine, an extremely bitter plant-derived toxic amino acid ([Shim et al. 2015](#)).

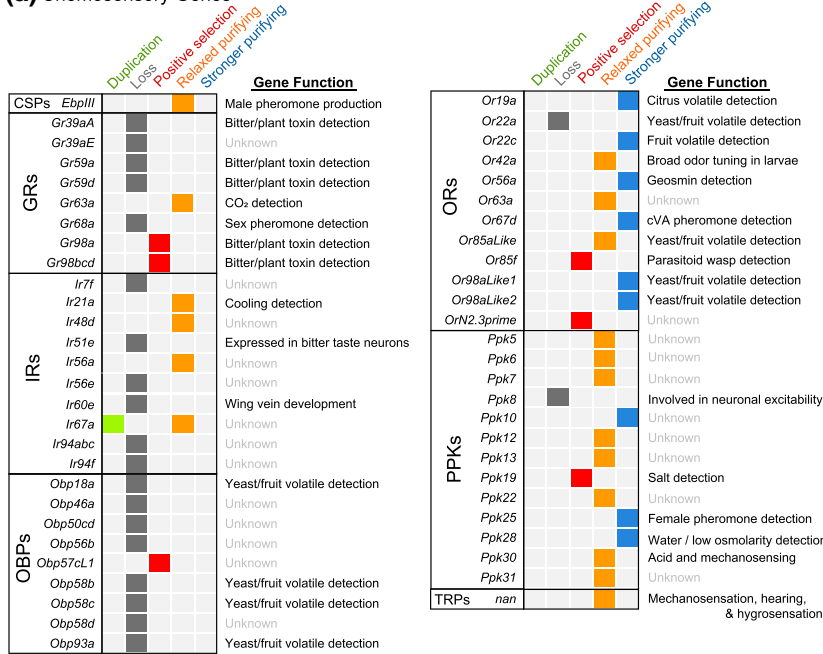
Among genes that did not show a strong indication of positive selection but were found to be evolving at significantly different rates between herbivores and background branches, many (20/35) had higher foreground dN/dS values ([Table 1](#)), indicating relaxation from selective constraints among the herbivores, whereas the remaining experienced stronger purifying selection. Many ORs 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. We found that almost all ORs experiencing relaxed purifying selection (*Or42a1*, *Or85aLike1*) or stronger purifying selection (*Or22c*, *Or98aLike1*, *Or98aLike2*) detect esters and alcohols, 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 et al. 2022](#)). Additionally, there was a stronger selection of *Or56a*, and *Or19a*, which detect geosmin and terpenes, respectively ([Stensmyr et al. 2012](#); [Dweck et al. 2013](#)).

Detoxification genes that experienced relaxed purifying selection in the ancestral herbivore lineage were *Cyp310a*, *Cyp6a22*, and *Cyp6u1*, and those that experienced stronger purifying selection were *GstE9*, *GstO2*, *Cyp28a5*, *Cyp6a16*, *Ugt301d1*, and *Ugt302c1* ([Table 1](#)). The majority of these have been found to be upregulated in response to toxin consumption, although some have putative functions in development, hormone metabolism, cold tolerance, and olfaction ([Fig. 3](#)).

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

(a) Chemosensory Genes



(b) Detoxification Genes

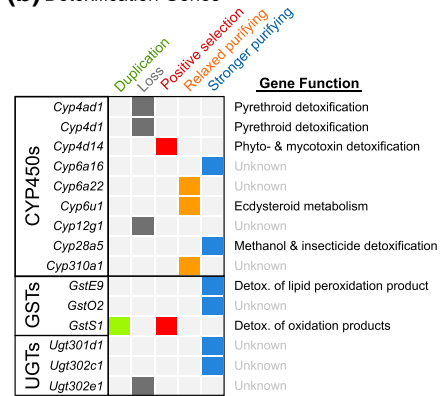


Fig. 3. Genes involved in the detection of bitter compounds and yeast/fruit odorants disproportionately experienced gene loss or divergent selection in the ancestral *Scaptomyza* herbivore lineage. Summary of (a) chemosensory and (b) detoxification genes that were duplicated, lost, or experienced divergent selection (positive selection or relaxed or stronger purifying selection). Gene duplications and losses are based on gene trees generated through RAXML, and genes that experienced changing selection pressures were identified with PAML. Gene counts from each species are shown in [Supplementary Table 7](#) in [Supplementary File 2](#). Gene functions based on published data for *D. melanogaster* orthologs (sources can be found in [Supplementary Table 10](#) in [Supplementary File 2](#)).

gene family evolution were confounded by demographic events, we examined 3 lines of evidence from our analyses.

First, the randomly selected set of orthologous clusters (~200 genes) exhibited gain, loss, and turnover rates that were not significantly higher among herbivores than in other *Drosophila* ($\lambda_{\text{herb}} = 0.002$; $\lambda_{\text{nonherb}} = 0.003$; $q\text{-value} < 0.02$, [Fig. 2](#)). These results suggest demographic processes were not the underlying cause of elevated rates of turnover among chemosensory and detoxification genes because these demographic processes would have generated similar patterns genome-wide.

Second, we inferred the level of nucleotide diversity (π) in an *S. flava* population (collected in 2013 from Belmont, MA, USA), using pooled whole-genome sequencing (methods described in the [Supplementary Methods in Supplementary File 1](#)). Autosomal nucleotide diversity, which is proportional to the coalescent effective population size, was similar between a single population of *S. flava* ($\pi = 0.0056$) and North American populations of *D. melanogaster* (the *D. melanogaster* Genetic Reference Panel; $\pi = 0.0060$ ([Mackay et al. 2012](#))). The relationship between physical distance and linkage disequilibrium was also similar between the 2 species and decayed quickly ([Peláez et al. 2022](#)), consistent with sharing similarly large coalescent effective population sizes ([Sved 1971](#)).

Third, we estimated the proportion of the genome composed of repetitive elements, which is sensitive to demographic shifts ([Bourgeois and Boissinot 2019](#)), and found that repeat content in *S. flava* is within the range observed across *Drosophila* species ([Supplementary Table 9 in Supplementary File 2](#)). These analyses do not preclude the possibility that demography has contributed to the gene family patterns observed, but they all suggest that herbivorous *Scaptomyza* do not have dramatically atypical demographic histories compared to other *Drosophila*.

Discussion

In this study, we sought to determine the extent to which the evolution of herbivory was correlated with patterns of molecular evolution across chemosensory and detoxification gene families in the *Drosophilidae*. We focused on a clade within the genus *Scaptomyza*—nested within the paraphyletic genus *Drosophila*—that evolved to feed as both leafminers (larvae) and leaf puncture-feeding adults (females) on living plants <15 mya ([Fig. 1a](#)). A great deal is known regarding how these gene families, especially chemosensory ones, evolve across *Drosophila* species in the context of diet shifts and diet specialization ([McBride 2007](#); [Gardiner et al. 2008](#); [Rane et al. 2019](#); [Reisenman et al. 2023](#)). However, we have far less insight into how these gene families evolve in response to a truly herbivorous niche shift (i.e. larval development is completed by feeding on living plants).

Here, we identified, through iterative curation and manual inspection, the full complement of several chemosensory and detoxification gene families for 5 *Scaptomyza* species—3 herbivorous species within the subgenus *Scaptomyza* (*S. flava*, *S. montana*, and *S. graminum*) and 2 nonherbivorous species from 2 other subgenera (*S. pallida* in *Parasaptomyza* 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 changes specific to the herbivorous lineages from those shared across the genus *Scaptomyza* and those associated with more recent host plant-specific specialization (*S. flava* and *S. montana* on Brassicales and *S. graminum* on Caryophyllaceae). Some limitations of this study are that only a single herbivore lineage was included and that some genetic changes may be driven by other evolutionary processes unrelated to the evolution of herbivory. We, therefore, focus our discussion

Table 1. Summary of selection analyses on chemosensory and detoxification genes.

