Title: Evolution of olfactory receptors tuned to mustard oils in herbivorous Drosophilidae

Authors: #Teruyuki Matsunaga[1*], #Carolina E. Reisenman[2], #Benjamin Goldman-Huertas[3], Philipp Brand[4], Kevin Miao[1], Hiromu C. Suzuki[1], Kirsten I. Verster[1], Santiago R. Ramírez[4], Noah K. Whiteman[1*]

#Equal contributions

Affiliations:

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

2 Department of Molecular and Cell Biology, University of California Berkeley, Berkeley, CA

3 Department of Molecular and Cellular Biology, University of Arizona, Tucson, AZ

4 Department of Evolution and Ecology, University of California Davis, Davis, CA

[*]Correspondence to: whiteman@berkeley.edu (N.K.W.) and teru.matsu0208@berkeley.edu

Keywords:

Drosophila melanogaster, *Scaptomyza flava*, herbivory, evolution, olfaction, isothiocyanate, chemoreceptor, SSR, olfactory receptor, wasabi, Brassicaceae, Or67b, gene duplication, neofunctionalization, subfunctionalization, specialization, olfactory specialization

Abstract (150):

The diversity of herbivorous insects is attributed to their propensity to specialize on toxic plants. In an evolutionary twist, toxins betray the identity of their bearers if herbivores co-opt them as cues for finding host plants, but the mechanisms underlying this process are poorly understood. We focused on *Scaptomyza flava*, an herbivorous drosophilid specialized on isothiocyanate (ITC)-producing (Brassicales) plants, and identified three *Or67b* paralogs that were triplicated during the origin of herbivory. Using heterologous systems for the expression of olfactory receptors, we found that *S. flava* Or67bs, but not the homologs from microbe-feeding relatives, responded selectively to ITCs, each paralog detecting different subsets of ITCs. Consistently with this, *S. flava* was attracted to ITCs, as was *D. melanogaster* expressing *S. flava* Or67b3 in attractive olfactory circuits. Thus, our results show that plant toxins were co-opted as olfactory attractants through gene duplication and functional specialization (neofunctionalization and subfunctionalization) in drosophilids flies.

INTRODUCTION

Many plant molecules used in food, agriculture and medicine are hypothesized to have first evolved as defenses against natural enemies (Frankel, 1959). Among the most familiar are reactive electrophiles found in plants that produce a sensation of pain when eaten, including diallyl disulfide and thiosulfinates in Alliaceae (*e.g.*, onions and garlic), α, β-unsaturated aldehydes in Lauraceae (*e.g.*, cinnamon and olives), and isothiocyanates (ITCs) in Brassicales plants (*e.g.*, arugula, broccoli, papaya, radish, wasabi, watercress). These electrophiles all activate the 'wasabi taste receptor' TrpA1, which is conserved in flies and humans (Kang et al., 2010). Although ITCs are potent natural insecticides that are aversive to most insects, including *D. melanogaster (Lichtenstein et al., 1962)*, some insect species are actually specialized on Brassicales plants. This guild has been important in advancing the field of co-evolution (Berenbaum and Zangerl, 2008) because these insects have evolved ITC diversion or resistance mechanisms (Gloss et al., 2014; Winde and Wittstock, 2011). Brassicales specialists from many insect orders (*e.g.*, Diptera; Eckenrode and Arn, 1972, Heteroptera; Pivnick et al., 1991, Hemiptera; Demirel and Cranshaw, 2006, and Lepidoptera; Pivnick et al., 1994) can be trapped with ITC-baits in crop fields, revealing an evolutionary twist of fate in which animals are able to use ancestrally aversive electrophiles as olfactory cues for host-plant finding (Kergunteuil et al., 2012). However, the evolutionary mechanisms underlying the chemosensory adaptations to these toxic plants are generally unknown in herbivorous insects (Liu et al., 2020). Identifying these mechanisms can help us understand the evolution of extant herbivorous insect species, 90% of which are specialized on a limited set of toxic host plants (Bernays and Graham 1988).

A compelling lineage to investigate how specialist herbivorous insects evolve to co-opt plant defenses as host finding cues is provided by the drosophilid fly *Scaptomyza flava*. The *Scaptomyza* lineage is nested phylogenetically in the subgenus *Drosophila*, sister to the Hawaiian *Drosophila*

radiation (Lapoint et al., 2013). The subgenus *Scaptomyza* contains many herbivorous species -the adult females make feeding punctures in living leaves using their sclerotized ovipositors (Pelaez et al., 2020)-, and their larvae mine the living leaves of these plants (Whiteman et al., 2011), a life history highly unusual within the Drosophilidae. More remarkably, *S. flava* specializes on Brassicales plants (Mitchell-Olds, 2001) and, like humans and *D. melanogaster*, uses the mercapturic pathway to detoxify ITCs (Gloss et al., 2014). *S. flava* was first reported from *Arabidopsis thaliana* in North America as *S. flaveola (Chittenden, 1902)*. Thus, the powerful genomic and genetic tools of both the *Arabidopsis* and the *Drosophila* lineages can be utilized. Herbivory evolved ca. 10-15 million years ago in the *Scaptomyza* lineage (Figure 1A) and so it provides an unusually useful context to understand the evolutionary and functional mechanisms underlying chemosensory specialization on mustard plants.

We previously found that *S. flava* lost three olfactory receptor (*Or*) genes encoding canonical yeast-associated receptors after divergence from microbe-feeding drosophilid relatives, contemporaneous with the evolution of herbivory (Goldman-Huertas et al., 2015). The loss of these receptors helps to explain why *S. flava* is not attracted to the ancestral feeding substrate (microbes living on decaying plant tissues), as these genes are required for attraction to aliphatic esters in *D. melanogaster* (Semmelhack and Wang, 2009). However, it does not explain how some herbivorous *Scaptomyza* species are able to use toxic Brassicales plants as hosts. Thus, the mechanistic basis of the *S. flava*'s putative attraction to Brassicales plants (a gain of function phenotype), as well as the evolutionary path leading to this attraction, are unknown.

Gain of function through gene duplication and subsequent divergence plays an important role in evolutionary innovation and is frequently associated with trophic transitions (Ohno, 1970). Gene and whole genome duplications in Brassicales in the last 90 million years resulted in the evolution of glucosinolates, the precursors of ITC compounds (Edger et al., 2015). Reciprocally, in diverse

Brassicales-specialist pierid butterflies, nitrile specifier proteins that divert hydrolysis of aliphatic glucosinolates away from ITC production evolved in tandem, underpinning their species diversification (Wheat et al., 2007). Similarly, we seek to investigate whether duplication of chemosensory genes in the *S. flava* lineage contributed to attraction to specific host-plant volatiles, including ITCs.

Several scenarios have been proposed to explain how gene duplication contributes to the acquisition of novel gene functions (Ohno, 1970; Heidel-Fischer et al., 2019): (A) neofunctionalization, which occurs when one of the two duplicated genes (paralogs) acquires a new function after accumulating *de novo* mutations, while the other copy retains its ancestral function; (B) subfunctionalization *sensu stricto*, where mutations accumulate in both copies leading to partitioning of the ancestral function; and (C) specialization, which happens when subfunctionalization and neofunctionalization evolve simultaneously, yielding gene copies that are functionally different from one another and from the ancestral copy (He and Zhang, 2005; Assis and Bachtrog, 2013).

We took a genes-first, "reverse chemical ecology approach" (*e.g.* Choo et al., 2018) to address the evolutionary mechanisms that lead to olfactory host-plant specialization in *S. flava*. Three chemoreceptor protein families (olfactory receptors -Ors-, ionotropic receptors -Irs-, and nociceptive receptors –Trp-) are candidates for mediating responses to host-specific volatile electrophiles such as ITCs (Joseph and Carlson, 2015). In particular, insect Ors are collectively sensitive to a variety of odorants important for food and host finding, avoidance of predators, and animal communication, including aversive chemicals (*e.g.* Or56a and its ligand geosmin; Stensmyr et al., 2012), pheromones (*e.g.* Or67d cVA; Kurtovic et al., 2007, BmOR-1 bombykol; Butenandt, 1959; Sakurai et al., 2004), and host-, oviposition- and food-related attractive compounds (Or19a Limonene; Dweck et al., 2013, and Or22a host-substrate-specific esters; Dekker et al., 2006a; Auer et al., 2020; Mansourian et al., 2018; Linz et al., 2013). Accordingly, we scanned the genome sequence of *S. flava* to identify rapidly evolving

Ors and found a lineage-specific gene copy number expansion of the olfactory receptor gene *Or67b* that we named *Or67b1*, *Or67b2* and *Or67b3*. These three paralogs exhibit signatures of positive selection and duplicated likely within the last 15 million years, at the base of the mustard-feeding clade. In contrast, *Or67b* is present as a single copy under strong evolutionary constraint across various species of Drosophilidae (Guo and Kim, 2007; Robertson et al., 2003; Goldman-Huertas et al., 2015), including two focal species included here, the microbe-feeding close relative *S. pallida* and the more distant-relative *D. melanogaster*. Our *in vivo* functional characterization of all full-length Or67b proteins from these three species shows that the three *S. flava* Or67b proteins, but not the conserved Or67b proteins from its microbe-feeding relatives, responded selectively and sensitively to volatile ITCs. Concomitantly, we found a population of *S. flava* antennal olfactory sensory neurons (OSNs) sensitive to ITCs. In agreement with these results, *S. flava*, but not *S. pallida* or *D. melanogaster*, is attracted to odors from Brassicales plants including single ITC compounds. Moreover, expression of *Sfla* Or67b3 in two different attractive olfactory circuits in transgenic *D. melanogaster* indicates that this Or is sufficient to mediate behavioral attraction to ITCs. Finally, our genetic silencing experiments indicates that the Or67b olfactory circuit mediates attraction in *D. melanogaster*, and possibly in other drosophilid flies, including *S flava*. Of particular significance is that *Sfla* Or67b proteins selectively shifted their odor-receptive ranges from an ancestral olfactome of ketones, alcohols, and aldehydes (Kreher et al., 2005; Kreher et al., 2008; Galizia and Sachse, 2009; Montague et al., 2011) to ITCs, an entirely different chemical class of compounds. In summary, our results suggest a relatively simple sensory mechanism by which mustard-specialist herbivorous insects evolved olfactory attraction towards host-plant ITCs via gene duplication followed by neofunctionalization of a specific clade of Ors.

RESULTS

Phylogenetic analysis identified *Or67b* paralogs as candidates mediating the detection of ecologically relevant host-plant odorants

To search for duplicated chemosensory receptors in *S. flava*, we conducted a phylogenetic analysis of the Or protein sequences of *S. flava* and four other *Drosophila* species to identify *Scaptomyza*-specific losses and gains (Figure 1-figure supplement1). The Or topology was largely congruent with previous *Drosophila* Or trees (McBride et al., 2007), except for some deeper nodes which had low bootstrap support. In order to identify *Or* paralogs potentially evolving under positive selection in *S. flava*, we inferred the ratio of nonsynonymous to synonymous substitution rates (d_N/d_S), using a foreground/background branch test as implemented in PAML (Yang, 2007). Out of seventy-five *S. flava* branches tested, seven branch models, corresponding to paralogs of *Or63a*, *Or67b* and *Or98a* in *D. melanogaster*, indicated a foreground rate larger than one, consistent with episodic positive selection (supplementary file 1). Among these receptors, *Or63a* is only expressed in *D. melanogaster* larvae (Hoare et al., 2011), and the *Or98a*-like genes found in *Scaptomyza* have no *D. melanogaster* homologs and are not characterized functionally. In contrast, the *Drosophila* homolog of *Or67b* is expressed in adults, modulates oviposition behavior in *D. melanogaster* (Chin et al., 2018), and is implicated in selection of oviposition sites in *D. mojavensis (Crowley-Gall et al., 2016)*, making it a good candidate for the olfactory adaptation of *Scaptomyza* to a novel host niche. After expanding the representation of *Or67b* orthologs in a phylogenetic analysis and conducting branch tests in PAML on all branches, we found elevated d_N/d_S exclusively in the *S. flava* and *D. mojavensis Or67b* lineages (Figure 1B and supplementary file 1). Furthermore, synteny analysis confirmed that the regions upstream and downstream of the two non-syntenic copies are present in both *S. pallida* and *D. melanogaster*,