Gene	Model	dN/dS		LRT 2(Δ lnL)	q-value (FDR 5%)
Chemosensory Genes					
<i>Csp2</i>	Branch	$\omega_0 = 0.06$	$\omega_1 = 0.24$	7.14	0.05
<i>Gr63a</i>	Branch-site	$\omega_0 = 0.03$ (81%) $\omega_{2a} = 1$ (11%)	$\omega_1 = 1$ (7%) $\omega_{2b} = 1$ (1%)	16.52	<0.01
(<i>SmonGr98a1</i> , <i>SflaGr98a1</i> , <i>SgraGr98a2</i>)	Branch	$\omega_0 = 0.29$	$\omega_1 = 0.87$	7.74	0.04
	Branch-site	$\omega_0 = 0.22$ (71%) $\omega_{2a} = 32.44$ (2%)	$\omega_1 = 1$ (26%) $\omega_{2b} = 32.44$ (1%)	11.4	0.03
<i>Gr98bcd</i>	Branch-site	$\omega_0 = 0.16$ (73%) $\omega_{2a} = 179.55$ (0.004%)	$\omega_1 = 1$ (27%) $\omega_{2b} = 179.55$ (0.001%)	10.2	0.05
<i>Ir21a</i>	Branch	$\omega_0 = 0.11$	$\omega_1 = 0.26$	9.2	0.02
<i>Ir48d</i>	Branch	$\omega_0 = 0.18$	$\omega_1 = 0.38$	7.96	0.04
(<i>SflaIr56a</i> , <i>SmonIr56a</i>)	Branch	$\omega_0 = 0.33$	$\omega_1 = 0.98$	18.8	<0.001
(<i>Sgrair56a1</i> , <i>Sgrair56a2</i> , <i>Sgrair56a4</i> , <i>Sgrair56a5</i>)	Branch	$\omega_0 = 0.33$	$\omega_1 = 3.66$	7.98	0.03
	Branch (positive)	$\omega_0 = 0.34$	$\omega_1 = 1$	1.56	0.31
<i>Ir67a</i>	Branch	$\omega_0 = 0.3$	$\omega_1 = 0.6$	10.6	0.01
<i>Obp57cL1</i>	Branch-site	$\omega_0 = 0.21$ (59%) $\omega_{2a} = 148.51$ (3%)	$\omega_1 = 1$ (36%) $\omega_{2b} = 148.51$ (2%)	15.68	<0.01
<i>Or19a</i>	Branch	$\omega_0 = 0.3$	$\omega_1 = 0.09$	13.7	<0.01
<i>Or22c</i>	Branch	$\omega_0 = 0.18$	$\omega_1 = 0.05$	7.24	0.05
(<i>SflaOr42a1</i> , <i>SmonOr42a1</i> , <i>SgraOr42a1</i>)	Branch	$\omega_0 = 0.13$	$\omega_1 = 0.45$	11.46	<0.01
<i>Or56a</i>	Branch	$\omega_0 = 0.11$	$\omega_1 = 0.03$	8.98	0.02
<i>Or63a</i>	Branch	$\omega_0 = 0.2$	$\omega_1 = 0.76$	29.5	<0.0001
<i>Or67d</i>	Branch	$\omega_0 = 0.18$	$\omega_1 = 0.03$	15.18	<0.01
<i>Or85aLike</i>	Branch	$\omega_0 = 0.13$	$\omega_1 = 0.29$	6.98	0.05
<i>Or85f</i>	Branch-site	$\omega_0 = 0.13$ (88%) $\omega_{2a} = 46.61$ (1%)	$\omega_1 = 1$ (11%) $\omega_{2b} = 46.61$ (0.001%)	12.54	0.03
<i>Or98aLike1</i>	Branch	$\omega_0 = 0.19$	$\omega_1 = 0.05$	7.58	0.04
<i>Or98aLike2</i>	Branch	$\omega_0 = 0.2$	$\omega_1 = 0$	35.92	<0.0001
<i>OrN2.3prime</i>	Branch-site	$\omega_0 = 0.16$ (73%) $\omega_{2a} = 15.3$ (5%)	$\omega_1 = 1$ (20%) $\omega_{2b} = 15.3$ (1%)	15.6	<0.01
<i>Ppk5</i>	Branch	$\omega_0 = 0.11$	$\omega_1 = 0.36$	15.16	<0.01
<i>Ppk6</i>	Branch	$\omega_0 = 0.09$	$\omega_1 = 0.25$	10.35	0.01
<i>Ppk7</i>	Branch	$\omega_0 = 0.15$	$\omega_1 = 0.27$	7.75	0.04
<i>Ppk10</i>	Branch	$\omega_0 = 0.1$	$\omega_1 = 0.01$	12.12	<0.01
<i>Ppk12</i>	Branch	$\omega_0 = 0.18$	$\omega_1 = 0.38$	10.07	0.02
<i>Ppk13</i>	Branch	$\omega_0 = 0.05$	$\omega_1 = 0.19$	16.12	<0.01
<i>Ppk19</i>	Branch-site	$\omega_0 = 0.11$ (75%) $\omega_{2a} = 15.75$ (2%)	$\omega_1 = 1$ (23%) $\omega_{2b} = 15.75$ (1%)	13.98	0.01
<i>Ppk22</i>	Branch	$\omega_0 = 0.16$	$\omega_1 = 0.37$	7.7	0.04
<i>Ppk25</i>	Branch	$\omega_0 = 0.21$	$\omega_1 = 0.04$	16.82	<0.01
<i>Ppk28</i>	Branch	$\omega_0 = 0.1$	$\omega_1 = 0.03$	11.72	<0.01
<i>Ppk30</i>	Branch	$\omega_0 = 0.2$	$\omega_1 = 0.44$	12.2	<0.01
<i>Ppk31</i>	Branch	$\omega_0 = 0.12$	$\omega_1 = 0.39$	17.23	<0.01
<i>nan</i>	Branch	$\omega_0 = 0.03$	$\omega_1 = 0.09$	10.66	0.01
Detoxification Genes					
<i>Cyp28a5</i>	Branch	$\omega_0 = 0.13$	$\omega_1 = 0.05$	9.95	0.02
<i>Cyp310a1</i>	Branch	$\omega_0 = 0.19$	$\omega_1 = 0.57$	18.43	<0.001
<i>Cyp4d14</i>	Branch-site	$\omega_0 = 0.06$ (84%) $\omega_{2a} = 809.22$ (1%)	$\omega_1 = 1$ (15%) $\omega_{2b} = 809.22$ (0.001%)	11.44	0.03
<i>Cyp6a16</i>	Branch	$\omega_0 = 0.21$	$\omega_1 = 0.1$	7.57	0.04
<i>Cyp6a22</i>	Branch	$\omega_0 = 0.07$	$\omega_1 = 0.26$	21.17	<0.001
<i>Cyp6u1</i>	Branch	$\omega_0 = 0.14$	$\omega_1 = 0.61$	20.06	<0.001
<i>GstE9</i>	Branch	$\omega_0 = 0.15$	$\omega_1 = 0.01$	9.35	0.02
<i>GstO2</i>	Branch	$\omega_0 = 0.12$	$\omega_1 = 0$	10.3	0.01
(<i>SflaGstS1b</i> , <i>SmonGstS1b</i> , <i>SgraGstS1b</i>)	Branch	$\omega_0 = 0.09$	$\omega_1 = 3.27$	50.43	<0.0001
	Branch-site	$\omega_0 = 0.08$ (84%) $\omega_{2a} = 38.47$ (13%)	$\omega_1 = 1$ (2%) $\omega_{2b} = 38.47$ (0.004%)	14.04	0.01
<i>Ugt301D1</i>	Branch	$\omega_0 = 0.09$	$\omega_1 = 0.03$	7.22	0.05
<i>Ugt302C1</i>	Branch	$\omega_0 = 0.11$	$\omega_1 = 0.04$	7.42	0.05

PAML analyses under branch and branch-site models. Branch model: M0 vs 2-ratios; branch model (positive selection): 2-ratios vs 2-ratios ($\omega_1 = 1$); branch-site model: Model A vs Model A ($\omega_2 = 1$). Omega values are reported only for the alternative model. Detailed results can be found in [Supplementary Table 8](#) in [Supplementary File 2](#).

below on candidate genetic changes identified in studies on other drosophilid herbivore and specialist lineages, which provide compelling evidence for their role in these dietary shifts.

Despite consistent reductions in gene family sizes across all 3 herbivorous species ([Fig. 2b](#), [Supplementary Table 3](#) in [Supplementary File 2](#)), we did not find significant gene turnover,

duplications, or losses for most chemosensory and detoxification gene families along the ancestral herbivore branch ([Fig. 2c–e](#)), contrary to the long-held view that the transition to herbivory involves extensive expansions and contractions of these gene families ([McBride 2007](#); [Edger et al. 2015](#); [Johnson et al. 2018](#)). The only exceptions were GRs and OBPs which were lost at a significantly

higher rate (Fig. 2e). The lack of excessive copy number changes among the remaining gene families suggests that the initial transition to herbivory requires a more limited set of changes. This implies that novel key innovations related to sensing or detoxifying plant toxins may not be necessary for the evolution of herbivory, contrary to findings in other herbivore lineages (Berenbaum et al. 1996; Wheat et al. 2007) (although other innovations, like the plant-penetrating ovipositor, may be important (Peláez et al. 2022)). Instead in some herbivore lineages, like within *Scaptomyza*, the evolution of herbivorous feeding may have depended on pre-adaptation/exaptation. Nonherbivores, including microbe-feeding drosophilids, may have already evolved adaptations to find plants and deal with plant toxins, possibly from using dead or dying plant substrates ancestrally. This contrasts with individual herbivore lineages leading to extant *Scaptomyza* species that experienced high gene turnover rates within most of the gene families surveyed. Specialization or host plant switching—processes that play a critical role in driving herbivore diversification rates (Janz et al. 2006; Hardy and Otto 2014)—instead may have been responsible for these dramatic genomic changes.

The heavy loss of GRs and OBPs is consistent with the neural limitation hypothesis (Bernays 2001). This hypothesis posits 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. While our phylogenetic analyses could not elucidate precisely whether the identified genetic changes along the ancestral branch occurred before, during, or after herbivory evolved, several of the candidate genes have been implicated in other dietary transitions (discussed below), providing strong support for their involvement in the evolution of herbivory as well.