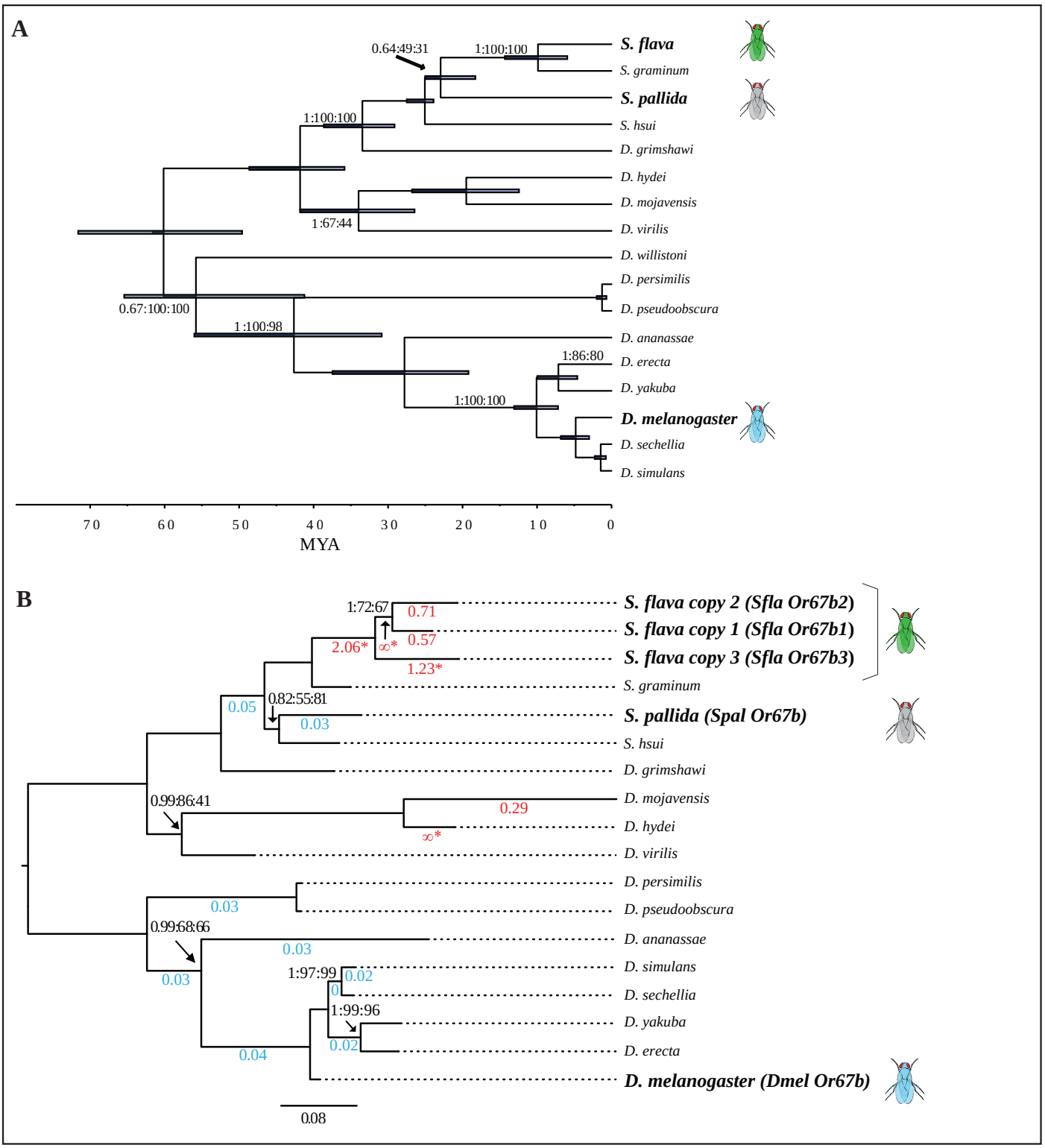

Figure 1 Maximum likelihood (ML) phylogeny of *Or67b* in Drosophilidae.
(A) Time-calibrated Bayesian chronogram of *Scaptomyza* and *Drosophila* spp. inferred from nine protein coding and two ribosomal genes. Source species for the *Or67b* coding sequences crossed into the empty neuron system are indicated with fly pictograms. Bars indicate 95% highest posterior density (HPD) age estimates in millions of years ago. Tree topology labeled as follows: Posterior Probability (PP):ML BS:Parsimony BS on branches. Scale bar proportional to MYA. (**B**) ML phylogeny reconstructed based on the coding sequence of *Or67b* orthologs from twelve *Drosophila* species, *S. pallida*, *S. hsui*, *S. graminum*, and *S. flava*. All bootstrap supports for the nodes are >80% and all posterior probabilities were >0.95 for the MrBayes tree. Branches with significant support (FDR p-value < 0.05) for d_N/d_S values different from the background rate are indicated with colored branch labels (blue where the foreground rate is less than the background, and red/enlarged fonts where d_N/d_S is greater than the background). Only *S. flava* and *D. mojavensis* branches have significantly elevated d_N/d_S according to branch model tests. *S. flava*, *S. pallida* and *D. melanogaster* labeled identically to (A). Scale bar units are substitutions per site.

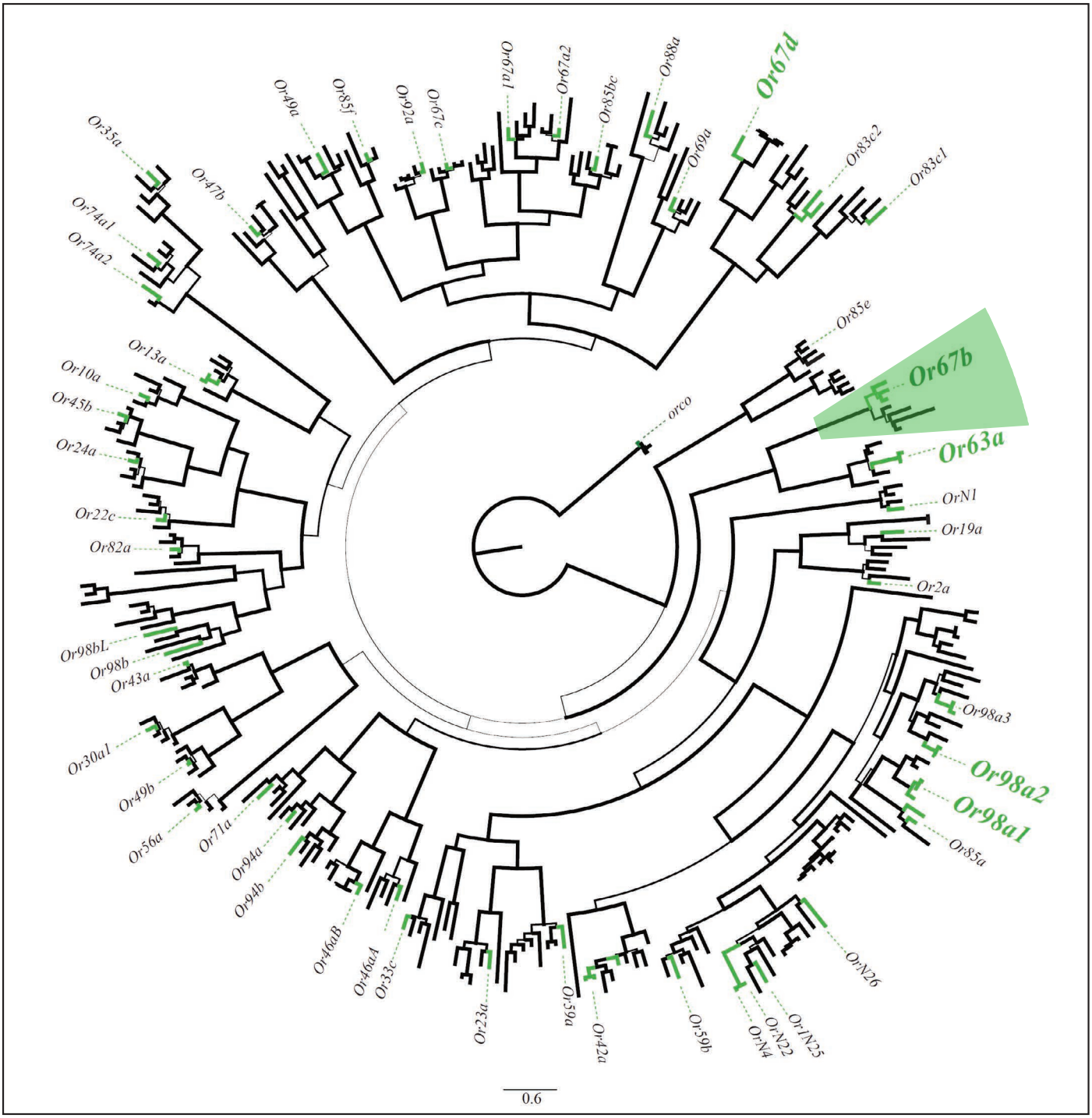

Figure supplement 1 Maximum likelihood (ML) phylogeny of *Ors* in Drosophilidae.

ML phylogeny reconstructed from protein translations of the *Ors* found in *S. flava*, *D. melanogaster*, *D. grimshawi*, *D. virilis* and *D. mojavensis* genomes. Line width of branches are proportional to bootstrap support. Green branches indicate *Scaptomyza Ors*. Enlarged gene names in bold include branches with estimated d_N/d_S >1. Scale bar units are substitutions per site.

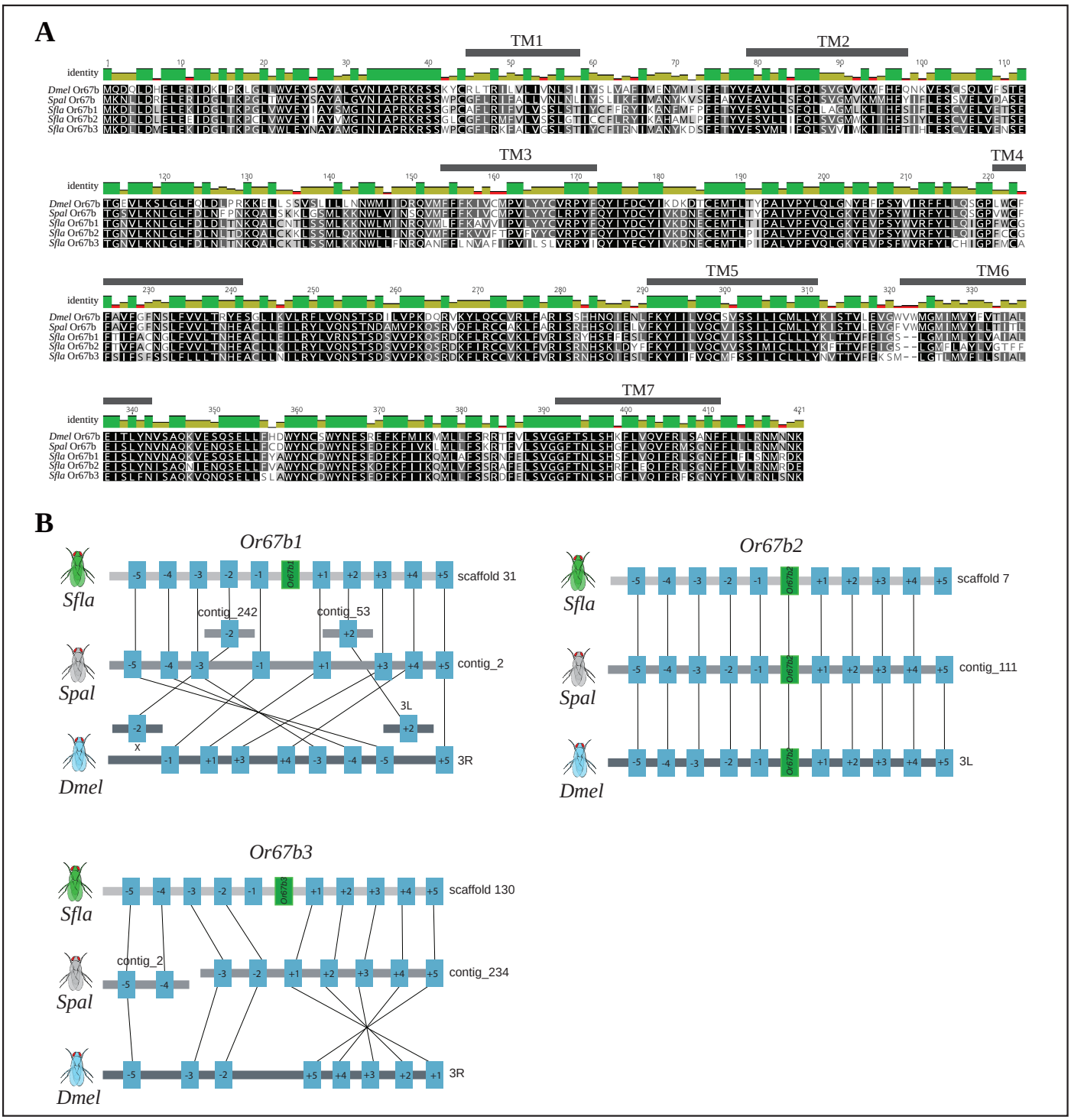

Figure supplement 2 Or67b protein alignment and micro-syntenic patterns of scaffolds from *S. flava, S. pallida,* and *D. melanogaster.*

(A) Alignments of Or67b proteins. Note that orthologs and paralogs share a large number of amino acids (black squares). Darker colors illustrate higher degrees of sequence similarity, and lighter colors denote residues with high variability in sequence across paralogs and orthologs. Also shown are each of the seven predicted transmembrane domains (TM1-7). **(B)** Micro-syntenic patterns of Or67b scaffolds. Five genes up and downstream of each *S. flava Or67b* ortholog are shown. We determined that there was only one copy of *Or67b* in both *S. pallida* and *D. melanogaster*. Note that in both *S. pallida* and *D. melanogaster*, many of the pGOIs are still present and syntenic with those in *S. flava*, suggesting that the absence of the other *Or67b* orthologs is not a consequence of mis-assembly but a *bona fide* absence.