Evolutionary patterns across GRs involved in detecting bitter compounds

Numerous studies have now shown that expansions of bitter GRs are typical of generalist arthropod species, to enable the detection of a wide variety of plant-derived bitter compounds, while oligophagous and monophagous herbivores derived from generalist ancestors tend to lose some of these bitter GRs. This pattern has been found in butterflies, moths, aphids, flies, beetles, sawflies, and mites (Smadja et al. 2009; Xu et al. 2016; Suzuki et al. 2018; Crava et al. 2020; Vertacnik et al. 2021). Our results suggest that the ancestral herbivore lineage experienced a loss in bitter detection because of losing many bitter GRs (paralogs of *Gr39aA*, *Gr59a*, *Gr59d*) and experiencing rapid evolution of other bitter GRs (*Gr98a* and *Gr98bcd*). These evolutionary genetic changes likely reduced the ability of these flies to detect bitter compounds, which would otherwise limit the intake of their host plants. A reduction in bitter sensitivity was found in *D. suzukii* (Dweck et al. 2021), which evolved herbivory (feeding on living, ripe fruit) on a similar time-scale as herbivorous *Scaptomyza* species (Suvorov et al. 2022). However, although *D. suzukii* is herbivorous, it is polyphagous in many plant families (Poyet et al. 2015), suggesting that reduction in bitter reception is a trait involved in the transition to herbivory, possibly independent of specialization. It is still not known to what extent bitter sensitivity has been reduced in herbivorous *Scaptomyza* specifically by these copy number and protein-coding changes, or whether transcriptional down-regulation of bitter GRs also plays a significant role, as it does in *D. suzukii* (Dweck et al. 2021).

With the exception of *Gr39aA*, these candidate bitter GRs do not encode “commonly expressed receptors”—a set of bitter GRs expressed in all bitter gustatory neurons (Weiss et al. 2011; Dweck

and Carlson 2020). This suggests that, despite a diet containing bitter, toxic compounds, *Scaptomyza* species are still able to detect bitter compounds generally, which is also true of specialist feeders on toxic hosts: *D. sechellia*, *D. erecta*, and *D. suzukii* (Dweck and Carlson 2020; Dweck et al. 2021). Bitter detection thus likely continues to be important for herbivores to differentiate toxin levels between individual leaves or host plants and distinguish old from young or healthy from damaged plants. Certainly, *S. flava* develops more slowly feeding on plants bearing aliphatic and indolic glucosinolates (Gloss et al. 2019), so differentiating between different bitter host-derived compounds could impact their fitness.

The lost copies of *Gr39a* (tandem duplicates of the isoform A-specific exon), *Gr59a*, and *Gr59d* are particularly interesting candidates to focus on for future study in relation to the evolution of toxin specialization, considering all 3 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 et al. 2007). The pattern of expansion of *Gr59a* and *Gr59d* in a generalist (*D. suzukii*) and contraction in specialists (*D. sechellia*, and *D. erecta*) suggests their involvement in mediating host breadth. Signals of positive selection have also been detected in *Gr59a* in *D. yakuba mayottensis*, which convergently evolved specialization on noni fruit along with *D. sechellia* (Ferreira et al. 2020). Similarly, losses of *Gr39aA* have occurred across both *D. sechellia* and *D. erecta* (McBride et al. 2007). *Gr39a* encodes several different isoforms, but only *Gr39aA* is expressed in all bitter gustatory neurons and is involved in the detection of many bitter compounds in *D. melanogaster* (Dweck and Carlson 2020). Because *D. melanogaster* only bears a single-copy of the A exon, it remains unknown whether these additional *Gr39aA* copies are expressed in species bearing multiple copies.

Notably, all 3 of these GRs (*Gr39aA*, *Gr59a*, and *Gr59d*) exhibit numerous tandem duplications. The remarkably high rate of turnover for *Gr59d* and *Gr39aA* across the sampled species (Supplementary Table 6 in Supplementary File 2) suggests that there may be transposable elements in the vicinity driving these duplications, which could be selected upon during dietary shifts to generate a dosage effect, where more copies may enable stronger bitter detection.

Concerted losses of and positive selection on genes involved in olfaction

Although we did not find excessively high levels of OR gene turnover at the base of the *Scaptomyza* herbivore clade, the ORs that were lost shared related and ecologically important functions, specifically in detecting yeast and fruit volatiles. Goldman-Huertas et al. (2015) showed that the evolution of herbivory in *S. flava* was associated with OR gene losses that reduced attraction toward their ancestral diet of yeast feeding. This was mediated through a behavioral loss of attraction to yeast volatiles, putatively ancestral feeding attractants, and with the stepwise loss of olfactory receptors (*Or22a*, *Or42b*, and *Or85d*) tuned to detect them (Goldman-Huertas et al. 2015). With full annotations from 4 other *Scaptomyza* species, we obtained results consistent with this finding, with the addition of 2 other yeast volatile-detecting ORs (*Or42a* and *Or85aLike*) that experienced relaxed purifying selection (Table 1, Supplementary Fig. 5 in Supplementary File 3). The loss of *Or22a*—which was the only OR loss shared exclusively by all herbivores in this study—strongly indicates an outsized role for *Or22a* in association with dietary shifts in conjunction with past studies on *Or22a* in various species. *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 sensitivity in *D. sechellia* to esters specific to noni fruit (*Morinda citrifolia*) (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* (Aguadé 2009; Mansourian et al. 2018); and finally, loss of sensitivity toward fermentation odors and increased sensitivity to certain leaf volatiles in the ripe fruit specialist *D. suzukii* (Keeseey et al. 2015). Collectively, this indicates the role of *Or22a* in driving attraction to fermenting plant tissues in drosophilids, which is no longer a niche used by ovipositing females of herbivorous species.

Interestingly, we also found a stronger selection on *Or56a* and *Or19a*. *Or56a* is a narrowly tuned OR in *Drosophila* to detect geosmin, an oviposition repellent produced by harmful microbes in rotting fruit (Stensmyr et al. 2012). Although we do not know whether *Or56a* still detects geosmin in *Scaptomyza*, it is possible that because herbivores require living, fresh plants for their larvae, geosmin-avoidance for oviposition may be even stronger. As for *Or19a*, which has been shown to be involved in the detection of citrus volatiles, specifically terpenes (Dweck et al. 2013), we suspect that this receptor in herbivores is under strong purifying selection because of its role in detecting other plant terpenes. Terpenes are one of the largest and most structurally diverse groups of herbivore-induced plant volatiles, and it is possible that the need to detect these compounds has resulted in a stronger purifying selection on *Or19a* in herbivorous *Scaptomyza*.

Previous divergence-based genomic 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), whereas here, we only found evidence of positive selection on *Or85f* and *OrN2.3prime* along the ancestral herbivore branch. This highlights the advantage of including additional closely related herbivorous species and nonherbivorous *Scaptomyza* species (i.e. *S. pallida* is ~22 million years diverged from the herbivores vs *D. grimshawi* at ~36 million years (Peláez et al. 2022)). In particular, there is now strong evidence that the triplication and strong positive selection on *Or67b* paralogs are 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. 2022).

The most striking herbivore-specific loss of genes occurred within the OBPs, which have been classified into 3 groups: classic OBPs that have 6 cysteine residues, minus-C with less than 6, and plus-C with more than 6 (Zhou et al. 2006). The majority of OBPs genes lost in herbivores (6/8) were among the plus-C class (Supplementary Fig. 13 in Supplementary File 3). The role of these lost plus-C OBPs is unclear, but some (*Obp58b*, *Obp58c*, and *Or85a*) are expressed only in the antennae and/or head of *D. melanogaster*, while others have been found additionally in the legs or body of adults (*Obp46a*, *Obp50c*, and *Obp93a*) (Zhou et al. 2004; Larter et al. 2016). We speculate that the reason for the excess loss of plus-C OBPs may be because they are more vulnerable to damage by plant toxins, like the Brassicales-specific isothiocyanates, which are highly electrophilic and attack nucleophilic sulfhydryl moieties of the cysteine residues (Kawakishi and Kaneko 1987).

Diverse functions of sensory genes associated with herbivory

In this study, we hypothesized that many 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 stimuli: carbon dioxide (relaxed purifying selection on *Gr63a*), salinity (positive selection on *Ppk19*), water/low osmolarity (stronger purifying selection on *Ppk28*), cooling (relaxed purifying selection on *Ir21a*), and pheromones (loss of *Gr68a*, strong purifying selection on *Ppk25*) (Bray and Amrein 2003; Liu et al. 2003; Kwon et al. 2007; Cameron et al. 2010; Ni et al. 2016). While dietary toxins play a significant role in driving the evolution of these gene families, it remains to be explored whether the leaf-mining lifestyle imposes other sensory changes. The leaf-mining larvae navigate through the leaf mesophyll, a fluid-filled cavity, where salt levels, temperature gradients, and risk of carbon dioxide poisoning may be significantly different than within decaying organic material. Additional studies will be needed to determine whether herbivorous *Scaptomyza* show altered responses to these stimuli and whether it is related to the leaf-mining/herbivorous lifestyle.