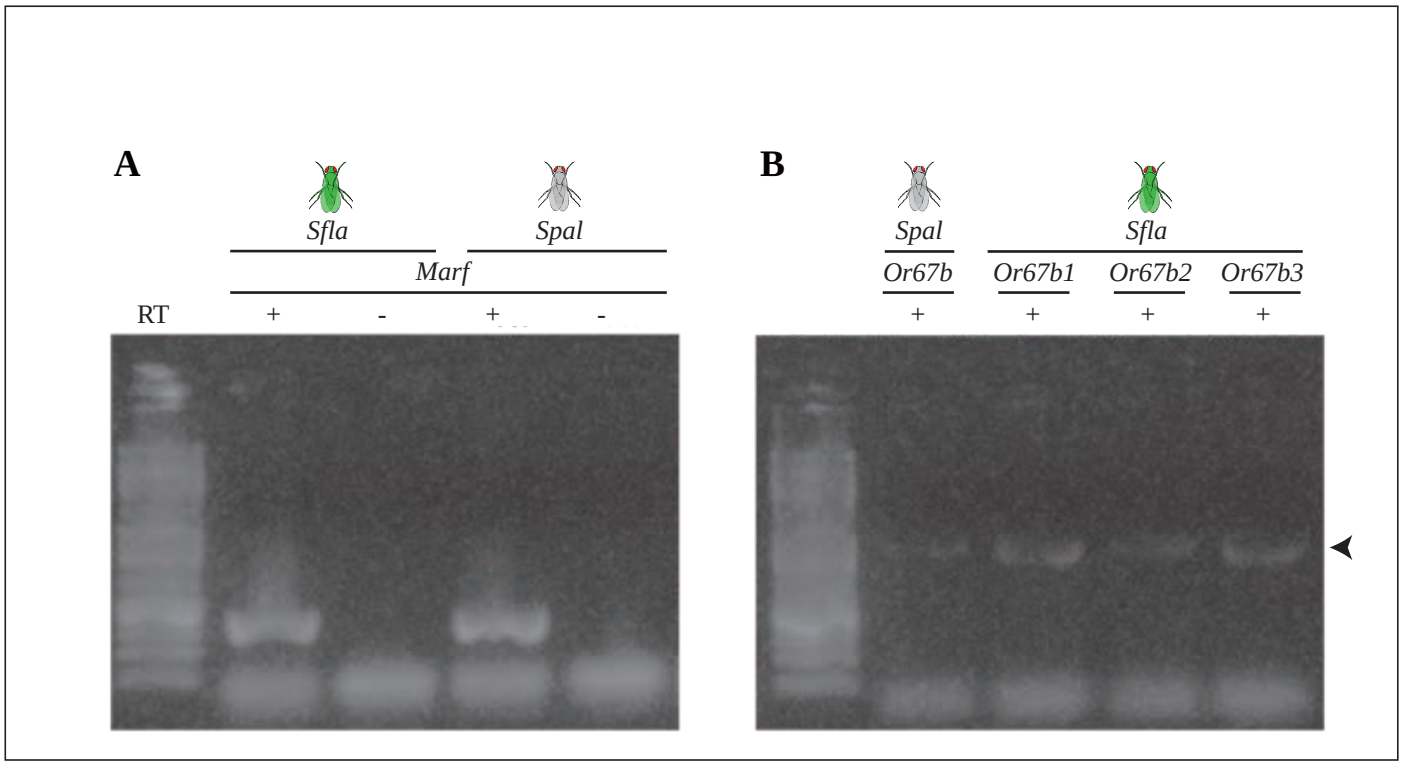

Figure supplement 3 *Or67bs* are expressed in *Scaptomyza* spp.

(A) Amplification of *Marf* from cDNA generated from whole body extracts of *S. flava* and *S. pallida* adults (+RT). As expected, *Marf* does not amplify in templates treated with DNaseI without reverse transcriptase (– RT), used as a negative control. **(B)** Amplification of *Or67b* genes from *S. pallida* and *S. flava* whole adult cDNA, which reveal *in vivo* transcription of *Or67b* genes in adult *Scaptomyza* (arrow). Ladder (firsts column) is GeneRuler™ 1 kb plus (ThermoFisher, USA).

suggesting that the gene's absence is not a consequence of a missing genomic region but an actual absence (Figure 1-figure supplement 2A, 2B).

S. flava Or67b proteins respond specifically to mustard plant volatiles

Because *Or67b* is triplicated exclusively in the *S. flava* lineage, among drosphilids with sequenced genomes, we next investigated whether *Sfla* Or67b proteins acquired novel sensitivity towards *S. flava*-specific niche odorants, consistent with positive selection during the evolution of mustard herbivory (Figure 1). First, we confirmed that all three *S. flava* paralogs and the *S. pallida* ortholog are expressed in adults (Figure 1-figure supplement 3). To study the odor-response profile of Or67b across species, we used a *D. melanogaster* mutant that either lacks its endogenous olfactory receptor Or22a in ab3A (antennal basiconic sensilla 3A) olfactory sensory neurons (OSNs), or its endogenous Or67d in at1 (antennal trichoid sensilla 1) OSNs (both known as “empty neuron” systems; Hallem and Carlson, 2006; Gonzalez et al., 2016; Figure 2A). We conducted single sensillum recordings (SSRs) in transgenic *D. melanogaster* flies expressing each of the *Sfla Or67b* paralogs, as well as the *Or67b* orthologs from *S. pallida* (*Spal*) and *D. melanogaster* (*Dmel*) in ab3A OSNs (or in at1 OSNs, see below; Figure 2). Endogenous deletions of *Or22a* and *Or67d* in ab3A and at1 OSNs were confirmed by the respective lack of responses to their native ligands, ethyl hexanoate and 11 *cis*-vaccenyl acetate (Figure 2-figure supplement 4). Thus, we used the “ab3A empty neuron” system for studying the olfactory responses of *Sfla* Or67b1, *Sfla* Or67b3, *Spal* Or67b and *Dmel* Or67b, and the “at1 empty neuron” system for studying the responses of *Sfla* Or67b2.

In order to test which odors specifically activate *Sfla* Or67b paralogs, we tested the responses of all Or67bs to stimulation with apple cider vinegar (a *D. melanogaster* attractant; Faucher et al., 2013), arugula leaves (*Eruca vesicaria*, a mustard plant), and tomato leaves (*Solanum lycopersicum*, a non-

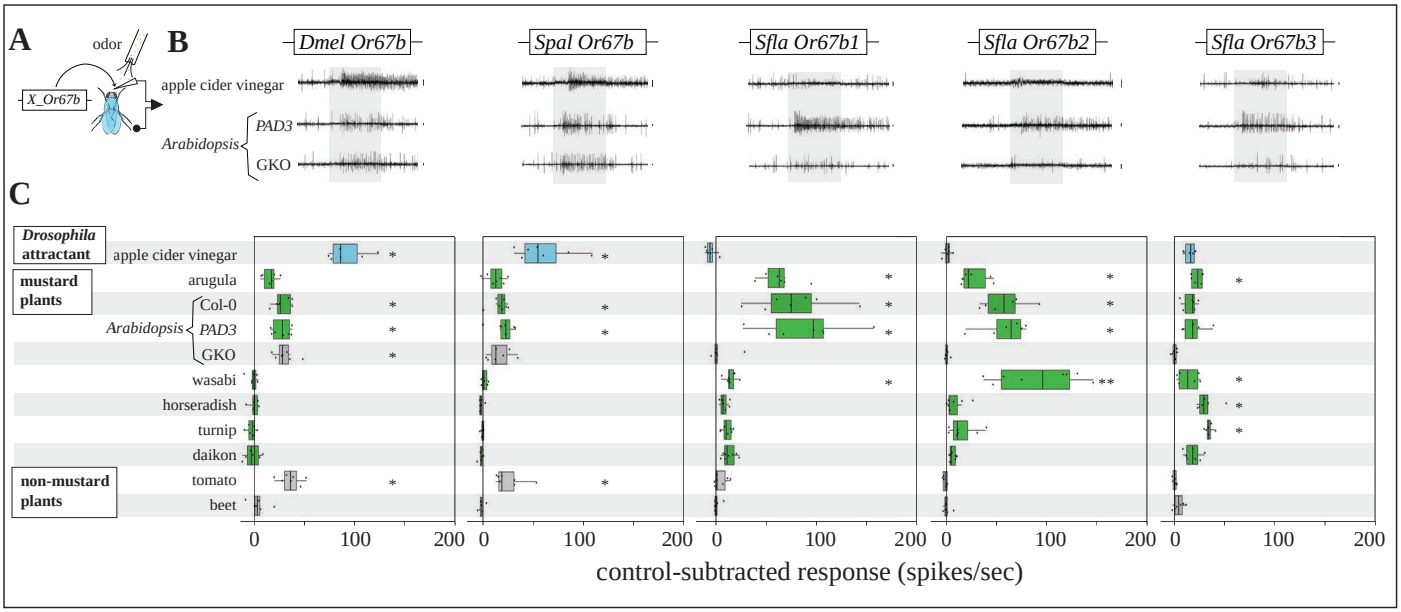

Figure 2 Responses of homologs Or67bs from *D. melanogaster*, *S. pallida* and *S. flava* expressed in the *D. melanogaster* empty neuron systems to stimulation with natural odor blends.
(A) Schematic representation of the single sensillum recording using two "empty neuron systems". Or67b proteins (*X_Or67b*, where *X* refers to the fly species) were expressed in a *D. melanogaster* mutant that lacks its endogenous *Or22a* in ab3A OSNs (Hallem and Carlson, 2006) or in a mutant that lacks *Or67d* in at1 OSNs (Kurtovic et al., 2007). **(B)** Representative electrophysiological recordings obtained from the targeted sensilla of flies expressing *Or67b* genes in OSNs in response to stimulation with apple cider vinegar, and *PAD3* and quadruple glucosinolate knockout (GKO) *Arabidopsis thaliana* mutant lines. The transparent rectangles indicate the onset and duration (1 sec) of the stimulation. Calibration bars: 10 mV. **(C)** Responses (net number of spikes/second, control-subtracted, n=6-9 obtained from 2-4 animals) evoked by stimulation with apple cider vinegar, mustard leaf volatiles (arugula and *A. thaliana*), mustard root volatiles (wasabi, horseradish, turnip and daikon), non-mustard leaf volatiles (tomato), and non-mustard root volatiles (beet, control). The outer edges of the horizontal bars represent the 25% and 75% quartiles, the vertical line inside the bars represents the median and the whiskers represent 10% and 90% quartiles; each dot represents an individual response. Asterisks indicate significant differences between the control-subtracted net number of spikes and a threshold median value (10 spikes/second), as explained in material and methods (one-sample signed rank tests; * $p<0.05$, ** $p<0.01$). Neurons expressing *Dmel Or67b* and *Spal Or67b*, but not those expressing any of the *S. flava Or67b* paralogs, significantly responded to stimulation with apple cider vinegar. Conversely, only neurons expressing *S. flava Or67b* paralogs responded to arugula volatiles (which bears ITCs). *Dmel* Or67b responded to all *A. thaliana* genotypes, while *Sfla* Or67b1-2 responded only to ITC-bearing *A. thaliana*, indicating that the presence of ITCs within plants is necessary to evoke responses from these two *S. flava* paralogs. Stimulation with wasabi root volatiles evoked significant responses from all *Sfla* Or67b paralogs but not from the *Dmel* or the *Spal* paralogs.

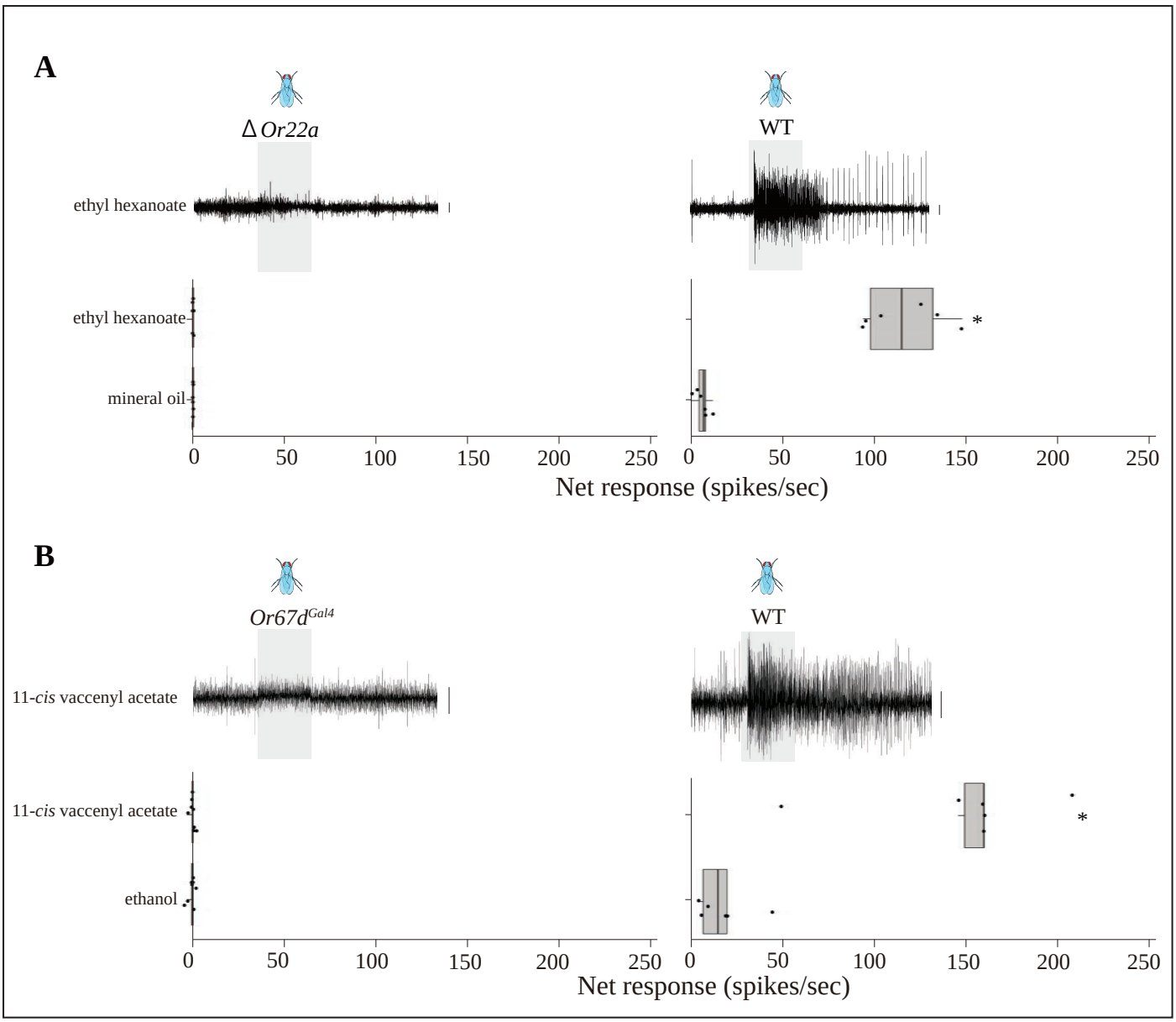

Figure supplement 4 Or22a and Or67b are not expressed in the ab3A and the at1 empty neuron systems.

(A) The lack of *Or22a* expression in the ab3A neuron in the M2-MD fly line ("empty ab3A neuron" mutant) was verified by the absence of electrophysiological responses to ethyl hexanoate 1:100 vol/vol (left). As expected, in wild-type flies (Canton-S), ab3A neurons respond strongly to simulation with this odor compound (right). The bottom part of the panel shows the population responses of mutant (left) and Canton-S flies (right). Data represents the net number of spikes in response to stimulation (n= 6-7, obtained from 3 females, dots: individual data points; boxes represent the 25 and 75% quartiles, whiskers represent the 10 and 90% quartiles, and the vertical line inside the boxes indicate the median). Differences between the responses of flies stimulated with the odorant and the control solvent were statistically different in the case of Canton-S flies (* $p<0.05$, Wilcoxon-matched pairs test) but not in the case of flies lacking Or22a ($p>0.05$). **(B)** Similarly, lack of Or67d expression in the at1 neuron in the *Or67d*^{GAL4} fly line ("empty at1 neuron" mutant) was verified by the lack of responses to 11-*cis* vaccenyl acetate 1:100 vol/vol (left; ethanol was used as the solvent control). The top shows representative electrophysiological recordings and the bottom the population response (data obtained and represented as explained in A). In Canton-S flies, at1 neurons respond strongly to stimulation with this compound (right, * $p<0.05$).

mustard plant that releases large quantities of volatile organic compounds). We found that OSNs expressing *Dmel Or67b* and *Spal Or67b*, but not those expressing any of the *Sfla Or67* paralogs, responded strongly to stimulation with apple cider vinegar (Figure 2B), indicating that the *Sfla* Or67b paralogs lost the ancestral odor selectivity. Volatiles from tomato activated both *Dmel* Or67b and *Spal* Or67b. In contrast, all *Sfla* Or67b paralogs responded to arugula, but not to tomato (Figure 2C).

Brassicales plants produce glucosinolates, some of which are hydrolyzed into ITCs upon tissue wounding (Fahey et al., 2001). These compounds are highly diverse (Mithen et al., 1995), probably as a consequence of differential selection by herbivore communities (Newton et al., 2009). In order to test if the responses of *Sfla* Or67bs to mustard plant volatiles are mediated by ITCs, we used an *A. thaliana* quadruple glucosinolate knockout (GKO) mutant that has nearly abolished expression of aliphatic and indolic glucosinolates and indole-derived camalexin (Beekwilder et al., 2008), (Sønderby et al., 2007), (Zhao et al., 2002). We found that stimulation with both wild-type *A. thaliana* (Col-0) and camalexin-deficient *phytoalexin deficient 3* (*PAD3*) mutant plants (Schuhegger et al., 2006), both of which produce glucosinolates, activated OSNs expressing *Sfla* Or67b1 or b2 but strikingly, stimulation with GKO mutant plants did not (Figure 2C). Responses to the three different *A. thaliana* genotypes differed for all *Sfla* Or67b paralogs (Kruskal-Wallis ANOVA, $p<0.005$ in all three cases). Responses to Col-0 and *PAD3* were not different from each other (Kruskal-Wallis ANOVAs followed by post-hoc Tukey tests; $p>0.8$ for all paralogs), but were different from the responses to the GKO genotype (post-hoc Tukey tests; $p<0.05$ in all cases). In contrast, stimulation with volatiles from any of the three *A. thaliana* genotypes evoked moderate responses from OSNs expressing *Dmel* Or67b that were not different from each other (Figure 2C; Kruskal-Wallis ANOVA, $p>0.8$). *Spal* Or67b responded to Col-0 and *PAD3* genotypes (one-sample signed rank tests; $p<0.05$ in both cases) but not to GKO ($p>0.05$; Figure 2C). However, the responses to stimulation with volatiles from the three different plant genotypes was similar

(median values = 19, 22 and 18 spikes/second; Mann-Whitney ANOVA; $p=0.65$). These results thus demonstrate that the *Sfla* Or67b paralogs are selectively activated by ITCs, the signature mustard plants compounds. In contrast, *Dmel* Or67b was activated by non-ITC vegetative mustard and non-mustard plant volatiles, including herbivore-induced green leaf volatiles (GLVs), consistent with previous studies (Rasmann et al., 2005; Degen et al., 2004).

To test plant-tissue specificity, we used as stimuli aqueous root homogenates of four mustard plant species, including wasabi (*Eutrema japonicum*), horseradish (*Armoracia rusticana*), turnip (*Brassica rapa*), daikon (*Raphanus sativus*), and beet (beet, *Beta vulgaris*), a non-mustard species as a control. We found that OSNs expressing *Sfla* Or67b1-2 (but not OSNs expressing *Dmel* Or67b or *Spal* Or67b) responded selectively to stimulation with wasabi root volatiles (Figure 2C). *Sfla Or67b3* responded also to horseradish and turnip (one-sample signed rank tests, $p<0.05$ in both cases; $p=0.063$ for daikon).

Additionally, we found that OSNs expressing *Sfla* Or67b paralogs differ in their responsiveness to mustard plant volatiles. Stimulation with arugula leaves, for instance, elicited stronger responses in neurons expressing *Sfla* Or67b1 (64 spikes/sec; median) than in neurons expressing *Sfla* Or67b2 (23 spikes/sec) or *Sfla* Or67b3 (26 spikes/sec; Kruskal-Wallis ANOVA, $p<0.006$; post-hoc Tukey tests, $p<0.05$ in both cases; Figure 2C). In contrast, stimulation with wasabi root volatiles produced stronger responses in neurons expressing *Sfla* Or67b2 (96 spikes/sec) than in *Sfla* Or67b1 (14 spikes/sec) or *Sfla* Or67b3 (13 spikes/sec) (Kruskal-Wallis ANOVA, $p<0.001$; post-hoc Tukey tests, $p<0.005$ in both cases). Stimulation with turnip root volatiles produced the strongest responses in neurons expressing *Sfla* Or67b3 (34 spikes/sec), while daikon root volatiles evoked the strongest responses from *Sfla* Or67b1 and b3 (respectively 12 and 19 spikes/second; Figure 2C).

S. flava Or67b paralogs respond selectively to ITCs

We next investigated the odor-tuning profiles of Or67b proteins across species by testing the responses of these proteins to a panel of 42 odorants (all tested at 1:100 vol/vol), which included ITCs, nitriles, GLVs, and *Dmel* Or67b activators with different functional groups (Münch and Galizia, 2016; Figure 3A).

D. melanogaster activators and GLVs evoked responses from *Dmel* Or67b (13/17 odorants; one-sample signed rank tests, $p<0.05$ for 13 odorants) and less so from *Spal* Or67b (5/17 odorants; $p<0.05$ for 5 odorants). In contrast, only one compound (*trans*-2-hexen-1-ol) evoked responses from *Sfla* Or67b1 (median=22 spikes/second; $p<0.05$), and none evoked responses from *Sfla* Or67b2 and *Sfla* Or67b3 (Figure 3A; $p>0.05$ in all cases). While various ITCs evoked responses from each of the *S. flava* paralogs (7/10 ITCs in all three cases; $p<0.05$), none evoked responses from *Dmel* Or67b or *Spal* Or67b ($p>0.05$ in all cases; Figure 3A, B). Tuning curves (Figure 3B) further demonstrated that non-ITC compounds evoke the strongest responses from both *Dmel* Or67b and *Spal* Or67b (acetophenone and *cis*-3-hexenyl-butyrate, respectively, center of the distribution, yellow bars), while all the three *Sfla* Or67b paralogs display the strongest responses to ITC compounds (center of the distribution, green bars). Also, *Sfla* Or67b1 has the narrower odorant-receptive range (*i.e.* responds to a narrow set of odorant -ITC- compounds), indicated by the highest kurtosis and sharp peak of the tuning curve. Overall, these results demonstrate that *Dmel* Or67b and *Spal* Or67b have similar odor-response profiles (including lack of responses to ITCs), and that the *Sfla* Or67b paralogs are all selectively responsive to ITCs.

Figure 3 Responses of homologs Or67bs from *D. melanogaster*, *S. pallida* and *S. flava* expressed in the *D. melanogaster* empty neuron systems to stimulation with single odorants.
Experiments were conducted and analyzed as in Figure 2. **(A)** Responses evoked by stimulation with single odorants (tested at 1:100 vol/vol) categorized as follows: *Dmel* Or67b activators (Database of Odor Responses; Münch and Galizia, 2016; blue), green leaf volatiles (GLVs; gray), ITCs (green), benzyl thiocyanate (yellow), nitrile (pink), and TrpA1 activators (purple). ITCs strongly activate OSNs expressing *Sfla Or67b* paralogs and for the most part these neurons do not respond to any of the other odorants used. *Spal* Or67b has an odorant response profile similar to that of *Dmel* Or67b, responding mostly to stimulation with *D. melanogaster* activators and GLVs. Asterisks indicate significant differences as explained in Figure 2. **(B)** Tuning curves for each Or67b, showing the distribution of median responses to the 42 odorants tested (color-coded as in **A)**. The odorants are displayed along the horizontal axis according to the net responses they elicit from each Or. The odorants (numbers) eliciting the strongest responses for each Or are located at the center of the distribution and weaker activators are distributed along the edges; the order of the odorants is thus different for the five Ors (numbers refer to the chemical structures shown in bottom). Note that the strongest responses (center of the distribution) from *Dmel* Or67b and *Spal* Or67b are evoked by *D. melanogaster* activators and GLVs (blue and gray bars), while the strongest responses from all *Sfla* Or67b paralogs are evoked by ITCs (green bars). The tuning breadth of each Or is quantified by the kurtosis value (k) of the distribution (Willmore and Tolhurst, 2001), with higher values indicating narrower odor-response profiles. The chemical structure of the top seven *Sfla* Or67b3 activators, as well as AITC, citronellal, acetophenone, *cis*-3-hexenyl butyrate and benzyl thiocyanate (BTC) are shown at the bottom.

S. flava paralogs differ in their ITC selectivity

We then asked whether *Sfla* Or67b paralogs differed in their ITC selectivity, possibly allowing flies to differentiate between mustard plant species (different species vary in their ITC composition, see below). In order to test this, we stimulated OSNs expressing *Sfla* paralogs with various concentrations of eight selected ITCs that evoked strong responses (*i.e.* having median responses >100 spikes/second; Figure 3B). Interestingly, these ITCs are released from diverse mustard plant species, for instance butyl ITC (BITC), sec-butyl ITC (SBITC), isobutyl ITC (IBITC), and 3-methyl-thio propyl ITC (3MPITC: also known as iberverin) are produced by *A. thaliana* (Rohloff and Bones, 2005). Four-methyl-thio butyl ITC (4MBITC: also known as erucin) is produced by arugula (Jirovetz et al., 2002), benzyl ITC is produced by turnip (Xue et al., 2020), allyl ITC (AITC) is produced by wasabi (Sultana et al., 2003), and phenethyl ITC (PITC) is produced by watercress (Gupta et al., 2014) and horseradish (Kim et al., 2015).