The role of gene duplications in evolving plant detoxification mechanisms

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. (2023)). 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 hypotheses and predictions that we might find numerous instances of gene duplications across chemosensory and detoxification gene families, our phylogenetic gene trees of each family revealed that duplications were more prevalent within individual herbivore species or among just the mustard specialists, with only 2 gene duplications (*GstS1* and *Ir67a*) shared by the 3 surveyed herbivores (the function of the latter is unknown in any species). This suggests that gene duplications may spur further specialization but may not play a prominent role in initial transitions to herbivory. There are many highly specialized herbivores that have evolved adaptations to plant secondary metabolites through duplicated detoxification genes (Wen et al. 2006; Fischer et al. 2008; Sehlmeier et al. 2010). But phenotypic plasticity and regulatory evolution driving changes in the expression of existing detoxification genes have also been found in other herbivores. For instance, the lepidopteran herbivore *Trichoplusia ni* (cabbage looper) responds to defenses from different plant families through the induction of different detoxification genes in their midgut (Herde and Howe 2014), while the two-spotted spider mite *Tetranychus urticae* has evolved transacting upregulation of some CYPs allowing it to feed on 250 plant families (Kurlovs et al. 2022).

Further investigation into the role of the duplication and rapid evolution of *GstS1* (Table 1) may clarify its possible involvement in enabling herbivory within the *Scaptomyza* lineage. Its *D. melanogaster* ortholog encodes an enzyme involved in detoxifying oxidation products, lipid peroxidation products, and organometallic compounds, with expression in the adult central nervous system, as well as in-flight muscle (Clayton et al. 1998; Singh et al. 2001; Whitworth et al. 2005; Saisawang et al. 2012; Vorojekina et al. 2017). *S. flava* larvae express both *GstS1* copies, but isothiocyanates, the defense compounds in mustard plants, do not induce strong up-regulation of either copy, reinforcing that this duplication is not specific to mustard feeding. Functional genetic testing in herbivorous drosophilids should reveal the fate and possible functional divergence of these 2 paralogs, specifically the specificity of each paralog toward diet-derived xenobiotics.

Conclusions

Genomic comparisons of older herbivorous lineages to distantly related nonherbivores, or across herbivorous lineages, have uncovered striking expansions and losses of genes involved in chemosensation and detoxification in arthropods (e.g. Goldman-Huertas et al. 2015; Rane et al. 2016; Calla et al. 2017; Johnson et al. 2018; Schoville et al. 2018). Yet, the lack of dense sampling among closely related taxa has, in many cases, precluded pinpointing the timing of these changes relative to the evolution of herbivory vs the evolution of host plant specificity. Here, using a comparative genomic approach across a growing number of drosophilid genome assemblies, particularly herbivorous and nonherbivorous *Scaptomyza*, we found accelerated protein evolution and gene losses among some chemosensory genes that may be linked to the evolution of herbivory. If confirmed through further testing, the patterns of evolution we identified would lend support to the hypothesis that the chemical composition of plant tissues drives herbivore genome evolution, an idea at the core of early theories on species interactions that motivated the development of co-evolutionary theory (Fraenkel 1959; Ehrlich and Raven 1964; Swain 1977; Berenbaum and Zangerl 2008). Similar comparative approaches in other young herbivore lineages may reveal the extent to which the genomic changes tied to herbivory in *Scaptomyza* species reflect general strategies underpinning the evolution of herbivory.

Data availability

The *S. flava* Illumina-only genome assembly (sfla_v1) is available as GenBank assembly accession GCA_003952975.1, and the PacBio/HiC/Illumina assembly (sfla_v2) has been deposited at GenBank under the accession JARNME000000000. Source data, scripts, and analysis output files are accessible as [Supplementary Dataset 1–4](#) in [Supplementary File 1](#) in the Dryad repository (doi:10.6078/D1841H); for further details, see “List of Supporting Datasets” in the [Supplementary Materials](#) in [Supplementary File 1](#).

[Supplemental material](#) available at G3 online.

Acknowledgements

We wish to acknowledge Kamalakar Chatla and Dovetail Genomics for assistance in assembling the *Scaptomyza flava* genome (sfla_v2). We also thank Aya Takahashi for sharing early sequence data for the *Scaptomyza graminum* genome assembly. Finally, we thank Susan Bernstein and Erika LaPlante for their assistance in rearing the *Scaptomyza flava* that was used for whole-genome sequencing.

Funding

This work was supported by the National Science Foundation (NSF) Graduate Research Fellowships to J.N.P., A.D.G., B.G.H., K.I.V., J.A., and T.K.O.; NSF award to A.D.G. (DEB 1405966); NSF DDIG to B.G.H. (DEB-1601355); NSF award to A.D.L.N. (IOS 1758532); NSF IGERT program traineeship to A.D.G., B.G.H., and T.K.O. (DGE-0654435); the National Institutes of Health for A.C.N.D., R.T.L., and D.H.H. (PERT fellowship 5K12GM000708-13); the United States Department of Agriculture to P.D.N. (AFRI Grant No. 2012-67012-19883); to S.C.G. by University of California Riverside start-up funds; and to N.K.W. by the National Geographic Society (9097-12), the NSF (DEB-1256758),

the John Templeton Foundation (#41855), and the National Institute of General Medical Sciences of the NIH (R35GM119816).

Conflicts of interest statement

The author(s) declare no conflict of interest.

Literature cited

- Aguadé M. Nucleotide and copy-number polymorphism at the odorant receptor genes Or22a and Or22b in *Drosophila melanogaster*. *Mol Biol Evol.* 2009;26(1):61–70. doi:10.1093/molbev/msn227.
- Almeida FC, Sánchez-Gracia A, Campos JL, Rozas J. Family size evolution in *Drosophila* chemosensory gene families: a comparative analysis with a critical appraisal of methods. *Genome Biol Evol.* 2014;6(7):1669–1682. doi:10.1093/gbe/evu130.
- Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 1997; 25(17):3389–3402. doi:10.1093/nar/25.17.3389.
- Armstrong J, Hickey G, Diekhans M, Fiddes IT, Novak AM, Deran A, Fang Q, Xie D, Feng S, Stiller J, et al. Progressive Cactus is a multiple-genome aligner for the thousand-genome era. *Nature.* 2020;587(7833):246–251. doi:10.1038/s41586-020-2871-y.
- Aryal B, Lee Y. Histamine avoidance through three gustatory receptors in *Drosophila melanogaster*. *Insect Biochem Mol Biol.* 2022;144: 103760. doi:10.1016/j.ibmb.2022.103760.
- Becher PG, Flick G, Rozpędowska E, Schmidt A, Hagman A, Lebreton S, Larsson MC, Hansson BS, Piškur J, Witzgall P, et al. Yeast, not fruit volatiles mediate *Drosophila melanogaster* attraction, oviposition and development. *Funct Ecol.* 2012;26(4):822–828. doi:10.1111/j.1365-2435.2012.02006.x.
- Berenbaum M. Coumarins and caterpillars: a case for coevolution. *Evolution.* 1983;37(1):163–179. doi:10.2307/2408184.
- Berenbaum MR, Favret C, Schuler MA. On defining “key innovations” in an adaptive radiation: cytochrome P450S and papilionidae. *Am Nat.* 1996;148:S139–S155. doi:10.1086/285907.
- Berenbaum MR, Zangerl AR. Facing the future of plant-insect interaction research: le retour à la “raison d’être”. *Plant Physiol.* 2008;146(3):804. doi:10.1104/pp.107.113472.
- Bernays EA. Evolution of feeding behavior in insect herbivores: success seen as different ways to eat without being eaten. *Bioscience.* 1998;48(1):35–44. doi:10.2307/1313226.
- Bernays EA. Neural limitations in phytophagous insects: implications for diet breadth and evolution of host affiliation. *Annu Rev Entomol.* 2001;46(1):703–727. doi:10.1146/annurev.ento.46.1.703.
- Bielawski JP, Yang Z. A maximum likelihood method for detecting functional divergence at individual codon sites, with application to gene family evolution. *J Mol Evol.* 2004;59(1):121–132. doi:10.1007/s00239-004-2597-8.
- Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for illumina sequence data. *Bioinformatics.* 2014;30(15):2114–2120. doi:10.1093/bioinformatics/btu170.
- Bourgeois Y, Boissinot S. On the population dynamics of junk: a review on the population genomics of transposable elements. *Genes (Basel).* 2019;10(6):419. doi:10.3390/genes10060419.
- Bray S, Amrein H. A putative *Drosophila* pheromone receptor expressed in male-specific taste neurons is required for efficient courtship. *Neuron.* 2003;39(6):1019–1029. doi:10.1016/S0896-6273(03)00542-7.
- Briscoe AD, Macias-Muñoz A, Kozak KM, Walters JR, Yuan F, Jamie GA, Martin SH, Dasmahapatra KK, Ferguson LC, Mallet J, et al.