In general, and as expected, odorant responses increased with increasing odorant concentration (Figure 4A, B). All *Sfla* Or67b paralogs had similar dose-responses when stimulated with BITC, while other ITCs elicited more or less Or-specific dose-responses. For instance, *Sfla* Or67b1 had a low concentration sensitivity threshold for both BITC and 4MPITC (under 10^{-4} vol/vol; one-sample signed rank tests; $p<0.05$ in both cases), and *Sfla* Or67b2 was the only paralog that significantly responded to stimulation with AITC. *Sfla* Or67b3 was the most broadly tuned of all paralogs, having low- to -medium concentration sensitivity thresholds (<1:1,000 vol/vol) for BITC, SBITC, PITC and 4MBITC, and responded to all ITCs (except AITC) at 1:100 vol/vol (Figure 4A, B). Interestingly, stimulation with PITC, which is emitted in large quantities by watercress (Gupta et al., 2014), a common *S. flava* host plant, evoked responses only from *Sfla* Or67b3 at all concentrations. We then analyzed the responses of the three *Sfla* Or67b paralogs using principal component analysis, and found that most ITCs distributed separately in the odor space when tested at 1:100 vol/vol, except for BITC, 4MBITC and 3MPITC,

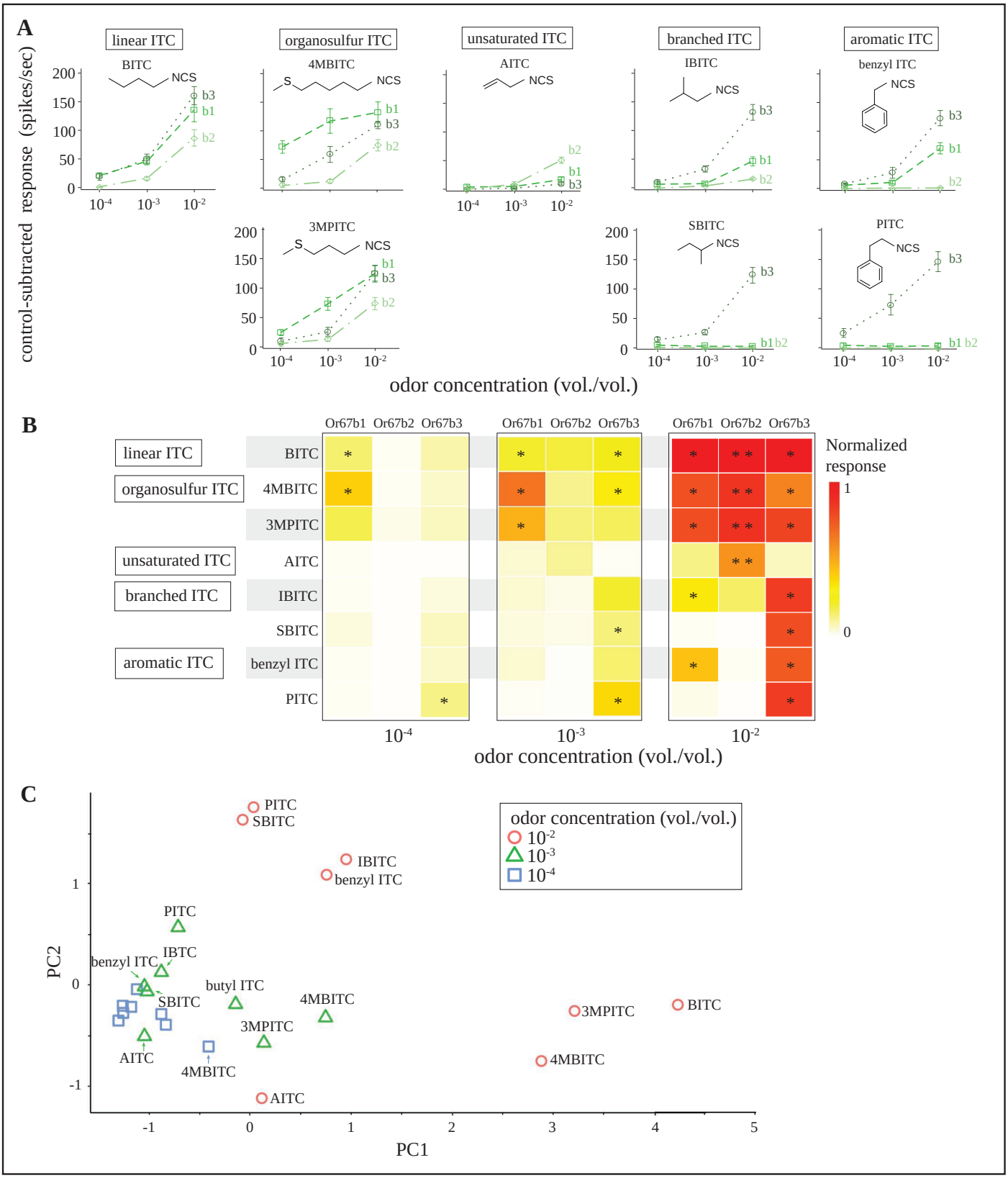

Figure 4 *Sfla* Or67b1-3 have distinct ITC breadth of tuning.
(A) Control-subtracted dose responses of *Sfla* Or67b1, *Sfla* Or67b2 and *Sfla* Or67b3 (abbreviated to b1, b2 and b3) to stimulation with increasing concentrations (vol/vol) of eight different ITCs (categorized according to molecular structure, top boxes; odorant abbreviations are as in Figure 3). Data represent the control-subtracted net number of spikes (average ± SE; n=6-8, obtained from 2-3 animals). **(B)** Heatmap of dose-responses (median, color-coded) from *Sfla* Or67b1, *Sfla* Or67b2, and *Sfla* Or67b3 normalized by each paralog's median response to 1:100 vol/vol of BITC (the strongest ITC activator across paralogs). Asterisks indicate significant differences as explained in Materials and Methods (One-sample signed rank tests; * $p<0.05$; ** $p<0.01$). The strongest responses were evoked by the highest concentration, with many ITCs evoking significant responses from all paralogs particularly *Sfla* Or67b1 and b3; the number of stimuli that evoked significant responses decreased with decreasing odorant concentration. **(C)** Responses of Or67b paralogs in odor space, generated by principal component analysis of median responses from the three *S. flava* paralogs to eight ITCs tested at 1:100 (red circles), 1:1,000 (green triangles), and 1:10,000 vol/vol (blue rectangles). The axes display the two first principal components (PC1 and PC2) which respectively explain 75% and 17% of the variance.

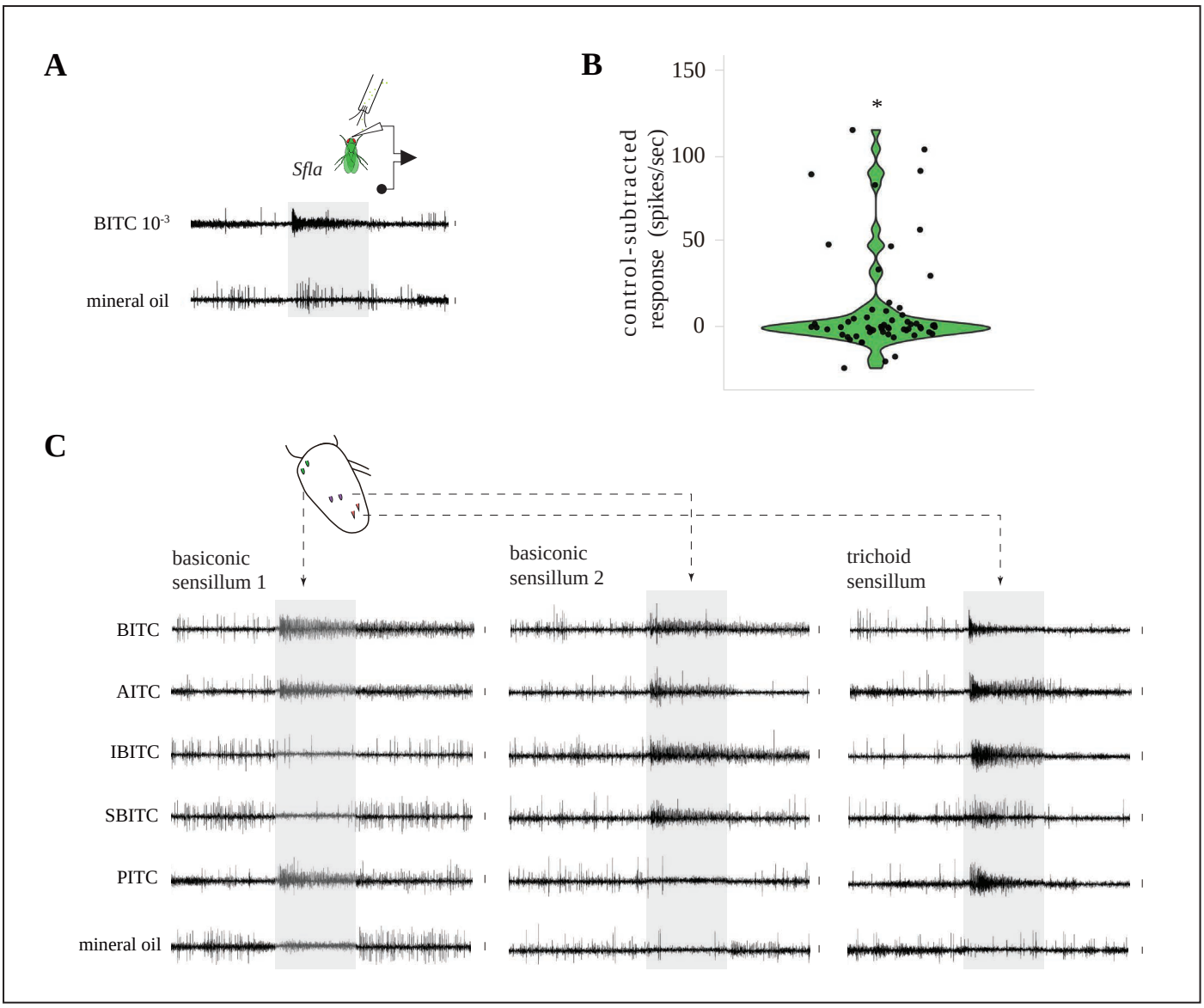

Figure supplement 5 Three different antennal classes of OSNs respond to ITCs in *S. flava*.

(A) Representative electrophysiological recording obtained from an antennal *S. flava* OSN in response to stimulation with BITC 1:1,000 vol/vol and the mineral oil solvent control. The gray bars indicate the onset and duration (1 second) of stimulation. Calibration bars (vertical lines to the right) = 10 mV. **(B)** Responses of *S. flava* individual antennal OSNs (net number of spikes, control-subtracted; n=60; 3 individuals) in response to BITC 1:1,000 vol/vol. Most OSNs (73%) showed little or no response (<10 spikes /second), but 7% showed medium responses (between 30 and 50 spikes/second) and strikingly, 10% show strong responses (between 57 and 115 spikes/second). The asterisk indicates significant differences between the control-subtracted net number of spikes and a threshold median value (10 spikes/second), as explained in material and methods (one-sample signed rank test; * $p<0.05$). **(C)** Responses of three *S. flava* antennal sensilla to stimulation with the odorants indicated (at 1:100 vol/vol, as used in experiments with the heterologous system; Fig. 3). Basiconic sensilla type 1, located proximally, houses an OSN responding to all ITCs except IBITC and SBITC; basiconic sensilla type 2, located in the middle of the antennae, houses an OSN responsive to all ITCs except PITC; trichoid sensilla located distally houses an OSN responsive to all ITCs except SBITC. Thus, these three sensilla types contain OSNs having different ITC-receptive ranges according to side chain compound variations, as shown for Or67b paralogs (Figures 3 and 4).

which were separated best at 1:1,000 vol/vol (Fig. 3B, right). Thus, these results show that the three paralogs have different ITC-specific sensitivity and selectivity.

***S. flava* antennal OSNs are responsive to ITCs**

Sfla Or67b paralogs selectively respond to diverse ITCs when expressed in the heterologous system (Figure 4). We next investigated whether ITC-sensitive Ors are indeed present in wild-type *S. flava* flies. To test this, we recorded the olfactory responses of *S. flava* antennal basiconic and trichoid sensilla (n= 60 OSNs recorded from 27 sensilla in three different *S. flava* individuals; Figure 4-figure supplement 5) to stimulation with BITC 1:1,000 vol/vol. We chose this compound at this low, naturally-occurring concentration, because it evokes responses from all *S. flava* paralogs (Figure 4B; middle panel). We found that 17% of recorded OSNs showed medium to very strong responses (median=70, ranging from 30 to 115 spikes/second) to stimulation with BITC (Figure 4-figure supplement 5A, B).

We next investigated if wild-type *S. flava* OSNs show distinct selectivity to ITCs, as observed for each of the three *S. flava* Or67b paralogs in the heterologous systems (Figure 4). For this, we tested the responses of *S. flava* sensilla to stimulation with eight ITCs: BITC, AITC, IBITC, SBITC and PITC (all tested at 1:100 vol/vol, as used in experiments with the heterologous systems; Figure 4-figure supplement 5C). We found three types of OSNs that responded to ITCs but had different selectivity: (1) basiconic sensilla located proximally, with OSNs responding to all ITCs except IBITC or PITC: (2) basiconic sensilla located more distally, near the middle of the antennae, bearing OSNs responding to all ITCs except PITC: and (3) trichoid sensilla located distally, with OSNs responding to all ITCs except SBITC. Thus, wild-type *S. flava* has at least three OSNs housed in different antennal sensilla with differential ITC-receptive ranges. However, their odorant selectivity does not perfectly match that of the OSNs expressing the *S. flava* paralogs in the empty neuron heterologous systems (compare with Figure

4B). This is not unexpected, as the native system likely houses other accessory elements and proteins such as OBPs that might modulate OSN responses (Ronderos et al., 2014; Larter et al., 2016).

Activation of *S. flava* Or67b paralogs requires the presence of the ITC functional group

The finding that *Sfla* Or67b paralogs respond to ITCs with differential selectivity (Figure 4) suggests that not only the ITC functional group (–N=C=S), but also the size of the molecule, determine binding specificity. These two factors are indeed important to evoke responses from TrpA1, a gustatory receptor that is reported to detect electrophiles (Hinman et al., 2006; Wei et al., 2015). We started by investigating if *Sfla* Or67b paralogs are also activated by non-ITC electrophiles. We found that none of the eleven tested electrophiles evoked responses from *Sfla* Or67b1 or b3 (Figure 3A). However, stimulation with three of those electrophiles (citronellal, S-Allyl 2-propene-1-sulfinothionate, and diallyl disulfide) increased activity from *Sfla* Or67b2 (median values = 76, 22 and 17 spikes/second, respectively; Figure 3A and B, 4th panel) albeit their responses were not significant (p-values= 0.063-0.078). Overall, these results suggest that *Sfla* Or67b are narrowly tuned to ITCs but not to electrophiles generally, unlike TrpA1.

We next asked whether the presence of the ITC functional group (–N=C=S) is necessary for evoking responses from *Sfla* paralogs. We tested the responses of *Sfla* Or67bs to two linkage isomers, BITC (which bears the ITC functional group, –N=C=S) and benzyl thiocyanate BTC (which bears the thiocyanate functional group, -S≡C-N) (Figure 3B). BITC produced robust activity in neurons expressing any of the three paralogs (one-sample signed rank tests, $p<0.05$), while BTC had no effect ($p>0.05$; Figure 3A, B). This differential activation pattern is thus due to the presence of different functional groups (ITC vs. TC), as these compounds are not only isomers but also have similar volatilities.

***S. flava* is attracted to mustard plant odors and volatile ITCs**

Because *S. flava* paralogs, but not the *Spal* and *Dmel* orthologs, respond selectively to ITCs, we hypothesize that *S. flava*, but neither *D. melanogaster* nor *S. pallida*, are attracted to these odorants. We tested this using a dual-choice olfactory assay (based on Reisenman et al., 2013; Figure 5 - figure supplement 6) in which flies are allowed to choose between the odorless and the odor-laden arm of a "y-maze" olfactometer, a behavioral assay that is thought to measure long-distance olfactory orientation (*e.g.* Auer et al., 2020). We found that *S. flava* is attracted to arugula and *A. thaliana* leaf odors (two-tailed Binomial tests, in all cases $p<0.05$; Figure 5), while the closely related species *S. pallida* was not attracted to these odors ($p>0.05$ in all cases, Figure 5).

Because both *S. flava* Or67b paralogs and *S. flava* antennal OSNs strongly respond to ITCs, we tested the possibility that the fly's attraction to arugula and *A. thaliana* leaf volatiles is solely mediated by these odorants. To address this, we tested whether flies differentially oriented towards leaf volatiles from *A. thaliana* GKO and *PAD3*, mutants that differ in their ability to produce glucosinolates, the precursors of ITCs (Beekwilder et al., 2008; Sønderby et al., 2007; Zhao et al., 2002; Schuhegger et al., 2006). We found that the volatile organic compounds released by either of these two mutant plant lines were similarly preferred over clean air (Figure 5), indicating that other non-ITC host-plant VOCs can mediate olfactory attraction, at least at a distance. Importantly, non-host VOCs (such as those released from tomato leaves) did not attract or repel *S. flava* ($p>0.05$, Figure 5; males were attracted to volatiles from all plants; Figure 5 – figure supplement 7). Tomato volatiles were highly attractive to the closely related species *S. pallida*, which is in accordance with the fact that it can be reared in medium containing rotten tomato in the laboratory (Maca, 1972). *A. thaliana* leaf volatiles did not attract *S. pallida* but

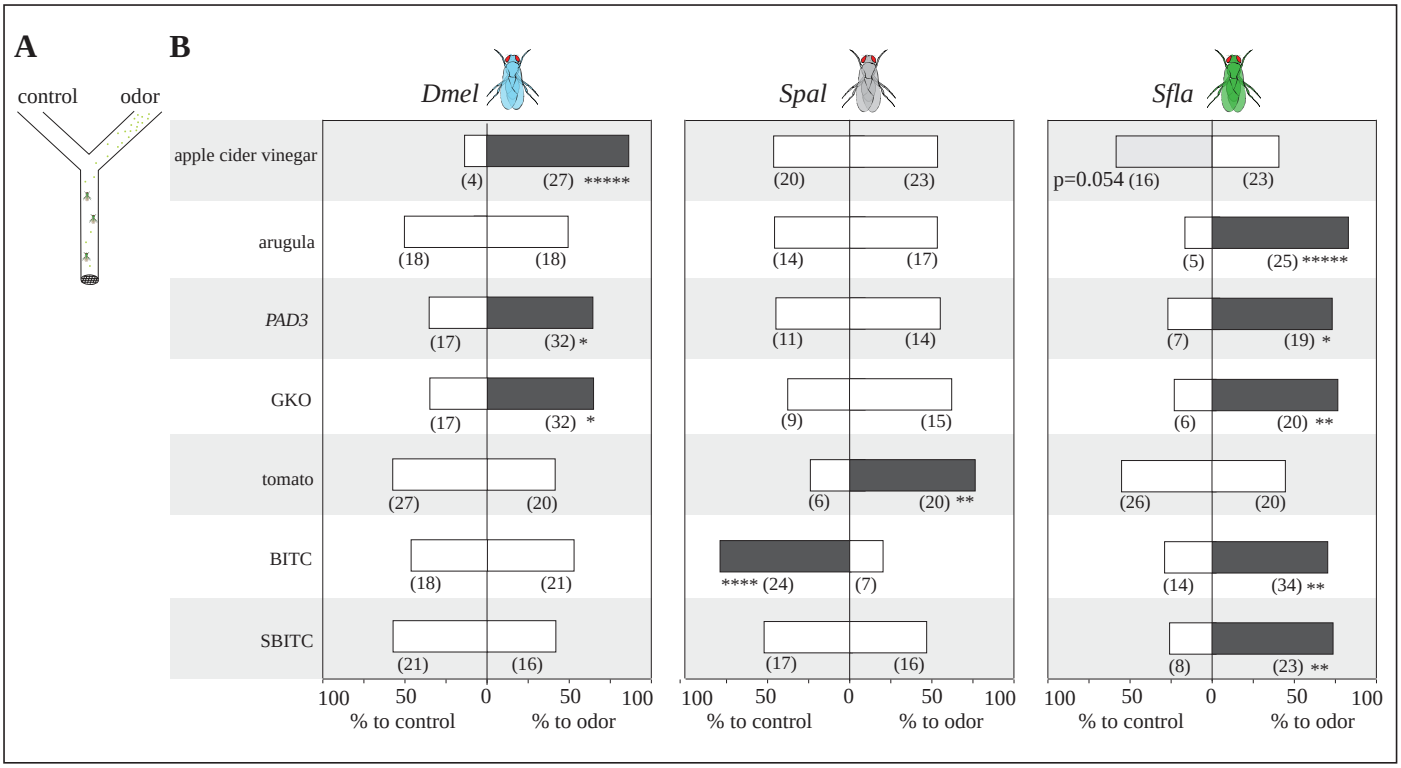

Figure 5 Olfactory behavioral responses of *S. flava* and its microbe-feeding relatives *S. pallida* and *D. melanogaster* to mustard plant volatiles and ITCs.

(A) Schematic representation of the dual choice y-maze used to test the olfactory responses of flies (see details in Figure 5-figure supplement 6). One arm of the maze offered constant odor/odorant airflow (apple cider vinegar, plant leaves, or a 1:100 vol/vol solution of single odorant compounds), while the control arm offered a constant odorless airflow. In each test a group of non-starved flies (n=3-4) was released at the base of the maze and allowed to choose between the two arms of the maze. Each test (maximum duration=5 min) ended when the first insect (out of all released) made a choice. **(B)** Olfactory behavioral responses of *D. melanogaster*, *S. pallida* and *S. flava* to the odors and odorants indicated to the left. Data represent the percentage of tests in which animals choose the odorous or odorless (*i.e.* control) arms of the maze; numbers between parenthesis indicate the number of tests with choices for one or the other arm. For each fly species and odor/odorant compound, data was analyzed using two-tailed Binomial tests (* $p<0.05$, ** $p<0.01$, **** $p<0.001$). *S. flava* was significantly attracted to volatiles from all mustard plant species and to both ITC compounds tested, and slightly avoided ($p=0.054$) apple cider vinegar volatiles; flies were not attracted neither repelled by leaf volatiles from the non-host plant tomato (see Figure 5-figure supplement 7 for tests with males). *D. melanogaster* was strongly attracted to apple cider vinegar and to a lesser extent to *A. thaliana* volatiles, but was not attracted neither repelled by arugula leaf volatiles or ITCs. *S. pallida* was only attracted to tomato odors and significantly avoided BITC. The species-specific behavioral responses to the various odors and odorants tested reflect niche partitioning and to a lesser extent, the distinct odorant-receptive range of Or67b homologs.

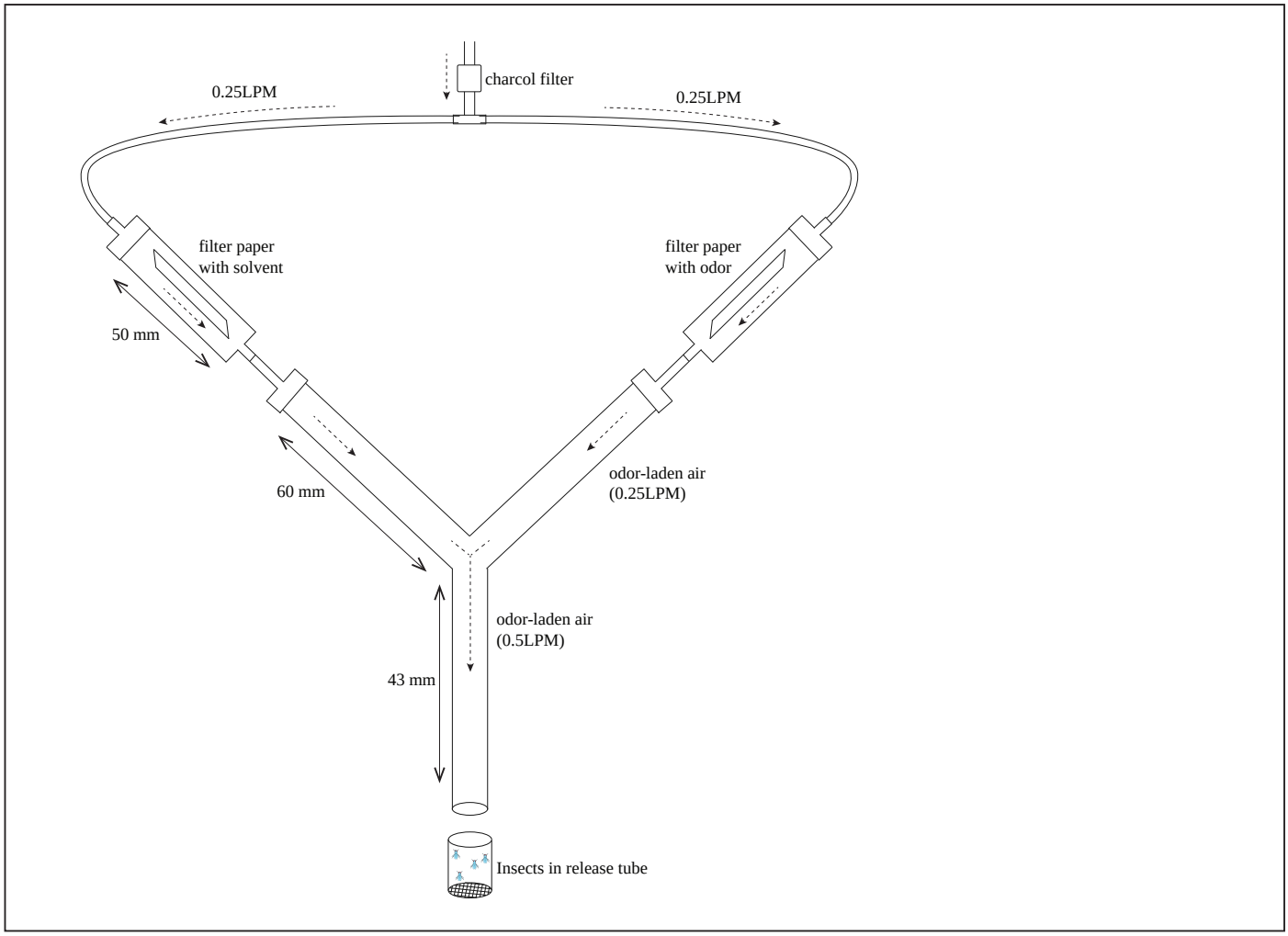

Figure supplement 6 Detailed schematic representation of the device used to test olfactory behavioral responses. The responses of insects were tested using a dual-choice "Y-shaped" olfactometer modified from one previously published (*e.g.* Reisenman et al. 2013). The "Y" part of the maze was a propylene connector; the open ends of the arms of the connector were each connected to a 1 or a 5 ml plastic syringe containing the odor/odorant stimulus or the control stimulus. Single odorants (or the solvent control) were loaded in a piece of filter paper which was placed in 1-ml syringes; plant leaves (or wet tissue paper as a control) were placed in 5-ml syringes. Charcoal-filtered air was delivered to the maze and adjusted to 0.5 liters per minute (LPM) using a flowmeter; thus, odor airflow in each arm was 0.25 LPM, and at the base of the maze again 0.5 LPM. Three-four insects were placed in individual open-top releasing containers 2 hours before experiments with a piece of tissue paper soaked in distilled water (in the case of *D. melanogaster*) or ca. 20 hours with a cotton piece soaked in honey water solution (in the case of *S. flava*). For experiments in Figure 6C, transgenic flies were placed in individual releasing containers with a piece of tissue paper embedded in water 24 hours before experiments. Before each test, the release container was placed during 40-60 seconds in ice to slow down insect activity. Each test started when the open-top of the insect container was carefully slid into the open end of the long arm of the "Y". Thus, upon being released, insects could walk upwind towards the "decision point" (intersection of the short and long arms of the "Y") and turn towards either the odor-laden or the odorless arm of the maze. A choice was considered as such only if the insect walked past at least 1 cm into the arm, orienting upwind. Although 3-4 insects were released in each test (to increase the possibility that at least one insect made a choice), only the first choice (control arm or odorous arm) was recorded. If a second insect made a choice for the other arm within 5-7 seconds of the first insect, the test was discarded. Each test lasted a maximum of five minutes, and each group of insects were used only once. The position of the control and the test arms was switched every one or two tests to control for positional asymmetries. The whole device was illuminated with white light or green light (in the case of tests with *S. flava*).