- Female behaviour drives expression and evolution of gustatory receptors in butterflies. *PLoS Genet.* 2013;9(7):e1003620. doi:10.1371/journal.pgen.1003620.
- Calla B, Noble K, Johnson RM, Walden KKO, Schuler MA, Robertson HM, Berenbaum MR. Cytochrome P450 diversification and host-plant utilization patterns in specialist and generalist moths: birth, death and adaptation. *Mol Ecol.* 2017;26(21):6021–6035. doi:10.1111/mec.14348.
- Cameron P, Hiroi M, Ngai J, Scott K. The molecular basis for water taste in *Drosophila*. *Nature.* 2010;465(7294):91–95. doi:10.1038/nature09011.
- Carscadden KA, Batstone RT, Hauser FE. Origins and evolution of biological novelty. *Biol Rev.* 2023. doi:10.1111/brv.12963.
- Church SH, Extavour CG. Phylotranscriptomics reveals discordance in the phylogeny of Hawaiian *Drosophila* and *Scaptomyza* (Diptera: Drosophilidae). *Mol Biol Evol.* 2022;39(3):msac012. doi:10.1093/molbev/msac012.
- Drosophila* 12 Genomes Consortium; Clark AG, Eisen MB, Smith DR, Bergman CM, Oliver B, Markow TA, Kaufman TC, Kellis M, Gelbart W, et al. Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature.* 2007;450(7167):203–218. doi:10.1038/nature06341.
- Clayton JD, Cripps RM, Sparrow JC, Bullard B. Interaction of troponin-H and glutathione S-transferase-2 in the indirect flight muscles of *Drosophila melanogaster*. *J Muscle Res Cell Motil.* 1998;19(2):117–127. doi:10.1023/A:1005304527563.
- Crava CM, Zanini D, Amati S, Sollai G, Crnjar R, Paoli M, Rossi-Stacconi MV, Rota-Stabelli O, Tait G, Haase A, et al. Structural and transcriptional evidence of mechanotransduction in the *Drosophila suzukii* ovipositor. *J Insect Physiol.* 2020;125:104088. doi:10.1016/j.jinsphys.2020.104088.
- David JR, Lemeunier F, Tsacas L, Yassin A. The historical discovery of the nine species in the *Drosophila melanogaster* species subgroup. *Genetics.* 2007;177(4):1969–1973. doi:10.1534/genetics.104.84756.
- Dekker T, Ibba I, Siju KP, Stensmyr MC, Hansson BS. Olfactory shifts parallel superspecialism for toxic fruit in *Drosophila melanogaster* sibling. *D sechellia*. *Curr Biol.* 2006;16(1):101–109. doi:10.1016/j.cub.2005.11.075.
- Dermauw W, Osborne EJ, Clark RM, Grbić M, Tirry L, Van Leeuwen T. A burst of ABC genes in the genome of the polyphagous spider mite *tetranychus urticae*. *BMC Genomics.* 2013;14(1):317. doi:10.1186/1471-2164-14-317.
- Dulai KS, von Dornum M, Mollon JD, Hunt DM. The evolution of trichromatic color vision by opsin gene duplication in new world and old world primates. *Genome Res.* 1999;9(7):629–638. doi:10.1101/gr.9.7.629.
- Durkin SM, Chakraborty M, Abrieux A, Lewald KM, Gadau A, Svetec N, Peng J, Kopyto M, Langer CB, Chiu JC, et al. Behavioral and genomic sensory adaptations underlying the pest activity of *Drosophila suzukii*. *Mol Biol Evol.* 2021;38(6):2532–2546. doi:10.1093/molbev/msab048.
- Dweck HK, Talross GJ, Wang W, Carlson JR. Evolutionary shifts in taste coding in the fruit pest *Drosophila suzukii*. *Elife.* 2021;10:e64317. doi:10.7554/eLife.64317.
- Dweck HKM, Carlson JR. Molecular logic and evolution of bitter taste in *Drosophila*. *Curr Biol.* 2020;30(1):17–30.e3. doi:10.1016/j.cub.2019.11.005.
- Dweck HKM, Ebrahim SAM, Kromann S, Bown D, Hillbur Y, Sachse S, Hansson BS, et al. Olfactory preference for egg laying on citrus substrates in *Drosophila*. *Curr Biol.* 2013;23(24):2472–2480. doi:10.1016/j.cub.2013.10.047
- Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 2004;32(5):1792–1797. doi:10.1093/nar/gkh340.
- Edger PP, Heidel-Fischer HM, Bekaert M, Rota J, Glöckner G, Platts AE, Heckel DG, Der JP, Wafula EK, et al. The butterfly plant arms-race escalated by gene and genome duplications. *Proc Natl Acad Sci U S A.* 2015;112(27):8362–8366. doi:10.1073/pnas.1503926112.
- Ehrlich PR, Raven PH. Butterflies and plants: a study in coevolution. *Evolution.* 1964;18(4):586–608. doi:10.2307/2406212.
- Farrell BD. Inordinate fondness” explained: why are there so many beetles? *Science.* 1998;281(5376):555–559. doi:10.1126/science.281.5376.555.
- Ferreira EA, Lambert S, Verrier T, Marion-Poll F, Yassin A. Soft selective sweep on chemosensory genes correlates with ancestral preference for toxic noni in a specialist *Drosophila* population. *Genes (Basel).* 2020;12(1):32. doi:10.3390/genes12010032.
- Fiddes IT, Armstrong J, Diekhans M, Nachtweide S, Kronenberg ZN, Underwood JG, Gordon D, Earl D, Keane T, Eichler EE, et al. Comparative annotation toolkit (CAT)—simultaneous clade and personal genome annotation. *Genome Res.* 2018;28(7):1029–1038. doi:10.1101/gr.233460.117.
- Finet C, Kassner VA, Carvalho AB, Chung H, Day JP, et al. *Drosophila*: resources for drosophilid phylogeny and systematics. *Genome Biol Evol.* 2021;13(8):evab179. doi:10.1093/gbe/evab179.
- Fischer HM, Wheat CW, Heckel G, Vogel H. Evolutionary origins of a novel host plant detoxification gene in butterflies. *Mol Bio Evol.* 2008;25(5):809–820. doi:10.1093/molbev/msn014.
- Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, Smit AF. Repeatmodeler2 for automated genomic discovery of transposable element families. *Proc Natl Acad Sci U S A.* 2020;117(17):9451–9457. doi:10.1073/pnas.1921046117.
- Forister ML, Novotny V, Panorska AK, Bajé L, Basset Y, Butterill PT, Cizek L, Coley PD, Dem F, Diniz IR, et al. The global distribution of diet breadth in insect herbivores. *Proc Natl Acad Sci U S A.* 2015;112(2):442–447. doi:10.1073/pnas.1423042112.
- Fraenkel GS. The raison d’être of secondary plant substances. *Science.* 1959;129(3361):1466–1470. doi:10.1126/science.129.3361.1466.
- Futuyma DJ, Agrawal AA. Macroevolution and the biological diversity of plants and herbivores. *Proc Natl Acad Sci U S A.* 2009;106(43):18054–18061. doi:10.1073/pnas.0904106106.
- Gago-Dominguez M, Jiang X, Castelao JE. Lipid peroxidation, oxidative stress genes and dietary factors in breast cancer protection: a hypothesis. *Breast Cancer Res.* 2007;9(1):201. doi:10.1186/bcr1628.
- Gardiner A, Barker D, Butlin RK, Jordan WC, Ritchie MG. *Drosophila* chemoreceptor gene evolution: selection, specialization and genome size. *Mol Ecol.* 2008;17(7):1648–1657. doi:10.1111/j.1365-294X.2008.03713.x.
- Gloss AD, Brachi B, Feldmann MJ, Groen SC, Bartoli C, Gouzy J, LaPlante ER, Meyer CG, Pyon HS, Rogan SC, et al. Genetic variants affecting plant size and chemical defenses jointly shape herbivory in *Arabidopsis*. *bioRxiv* 156299. <https://doi.org/10.1101/156299>, 27 June 2017, preprint: not peer reviewed.
- Gloss AD, Nelson Dittrich AC, Lapoint RT, Goldman-Huertas B, Verster KI, Pelaez JL, Nelson ADL, Aguilar J, Armstrong E, Charboneau JLM, et al. Evolution of herbivory remodels a *Drosophila* genome. *bioRxiv* 767160. <https://doi.org/10.1101/767160>, 12 September 2019, preprint: not peer reviewed.
- Gloss AD, Vassão DG, Hailey AL, Nelson Dittrich AC, Schramm K, Reichelt M, Rast TJ, Weichsel A, Cravens MG, Gershenzon J, et al. Evolution in an ancient detoxification pathway is coupled with a transition to herbivory in the Drosophilidae. *Mol Biol Evol.* 2014;31(9):2441–2456. doi:10.1093/molbev/msu201.
- Goldman-Huertas B, Mitchell RF, Lapoint RT, Faucher CP, Hildebrand JG, Whiteman NK. Evolution of herbivory in Drosophilidae linked to loss of behaviors, antennal responses,