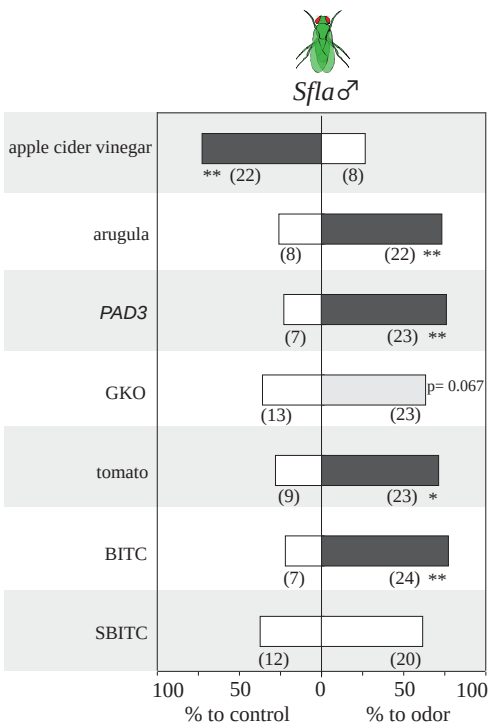

Figure supplement 7 Behavioral responses of *S. flava* males.
Experiments were conducted with mated males as shown in Figure 5. Data indicate the percentage of insects that choose one or the other arm of the maze (x-axis) and the number of choices (numbers in parentheses). Asterisks indicate significant differences from the 50% random expectation (Binomial tests; * $p<0.05$, ** $p<0.01$).

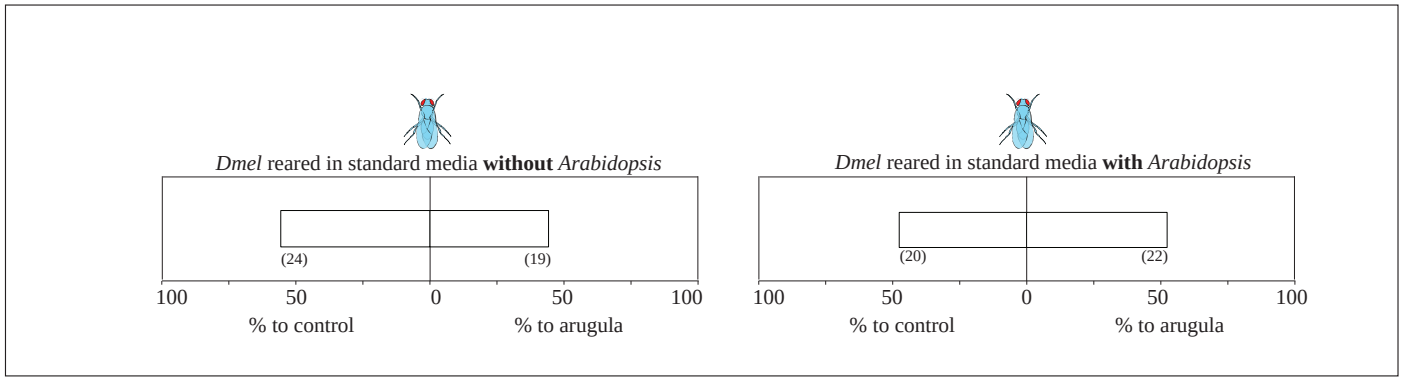

Figure supplement 8 Behavioral responses of *D. melanogaster* reared in the presence (left) or absence (right) of *A. thaliana*.

Experiments were conducted as explained in Figure 5. One arm of the maze offered arugula volatiles, while the other arm offered clean air. Data represent the percentage of tests in which animals choose the odorous or odorless arms of the maze; the numbers between parentheses indicate the number of tests with choices for one or the other arm. In both cases, flies were not attracted nor repelled by arugula volatiles (Binomial tests, $p>0.05$ in both cases). These results indicate that larval experience with intact mustard plants does not suffice to produce orientation towards mustard plant volatiles in adult *D. melanogaster*, and suggest that the orientation of adult *S. flava* towards mustard plant volatiles and ITCs is innate and does not result from experience with these compounds during the larval stage.

attracted the more distant species *D. melanogaster* (Figure 5). Thus, although ITCs might not be required for long-distance olfactory attraction towards hostplants in *S. flava*, the differential olfactory behavior of *S. pallida* and *S. flava* towards mustard plants would ensure niche partitioning.

As expected, *D. melanogaster* was strongly attracted to apple cider vinegar volatiles (Figure 5), a fermented odor source, in agreement with previous studies (Faucher et al., 2013). Both *S. flava* and *S. pallida*, in contrast, distributed at random between the odor-laden and odorless arm of the maze. *S. flava* slightly preferred the odorless arm of the maze (59%; $p=0.054$; Figure 5; $p<0.05$ in the case of *S. flava* males, Figure 5 - figure supplement 7). Overall, these results suggest that *S. flava* is specifically attracted to mustard plant volatiles and tends to avoid odor sources that attract *D. melanogaster*, which are characterized by the presence of acetic acid and various ester, carbonyl, and hydroxyl-containing compounds (Aurand et al., 1966; Cousin et al., 2017).

We then investigated whether ITCs alone can mediate olfactory attraction in *S. flava*. We used a "reverse chemical ecology approach" (*e.g.* Choo et al., 2018), in which behavior is probed based on the identification of bio-active semiochemicals probed in neurons (whether in native or heterologous systems). Thus, we chose BITC and SBITC because these compounds evoked distinct odor responses from *Sfla* Or67b paralogs (Figure 4; BITC strongly activates all *S. flava* paralogs, while SBITC activates only *Sfla* Or67b3). Furthermore, these compounds are chain isomers and have similar chemical properties, including volatility. We found that *S. flava* was attracted to both BITC and SBITC ($p<0.05$ in both cases; males were also significantly attracted to BITC but not to SBITC, Figure 5-figure supplement 7). Interestingly, *S. pallida* avoided BITC ($p<0.05$; Figure 5), which correlates with the finding that this species sometimes feed on decomposing – but not intact – mustard plant leaves (*S. pallida* was neither attracted nor repelled by SBITC, $p>0.05$; Figure 5). As observed in experiments with

whole plants, these results again reinforce the conclusion that the differential species-specific olfactory behavior towards mustard plants and ITCs contributes to niche partitioning between these two closely related *Scaptomyza* species. As predicted, based on the lack of responses to stimulation with ITCs (Figure 3; Goldman-Huertas et al., 2015), *D. melanogaster* flies were not attracted neither repelled by any of the two ITCs tested ($p>0.05$ in both cases; Figure 5). *D. melanogaster* flies were also not attracted to ITCs in dual-choice trap assays (Wilcoxon-matched pairs test on preference indexes, $p=0.84$, $n=35$ tests each with 15-20 female flies; data not shown). In summary, these experiments demonstrate that ITCs are sufficient for olfactory orientation in *S. flava*, and that these compounds do not evoke olfactory-oriented responses in its microbe-feeding relatives.

We explored the possibility that odorant exposure during the larval stage affects adult choice ("Hopkins host selection principle"; Anderson et al., 2013; Anderson and Anton, 2014; Barron, 2001). Because *S. flava* are mustard-plant obligated herbivores (Gloss et al., 2019) and cannot be reared in artificial media, we investigated this possibility by rearing parallel cohorts of *D. melanogaster* on standard fly food in the presence or absence of mustard plants, and tested the orientation of flies towards arugula leaf volatiles. We found that both groups of flies showed no preference nor deterrence towards mustard plant volatiles ($p>0.05$ in both cases, Figure 5-figure supplement 8). Thus, these results are consistent with the hypothesis that the olfactory attraction of adult *S. flava* towards mustard plant volatiles is innate.

***Sfla* Or67b3 is sufficient to evoke attraction to ITCs in an attractive olfactory circuit**

Because *SflaOr67b* paralogs selectively respond to ITCs, we asked if these receptors are necessary and/or sufficient to mediate olfactory attraction to these odor compounds. In principle this could be readily tested using Or67b knock-out *S. flava* mutants generated via CRISPR-Cas9.

Unfortunately, this technology is not yet suitable for *S. flava*, as females oviposit eggs inside plant leaves, making it difficult to inject embryos. Instead, we conducted a "gain of function" experiment (*sensu* Karageorgi et al. 2019) in which we tested the olfactory responses of *D. melanogaster* flies expressing *S. flava* paralogs in OSNs which lack expression of their cognate receptors. We choose the Or22a olfactory circuit because it is important for olfactory attraction in diverse drosophilids (Mansourian et al., 2018; Dekker et al., 2006a; Linz et al., 2013). In addition, *S. flava* (but not *S. pallida*) lost its cognate *Or22a* (Goldman-Huertas et al., 2015), raising the possibility that this species expresses one of the triplicated Or67b paralogs in these OSNs. We focused on *Sfla* Or67b3, as this Or has a broader sensitivity and sensitivity to ITC compounds than *Sfla* Or67b1 and b2 (Figures 3 and 4). As before, we used a dual-choice olfactometer and tested flies in which the expression of *Sfla* Or67b3 or *Dmel* Or67b is under the control of GAL4 in the ab3A neuron, as well as the three parental control lines. We tested flies with a lower concentration of BITC (1:1,000 vol/vol) in order to avoid compensatory/un-specific responses caused by the lack of Or22a in *S. flava*. While flies from all the three genetic control groups did not prefer either arm of the maze ($p>0.05$ in all cases, Figure 6A), *D. melanogaster* flies expressing *Sfla* Or67b3 in the ab3A neuron, but not those expressing *Dmel* Or67b, preferred the odor-laden arm of the maze ($p<0.05$, Figure 6A). Therefore, *Sfla* Or67b3 can confer odor sensitivity to ITCs when expressed in the *D. melanogaster* Or22a attractive olfactory circuit.

Or67b OSNs mediate attraction in *D. melanogaster*

We also tested the possibility that *Sfla* Or67b3 can confer ITC attraction in the Or67b olfactory circuit. In this experiment (unlike the Or22a experiments) flies did not lack expression of the endogenous Or67b. While control GAL4 flies were neither attracted nor repelled by BITC, flies

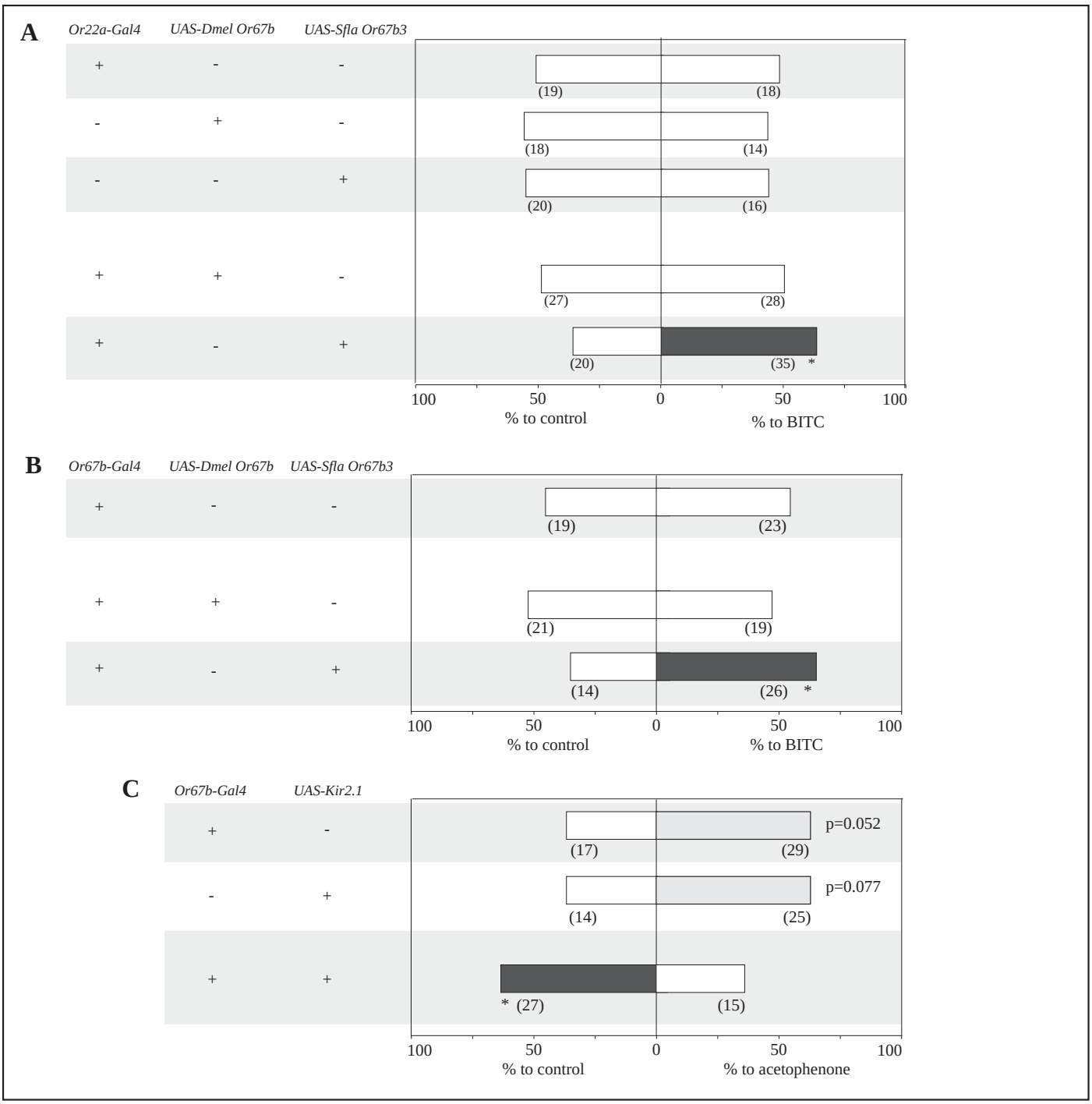

Figure 6 Ectopic expression of *Sfla* Or67b3 in Or22a OSNs or Or67b OSNs conferred behavioral attraction to BITC in *D. melanogaster*.