- odorant receptors, and ancestral diet. *Proc Natl Acad Sci U S A*. 2015;112(10):3026–3031. doi:10.1073/pnas.1424656112.
- Good RT, Gramzow L, Battlay P, Sztal T, Batterham P, Robin C. The molecular evolution of cytochrome P450 genes within and between *Drosophila* species. *Genome Biol Evol*. 2014;6(5):1118–1134. doi:10.1093/gbe/evu083.
- Grbić M, Van Leeuwen T, Clark RM, Rombauts S, Rouzé P, Grbić V, Osborne EJ, Dermauw W, Thi Ngoc PC, Ortego F, et al. The genome of *tetranychus urticae* reveals herbivorous pest adaptations. *Nature*. 2011;479(7374):487–492. doi:10.1038/nature10640.
- Groen SC, Whiteman NK. Using *Drosophila* to study the evolution of herbivory and diet specialization. *Curr Opin Insect Sci*. 2016;14:66–72. doi:10.1016/j.cois.2016.01.004.
- Hallam EA, Ho MG, Carlson JR. The molecular basis of odor coding in the *Drosophila* antenna. *Cell*. 2004;117(7):965–979. doi:10.1016/j.cell.2004.05.012.
- Han MV, Thomas GWC, Lugo-Martinez J, Hahn MW. Estimating gene gain and loss rates in the presence of error in genome assembly and annotation using CAFE 3. *Mol Biol Evol*. 2013;30(8):1987–1997. doi:10.1093/molbev/mst100.
- Hardy NB, Otto SP. Specialization and generalization in the diversification of phytophagous insects: tests of the musical chairs and oscillation hypotheses. *Proc Biol Sci*. 2014;281(1795):20132960. doi:10.1098/rspb.2013.2960.
- Heidel-Fischer HM, Vogel H. Molecular mechanisms of insect adaptation to plant secondary compounds. *Curr Opin Insect Sci*. 2015;8:8–14. doi:10.1016/j.cois.2015.02.004.
- Herde M, Howe GA. Host plant-specific remodeling of midgut physiology in the generalist insect herbivore *Trichoplusia ni*. *Insect Biochem Mol Biol*. 2014;50:58–67. doi:10.1016/j.ibmb.2014.03.013.
- Hickner PV, Rivaldi CL, Johnson CM, Siddappaji M, Raster GJ, Syed Z. The making of a pest: insights from the evolution of chemosensory receptor families in a pestiferous and invasive fly, *Drosophila suzukii*. *BMC Genomics*. 2016;17(1):648. doi:10.1186/s12864-016-2983-9.
- Holt C, Yandell M. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics*. 2011;12(1):2185. doi:10.1186/1471-2105-12-491.
- Janz N, Nylin S, Wahlberg N. Diversity begets diversity: host expansions and the diversification of plant-feeding insects. *BMC Evol Biol*. 2006;6(1):4. doi:10.1186/1471-2148-6-4.
- Johnson RM, Harpur BA, Dogantzis KA, Zayed A, Berenbaum MR. Genomic footprint of evolution of eusociality in bees: floral food use and CYPome “blooms”. *Insectes Soc*. 2018;65(3):445–454. doi:10.1007/s00040-018-0631-x.
- Johnson M, Zaretskaya I, Raytselis Y, Merezuk Y, McGinnis S, Madden TL. NCBI BLAST: a better web interface. *Nucleic Acids Res*. 2008;36(Web Server issue):W5–W9. doi:10.1093/nar/gkn201.
- Jones CD. The genetic basis of *Drosophila sechellia*'s Resistance to a host plant toxin. *Genetics*. 1998;149(4):1899–1908. doi:10.1093/genetics/149.4.1899.
- Katoh T, Izumitani HF, Yamashita S, Watada M. Multiple origins of Hawaiian drosophilids: phylogeography of *Scaptomyza hardy* (Diptera: Drosophilidae). *Entomol Sci*. 2017;20(1):33–44. doi:10.1111/ens.12222.
- Katoh K, Misawa K, Kuma K-I, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res*. 2002;30(14):3059–3066. doi:10.1093/nar/gkf436.
- Katoh K, Standley DM. MAFFT Multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 2013;30(4):772–780. doi:10.1093/molbev/mst010.
- Kawahara AY, Plotkin D, Espeland M, Meusemann K, Toussaint EFA, Donath A, Gimmich F, Frandsen PB, Zwick A, dos Reis M, et al. Phylogenomics reveals the evolutionary timing and pattern of butterflies and moths. *Proc Natl Acad Sci U S A*. 2019;116(45):22657–22663. doi:10.1073/pnas.1907847116.
- Kawakishi S, Kaneko T. Interaction of proteins with allyl isothiocyanate. *J Agric Food Chem*. 1987;35(1):85–88. doi:10.1021/jf00073a020.
- Keeseey IW, Knaden M, Hansson BS. Olfactory specialization in *Drosophila suzukii* supports an ecological shift in host preference from rotten to fresh fruit. *J Chem Ecol*. 2015;41(2):121–128. doi:10.1007/s10886-015-0544-3.
- Kim BY, Wang J, Miller DE, Barmina O, Delaney EK, Thompson A, Comeault AA, Peede D, D'Agostino ERR, Pelaez J, et al. Highly contiguous assemblies of 101 drosophilid genomes. *Elife*. 2021;10:e66405. doi:10.7554/eLife.66405.
- Kolmogorov M, Raney B, Paten B, Pham S. Ragout—a reference-assisted assembly tool for bacterial genomes. *Bioinformatics*. 2014;30(12):i302–i309. doi:10.1093/bioinformatics/btu280.
- Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol*. 2019;37(5):540–546. doi:10.1038/s41587-019-0072-8.
- Kurlovs AH, De Beer B, Ji M, Vandenhole M, De Meyer T, Feyereisen R, Clark RM, Van Leeuwen T. Abundant trans-driven variation in detoxification gene expression in the extreme generalist herbivore *Tetranychus urticae*. *bioRxiv* 499853. <https://doi.org/10.1101/2022.07.13.499853>, 13 July 2022, preprint: not peer reviewed.
- Kwon JY, Dahanukar A, Weiss LA, Carlson JR. The molecular basis of CO₂ reception in *Drosophila*. *Proc Natl Acad Sci U S A*. 2007;104(9):3574–3578. doi:10.1073/pnas.0700079104.
- Kwon JY, Dahanukar A, Weiss LA, Carlson JR. Molecular and cellular organization of the taste system in the *Drosophila larva*. *J Neurosci*. 2011;31(43):15300–15309. doi:10.1523/JNEUROSCI.3363-11.2011.
- Laetsch DR, Blaxter ML. Blobtools: interrogation of genome assemblies. *F1000Res*. 2017;6:1287. doi:10.12688/f1000research.12232.1.
- Lapoint RT, O'Grady PM, Whiteman NK. Diversification and dispersal of the Hawaiian Drosophilidae: the evolution of *Scaptomyza*. *Mol Phylogenet Evol*. 2013;69(1):95–108. doi:10.1016/j.ympev.2013.04.032.
- Larter NK, Sun JS, Carlson JR. Organization and function of *Drosophila* odorant binding proteins. *Elife*. 2016;5:e20242. doi:10.7554/eLife.20242.
- Letunic I, Bork P. Interactive tree of life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res*. 2021;49(W1):W293–W296. doi:10.1093/nar/gkab301.
- Li W, Schuler MA, Berenbaum MR. Diversification of furanocoumarin-metabolizing cytochrome P450 monooxygenases in two papilionids: specificity and substrate encounter rate. *Proc Natl Acad Sci U S A*. 2003;100(Suppl. 2):14593–14598. doi:10.1073/pnas.1934643100.
- Li L, Stoeckert CJ Jr, Roos DS. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res*. 2003;13(9):2178–2189. doi:10.1101/gr.1224503.
- Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragozcy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*. 2009;326(5950):289–293. doi:10.1126/science.1181369.
- Linz J, Baschwitz A, Strutz A, Dweck HK, Sachse S, Hansson BS, Stensmyr MC. Host plant-driven sensory specialization in *Drosophila erecta*. *Proc Biol Sci*. 2013;280(1760):20130626. doi:10.1098/rspb.2013.0626.

- Liu Y, Ge M, Zhang T, Chen L, Xing Y, Liu L, Li F, Cheng L. Exploring the correlation between deltamethrin stress and Keap1-Nrf2-ARE pathway from *Drosophila melanogaster* RNASeq data. *Genomics*. 2020;112(2):1300–1308. doi:10.1016/j.ygeno.2019.07.019.
- Liu L, Leonard AS, Motto DG, Feller MA, Price MP, Johnson WA, Welsh MJ. Contribution of *Drosophila* DEG/ENaC genes to salt taste. *Neuron*. 2003;39(1):133–146. doi:10.1016/S0896-6273(03)00394-5.
- Low WY, Ng HL, Morton CJ, Parker MW, Batterham P, Robin C. Molecular evolution of glutathione S-transferases in the genus *Drosophila*. *Genetics*. 2007;177(3):1363–1375. doi:10.1534/genetics.107.075838.
- Lynch VJ. Inventing an arsenal: adaptive evolution and neofunctionalization of snake venom phospholipase A2 genes. *BMC Evol Biol*. 2007;7:2. doi:10.1186/1471-2148-7-2.
- Lynch M, Conery JS. The evolutionary fate and consequences of duplicate genes. *Science*. 2000;290(5494):1151–1155. doi:10.1126/science.290.5494.1151.
- Mackay TFC, Richards S, Stone EA, Barbadilla A, Ayroles JF, Zhu D, Casillas S, Han Y, Magwire MM, Cridland JM, et al. The *Drosophila melanogaster* genetic reference panel. *Nature*. 2012;482(7384):173–178. doi:10.1038/nature10811.
- Mansourian S, Enjin A, Jirle EV, Ramesh V, Rehmann G, Becher PG, Pool JE, Stensmyr MC. Wild African *Drosophila melanogaster* are seasonal specialists on marula fruit. *Curr Biol*. 2018;28(24):3960–3968.e3. doi:10.1016/j.cub.2018.10.033.
- Martin NA. History of an invader, *Scaptomyza flava* (Fallen, 1823) (Diptera: Drosophilidae). *N Z J Zool*. 2004;31(1):27–32. doi:10.1080/03014223.2004.9518356.
- Martin NA. Ecological studies of *Scaptomyza flava* (Fallen, 1823) (Diptera: Drosophilidae) and its parasitoids. *N Z Entomol*. 2012;35(1):58–67. doi:10.1080/00779962.2012.649708.
- Marvaldi AE, Sequeira AS, O'Brien CW, Farrell BD. Molecular and morphological phylogenetics of weevils (coleoptera, curculionidae): do niche shifts accompany diversification? *Syst Biol*. 2002;51(5):761–785. doi:10.1080/10635150290102465.
- Matsunaga T, Reisenman CE, Goldman-Huertas B, Brand P, Miao K, Suzuki HC, Verster KI, Ramírez SR, Whiteman NK. Evolution of olfactory receptors tuned to mustard oils in herbivorous Drosophilidae. *Mol Biol Evol*. 2022;39(2):msab362. doi:10.1093/molbev/msab362.
- Maunsell SC, Burwell CJ, Morris RJ, McDonald WJF, Edwards ED, Oberprieler RG, Kitching RL. Host-plants of leaf-miners in Australian subtropical rainforest. *Aust Entomol*. 2017;56(4):403–411. doi:10.1111/aen.12252.
- McBride CS. Rapid evolution of smell and taste receptor genes during host specialization in *Drosophila sechellia*. *Proc Natl Acad Sci U S A*. 2007;104(12):4996–5001. doi:10.1073/pnas.0608424104.
- McBride CS, Arguello JR, O'Meara BC. Five *Drosophila* genomes reveal nonneutral evolution and the signature of host specialization in the chemoreceptor superfamily. *Genetics*. 2007;177(3):1395–1416. doi:10.1534/genetics.107.078683.
- Mitter C, Farrell B, Wiegmann B. The phylogenetic study of adaptive zones: has phytophagy promoted insect diversification? *Am Nat*. 1988;132(1):107–128. doi:10.1086/284840.
- Ni L, Klein M, Svec KV, Budelli G, Chang EC, Ferrer AJ, Benton R, Samuel ADT, Garrity PA. The ionotropic receptors IR21a and IR25a mediate cool sensing in *Drosophila*. *Elife*. 2016;5:e13254. doi:10.7554/eLife.13254.
- O'Grady P, Desalle R. Out of Hawaii: the origin and biogeography of the genus *Scaptomyza* (Diptera: Drosophilidae). *Biol Lett*. 2008;4(2):195–199. doi:10.1098/rsbl.2007.0575.
- O'Grady PM, DeSalle R. Phylogeny of the genus *Drosophila*. *Genetics*. 2018;209(1):1–25. doi:10.1534/genetics.117.300583.
- Ohno S. *Evolution by Gene Duplication*. New York: Springer Science & Business Media; 1970. doi:10.1007/978-3-642-86659-3.
- Okada T, Sasakawa M. Leaf-mining species of Drosophilidae (Diptera). *Akitu*. 1956;5:25–28.
- Peláez JN, Gloss AD, Ray JF, Chaturvedi S, Haji D, Charboneau JLM, Verster KI, Whiteman NK. Evolution and genomic basis of the plant-penetrating ovipositor: a key morphological trait in herbivorous Drosophilidae. *Proc Biol Sci*. 2022;289:20221938. doi:10.1098/rspb.2022.1938.
- Pfeiler E, Reed LK, Markow TA. Inhibition of alcohol dehydrogenase after 2-propanol exposure in different geographic races of *Drosophila mojavensis*: lack of evidence for selection at the Adh-2 locus. *J Exp Zool B Mol Dev Evol*. 2005;304(2):159–168. doi:10.1002/jez.b.21034.
- Poyet M, Le Roux V, Gibert P, Meirland A, Prévost G, et al. The wide potential trophic niche of the Asiatic fruit fly *Drosophila suzukii*: the key of its invasion success in temperate Europe? *PLoS One*. 2015;10(11):e0142785. doi:10.1371/journal.pone.0142785.
- Price MN, Dehal PS, Arkin AP. Fasttree 2—approximately maximum-likelihood trees for large alignments. *PLoS One*. 2010;5(3):e9490. doi:10.1371/journal.pone.0009490.
- Putnam NH, O'Connell BL, Stites JC, Rice BJ, Blanchette M, Calef R, Troll CJ, et al. Chromosome-scale shotgun assembly using an in vitro method for long-range linkage. *Genome Res*. 2016;26(3):342–350. doi:10.1101/gr.193474.115.
- Rane RV, Clarke DF, Pearce SL, Zhang G, Hoffmann AA, Oakeshott JG. Detoxification genes differ between cactus-, fruit-, and flower-feeding *Drosophila*. *J Hered*. 2019;110(1):80–91. doi:10.1093/jhered/esy058.
- Rane RV, Walsh TK, Pearce SL, Jermini LS, Gordon KH, Richards S, Oakeshott JG. Are feeding preferences and insecticide resistance associated with the size of detoxifying enzyme families in insect herbivores? *Curr Opin Insect Sci*. 2016;13:70–76. doi:10.1016/j.cois.2015.12.001.
- Reisenman CE, Wong J, Vedagarbha N, Lively C, Scott K. Taste adaptations associated with host specialization in the specialist *Drosophila sechellia*. *J Exp Biol*. 2023;226(3):jeb244641. doi:10.1242/jeb.244641.
- Rio B, Guy C, Françoise L, Daniel L. Evolution d'une spécialisation saisonnière chez *Drosophila erecta* (Dipt., Drosophilidae). *Ann Soc Entomol Fr*. 1983;19(2):235–248. doi:10.1080/21686351.1983.12278361.
- Roach MJ, Schmidt SA, Borneman AR. Purge haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. *BMC Bioinformatics*. 2018;19(1):460. doi:10.1186/s12859-018-2485-7.
- Saisawang C, Wongsantichon J, Ketterman AJ. A preliminary characterization of the cytosolic glutathione transferase proteome from *Drosophila melanogaster*. *Biochem J*. 2012;442(1):181–190. doi:10.1042/BJ20111747.
- Schoville SD, Chen YH, Andersson MN, Benoit JB, Bhandari A, Bowsher JH, Brevik K, Cappelle K, Chen MJM, Childers AK, et al. A model species for agricultural pest genomics: the genome of the Colorado potato beetle, *Leptinotarsa decemlineata* (Coleoptera: chrysomelidae). *Sci Rep*. 2018;8(1):1931. doi:10.1038/s41598-018-20154-1.
- Sehlmeyer S, Wang L, Langel D, Heckel DG, Mohagheghi H, Petschenka G, Ober D. Flavin-dependent monooxygenases as a detoxification mechanism in insects: new insights from the arctiids (Lepidoptera). *PLoS ONE*. 2010;5(5):e10435. doi:10.1371/journal.pone.0010435.
- Shim J, Lee Y, Jeong YT, Kim Y, Lee MG, Montell C, Moon SJ. The full repertoire of *Drosophila* gustatory receptors for detecting an aversive compound. *Nat Commun*. 2015;6(1):8867. doi:10.1038/ncomms9867.

- Simon J-C, d'Alençon E, Guy E, Jacquin-Joly E, Jaquier J, Nouhaud P, Peccoud J, Sugio A, Streiff R. Genomics of adaptation to host-plants in herbivorous insects. *Brief Funct Genomics*. 2015; 14(6):413–423. doi:10.1093/bfpg/elv015.
- Singh SP, Coronella JA, Beneš H, Cochrane BJ, Zimniak P. Catalytic function of *Drosophila melanogaster* glutathione S-transferase DmGSTS1-1 (GST-2) in conjugation of lipid peroxidation end products. *Eur J Biochem*. 2001;268(10):2912–2923. doi:10.1046/j.1432-1327.2001.02179.x.
- Smadja C, Shi P, Butlin RK, Robertson HM. Large gene family expansions and adaptive evolution for odorant and gustatory receptors in the pea aphid, *Acyrtosiphon pisum*. *Mol Biol Evol*. 2009;26(9): 2073–2086. doi:10.1093/molbev/msp116.
- Smit AFA, Hubley R, Green P. RepeatMasker Open-4.0. Institute for Systems Biology; 2013–2015. <http://www.repeatmasker.org>
- Southwood TRE. The insect/plant relationship—an evolutionary perspective. *Symp Roy Entomol Soc London*. 1972;6:3–30.
- Stamatakis A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*. 2006;22(21):2688–2690. doi:10.1093/bioinformatics/btl446.
- Stensmyr MC, Dweck HKM, Farhan A, Ibba I, Strutz A, Mukunda L, Linz J, et al. A conserved dedicated olfactory circuit for detecting harmful microbes in *Drosophila*. *Cell*. 2012;151(6):1345–1357. doi: 10.1016/j.cell.2012.09.046.
- Storey JD. Q-Value. 2008. <http://genomics.princeton.edu/storeylab/qvalue/>.
- Strong DR, Lawton JH, Southwood SR. *Insects on Plants. Community Patterns and Mechanisms*. Cambridge (MA): Blackwell Scientific Publications; 1984.
- Suvorov A, Kim BY, Wang J, Armstrong EE, Peede D, D'Agostino ERR, Price DK, Waddell PJ, Lang M, Courtier-Orgogozo V, et al. Widespread introgression across a phylogeny of 155 *Drosophila* genomes. *Curr Biol*. 2022;32(1):111–123.e5. doi:10.1016/j.cub.2021.10.052.
- Suzuki HC, Ozaki K, Makino T, Uchiyama H, Yajima S, Kawata M. Evolution of gustatory receptor gene family provides insights into adaptation to diverse host plants in nymphalid butterflies. *Genome Biol Evol*. 2018;10(6):1351–1362. doi:10.1093/gbe/evy093.
- Sved JA. Linkage disequilibrium and homozygosity of chromosome segments in finite populations. *Theor Popul Biol*. 1971;2(2): 125–141. doi:10.1016/0040-5809(71)90011-6.
- Swain T. Secondary compounds as protective agents. *Annu Rev Plant Physiol*. 1977;28(1):479–501. doi:10.1146/annurev.pp.28.060177.002403.
- Thompson JN. Coevolution and alternative hypotheses on insect/plant interactions. *Ecology*. 1988;69(4):893–895. doi:10.2307/1941238.
- Thompson JN. *The Coevolutionary Process*. Chicago (IL): University of Chicago Press; 1994.
- Van Nynatten A, Bloom D, Chang BSW, Lovejoy NR. Out of the blue: adaptive visual pigment evolution accompanies Amazon invasion. *Biol Lett*. 2015;11(7):20150349. doi:10.1098/rsbl.2015.0349.
- Vaser R, Sović I, Nagarajan N, Šikić M. Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res*. 2017; 27(5):737–746. doi:10.1101/gr.214270.116.
- Vertacnik KL, Herrig DK, Godfrey RK, Hill T, Geib SM, Unckless RL, Nelson DR, Linnen CR. Ecological correlates of gene family size in a pine-feeding sawfly genome and across Hymenoptera. *bioRxiv* 435331. <https://doi.org/10.1101/2021.03.14.435331>, 14 March 2021. preprint: not peer reviewed.
- Vertacnik KL, Linnen CR. Evolutionary genetics of host shifts in herbivorous insects: insights from the age of genomics. *Ann N Y Acad Sci*. 2017;1389(1):186–212. doi:10.1111/nyas.13311.
- Vorojeikina D, Broberg K, Love TM, Davidson PW, van Wijngaarden E, Rand MD. Editor's highlight: glutathione S-transferase activity moderates methylmercury toxicity during development in *Drosophila*. *Toxicol Sci*. 2017;157(1):211–221. doi:10.1093/toxsci/kfx033.
- Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One*. 2014;9(11):e112963. doi:10.1371/journal.pone.0112963.
- Wang Y, Coleman-Derr D, Chen G, Gu YQ. Orthovenn: a web server for genome wide comparison and annotation of orthologous clusters across multiple species. *Nucleic Acids Res*. 2015;43(W1): W78–W84. doi:10.1093/nar/gkv487.
- Weadick CJ, Chang BSW. Complex patterns of divergence among green-sensitive (RH2a) African cichlid opsins revealed by Clade model analyses. *BMC Evol Biol*. 2012;12(1):206. doi:10.1186/1471-2148-12-206.
- Weiss LA, Dahanukar A, Kwon JY, Banerjee D, Carlson JR. The molecular and cellular basis of bitter taste in *Drosophila*. *Neuron*. 2011;69(2):258–272. doi:10.1016/j.neuron.2011.01.001.
- Wen Z, Rupasinghe S, Niu G, Berenbaum MR, Schuler MA. CYP6B1 and CYP6B3 of the black swallowtail (*Papilio polyxenes*): adaptive evolution through subfunctionalization. *Mol Bio Evol*. 2006; 23(12):2434–2443. doi:10.1093/molbev/msl118.
- Wheat CW, Vogel H, Wittstock U, Braby MF, Underwood D, Mitchell-Olds T. The genetic basis of a plant–insect coevolutionary key innovation. *Proc Natl Acad Sci USA*. 2007;104(51):20427–20431. doi: 10.1073/pnas.0706229104.
- Whiteman NK, Gloss AD, Sackton TB, Groen SC, Humphrey PT, Lapoint RT, Sønderby IE, Halkier BA, Kocks C, Ausubel FM, et al. Genes involved in the evolution of herbivory by a leaf-mining, *Drosophila* fly. *Genome Biol Evol*. 2012;4(9):900–916. doi:10.1093/gbe/evs063.
- Whiteman NK, Groen SC, Chevasco D, Bear A, Beckwith N, Gregory TR, Denoux C, Mammarella N, Ausubel FM, Pierce NE. Mining the plant–herbivore interface with a leafmining *Drosophila* of *Arabidopsis*. *Mol Ecol*. 2011;20(5):995–1014. doi:10.1111/j.1365-294X.2010.04901.x.
- Whitworth AJ, Theodore DA, Greene JC, Benes H, Wes PD, Pallanck LJ. Increased glutathione S-transferase activity rescues dopaminergic neuron loss in a *Drosophila* model of Parkinson's Disease. *Proc Natl Acad Sci U S A*. 2005;102(22):8024–8029. doi:10.1073/pnas.0501078102.
- Wiens JJ, Lapoint RT, Whiteman NK. Herbivory increases diversification across insect clades. *Nat Commun*. 2015;6:8370. doi:10.1038/ncomms9370.
- Xu L, Dong Z, Fang L, Luo Y, Wei Z, Guo H, Zhang G, Gu YQ, Coleman-Derr D, Xia Q, et al. Orthovenn2: a web server for whole-genome comparison and annotation of orthologous clusters across multiple species. *Nucleic Acids Res*. 2019;47(W1): W52–W58. doi:10.1093/nar/gkz333.
- Xu W, Papanicolaou A, Zhang H-J, Anderson A. Expansion of a bitter taste receptor family in a polyphagous insect herbivore. *Sci Rep*. 2016;6(1):23666. doi:10.1038/srep23666.
- Yang Z. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Mol Biol Evol*. 1998;15(5): 568–573. doi:10.1093/oxfordjournals.molbev.a025957.
- Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol*. 2007;24(8):1586–1591. doi:10.1093/molbev/msm088.
- Yang Z, Nielsen R. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Mol Biol Evol*. 2002;19(6):908–917. doi:10.1093/oxfordjournals.molbev.a004148.

- Yassin A, Debat V, Bastide H, Gidaszewski N, David JR, Pool JE. Recurrent specialization on a toxic fruit in an island *Drosophila* population. *Proc Natl Acad Sci U S A*. 2016;113(17):4771–4776. doi:10.1073/pnas.1522559113.
- You M, Yue Z, He W, Yang X, Yang G, Xie M, Zhan D, Baxter SW, Vasseur L, Gurr GM, et al. A heterozygous moth genome provides insights into herbivory and detoxification. *Nat Genet*. 2013;45(2):220–225. doi:10.1038/ng.2524.
- Zhang J, Nielsen R, Yang Z. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Mol Biol Evol*. 2005;22(12):2472–2479. doi:10.1093/molbev/msi237.
- Zhang C, Rabiee M, Sayyari E, Mirarab S. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics*. 2018;19(Suppl. 6):153. doi:10.1186/s12859-018-2129-y.
- Zhou J-J, Huang W, Zhang G-A, Pickett JA, Field LM. “Plus-C” odorant-binding protein genes in two *Drosophila* species and the malaria mosquito *Anopheles gambiae*. *Gene*. 2004;327(1):117–129. doi:10.1016/j.gene.2003.11.007.
- Zhou J-J, Kan Y, Antoniw J, Pickett JA, Field LM. Genome and EST analyses and expression of a gene family with putative functions in insect chemoreception. *Chem Senses*. 2006;31(5):453–465. doi:10.1093/chemse/bjj050.

Editor: R. Anholt