(A) Behavioral responses of *D. melanogaster* flies expressing *Dmel* Or67b or *Sfla* Or67b3 in Or22a OSNs lacking their cognate olfactory receptor to stimulation with BITC (1:1,000 vol/vol). Experiments were conducted and analyzed as explained in the caption to Figure 5. Control flies (first three groups), as expected, were not attracted nor repelled by BITC. In contrast, flies expressing *Sfla* Or67b3, but not those expressing *Dmel* Or67b, were significantly attracted to the odorant (* $p<0.05$). **(B)** Same as **A**, but flies expressed *Dmel* Or67b or *Sfla* Or67b3 in Or67b OSNs (note that flies have the endogenous Or67b expressed in OSNs, in addition to the transgene). Only flies carrying the *S. flava* transgene were significantly attracted to BITC (* $p<0.05$). **(C)** Behavioral responses of *D. melanogaster* flies expressing a silencer of synaptic activity (Kir2.1) in Or67b OSNs, along with the responses of the two parental control lines (transgenes indicated to the left) in response to acetophenone (1:50,000 vol/vol), a strong *Dmel* Or67b activator (Figure 3; Münch and Galizia, 2016) which at low concentrations evokes activity from Or67b OSNs only (Münch and Galizia, 2016). Experiments were conducted as explained in the caption to Figure 5, but flies were starved 24 hs previous to testing. As observed for wild-type flies (Strutz et al., 2014), control flies tended to attract to low concentrations of acetophenone, while flies with Or67b OSNs silenced lost attraction and were instead repelled by the odorant (as wild-type flies tested with higher concentrations of acetophenone; see main text and Strutz et al., 2014). All these findings suggest that the Or67b circuit mediates olfactory attraction in various drosophilid flies including *S. flava* (Crowley-Gall et al., 2016)

expressing *Sfla* Or67b3, but not those expressing an additional copy of *Dmel* Or67b, were attracted to BITC (Figure 6B).

These results prompted us to investigate whether the Or67b olfactory circuit mediates olfactory attraction. To test this possibility, we took advantage of the known odorant receptive range of its cognate Or. *Dmel* Or67b is highly sensitive to acetophenone (Figure 3; Münch and Galizia, 2016), and in turn this odorant attract flies at low concentrations (but causes repellence at higher concentrations; Strutz et al., 2014). We first confirmed these results under our experimental conditions. We tested flies in the y-maze as before, and found that 24 hours-starved *D. melanogaster* (Canton-S strain) are attracted to lower concentrations of acetophenone (1:10,000 vol/vol; $p<0.05$; 75% odor attraction, n=26), but repelled by relatively high concentrations (1:100 and 1:1,000 vol/vol; Binomial tests, $p<0.05$ in both cases, 91% and 81% of flies respectively preferred the odorless arm of the maize, n=11 and n=16). These findings, along with those showing that low concentrations of acetophenone evoke activation only from Or67b, support the conclusion that attraction towards acetophenone is mediated by the Or67b olfactory circuit. We further tested this possibility using UAS-Kir2.1, an inwardly rectifying potassium channel, to inhibit neuronal activity (Baines et al., 2001) under the control of Or67b-Gal4. We found that flies carrying Kir2.1 in Or67b OSNs lost the attraction to low concentrations of acetophenone and instead showed odorant aversion (Figure 6C; Binomial test, $p<0.05$). Gal4- and UAS-control flies showed slight odorant attraction (63% in both cases; Figure 6C; $p=0.052$ and 0.077), consistent with our experiments with wild-type flies (see above) and previous studies (Strutz et al., 2014). We speculate that aversion to low odorant concentrations persists in experimental flies because in absence of Or67b other Ors sensitive to acetophenone (*e.g.* Or10a; Münch and Galizia, 2016), which might mediate repulsion, remained activated.

Discussion

The majority of herbivorous insects specialize on one or a few host plant lineages (Jaenike, 1990) that synthesize similar secondary compounds. While these toxins serve to defend against susceptible herbivores, these ancestrally aversive molecules become attractants to those herbivores that become specialized on a particular set of toxic host plants. Here we investigated how attraction to toxic host-plants evolves mechanistically using *S. flava*, a species within a recently-derived herbivorous fly lineage that attacks mustard plants and their relatives in the Brassicales. We first searched for changes in chemoreceptor proteins and found a candidate Or lineage (*Or67b*) likely under positive selection which was triplicated in a recent ancestor of *S. flava*, resulting in three divergent and paralogous olfactory receptor genes (*Sfla Or67b1-b3*; Figure 1). Next, using an heterologous system for expression of olfactory receptors, the two "empty neuron" systems of *D. melanogaster*, we investigated the odor-response profiles of these receptors. All three *Sfla* Or67b paralogs specifically respond to stimulation with mustard-plant odors and volatile ITCs (Figures 2 and 3). Furthermore, each paralog responded to a specific subset of ITCs (Figure 4). In contrast, *S. pallida* and *D. melanogaster* orthologs did not respond to ITCs, but instead showed strong responses to stimulation with apple cider vinegar and a broad range of aldehydes, alcohols and ketones (Figures 2 and 3), consistent with their microbe-feeding habits and previous findings (Galizia and Sachse, 2009). Consistent with these results, recordings from *S. flava* antennal sensilla revealed the presence of three OSN classes with different breadth of tuning to ITCs (Figure 4-figure supplement 5). In behavioral experiments, we found that *S. flava*, but neither *S. pallida* or *D. melanogaster*, is specifically attracted to host-plant volatiles, including volatile ITC compounds (Figure 5). Moreover, in "gain of function" experiments, expression of *S. flava* paralogs in two different *D. melanogaster* olfactory circuits was sufficient to confer odor-oriented responses to ITCs. Furthermore, these results along with previous findings, indicate that the Or67b olfactory circuit might

mediate olfactory attraction in various drosophilid species. Our results suggest an evolutionary mechanism, gene duplication and specialization, by which mustard specialists can evolve olfactory attraction towards ancestrally aversive chemical compounds.

Evolutionary path of olfactory receptor specialization

The gene triplication event that occurred in the *S. flava* lineage described here raises several molecular evolutionary questions. We found that the three *Sfla* paralogs selectively respond to different subsets of ITCs across a range of odorant concentrations (Figures 3 and 4). These results support the specialization model, as neither of the two outgroup Or67b proteins (*Spal* Or67b and *Dmel* Or67b) responded to ITCs (neofunctionalization), coupled with the fact that the *Sfla* Or67b paralogs are narrowly tuned to specific ITCs (subfunctionalization). Our conceptual model for this evolutionary hypothesis (Figure 7) predicts that the ancestral Or67b (labeled "a" in Figure 7) was broadly tuned to alcohols and ketones, and diverged to a mustard oil-specific (ITC-specific) Or67b ("b" in Figure 7) that lost sensitivity to these compounds, and gained responsiveness to ITCs in the *S. flava* lineage. This *Or* triplicated during evolution, ultimately giving rise to three Ors with distinct, but partially overlapping, ITC selectivity ("c" in Figure 7): *Sfla* Or67b1, which is most responsive to 4MBITC across a range of concentrations; *Sfla* Or67b2, which is sensitive to AITC; and *Sfla* Or67b3, which is most sensitive to SBITC and PITC. Interestingly, all *S. flava* paralogs are responsive to BITC, which is the structurally simplest of all the ITC compounds tested here.

***S. flava Or67b* triplication resulted in paralogs with distinct but overlapping ITC selectivity**

S. flava Or67b paralogs specifically responded to ITCs, but have different selectivity (Figures 3, 4 and 7). *Sfla* Or67b3 is the most promiscuous of all paralogs, responding to all ITCs across a broad

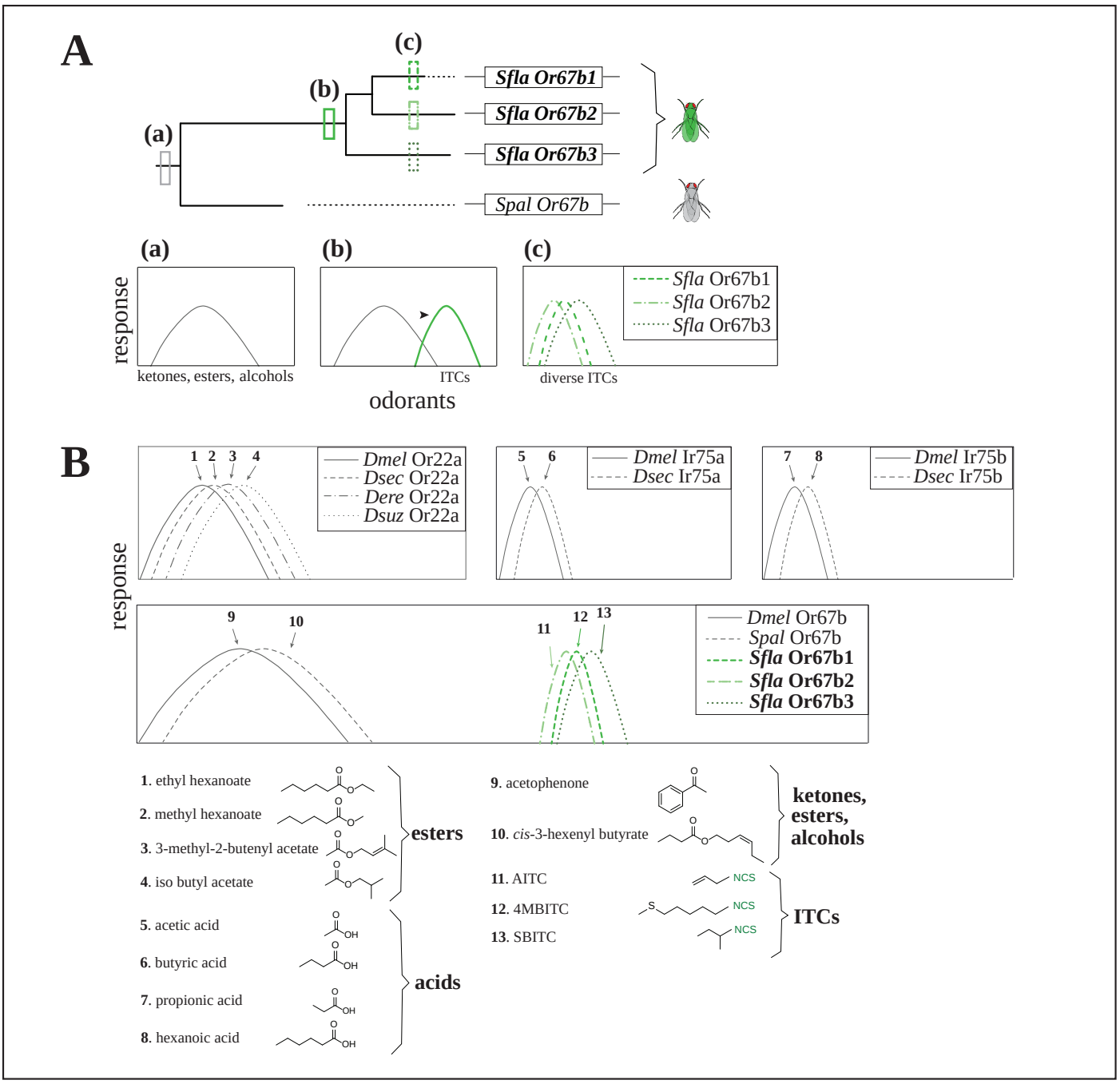

Figure 7 A model for the evolution of *Or67b* and comparison with the known evolution of other Or orthologs in drosophilid flies.

(A) Model for the evolution of *Scaptomyza Or67b*. The evolution of this Or begins with a shift in the ligand specificity of an ancestral Or67b (a) tuned to *Dmel* activators, GLVs, and ITCs (neofunctionalization, b). Subsequent gene triplication of *Sfla Or67b* gave rise to three paralog *Or67b* genes (*Sfla Or67b1*, *Sfla Or67b2*, and *Sfla Or67b3*; c), each of them having different but overlapping ITC odorant-receptive ranges (Figures 3 and 4). **(B)** Evolution of drosophilid orthologs with known ligand specificities (Or22a, Ir75a, Ir75b, top; and Or67b, bottom). The Or22a orthologs from *D. melanogaster* (*Dmel* Or22a), *D. sechellia* (*Dsec* Or22a), *D. erecta* (*Der* Or22a), and *D. suzukii* (*Dsuz* Or22a) are all strongly activated by species-specific host-derived esters (compounds 1-4; top left; Dekker et al., 2006b; Keesey et al., 2019; Linz et al., 2013; Mansourian et al., 2018). The Ir75a orthologs from *D. melanogaster* (*Dmel* Ir75a), and *D. sechellia* (*Dsec* Ir75a) are strongly activated respectively by the acid compounds 5 and 6 (Prieto-Godino et al., 2017). Similarly, the Ir75b orthologs from *D. melanogaster* (*Dmel* Ir75b), and *D. sechellia* (*Dsec* Ir75b) are respectively activated by the acids 7 and 8 (top right; Prieto-Godino et al., 2016). *Dmel* Or67b and *Spal* Or67b are strongly activated respectively by acetophenone and *cis*-3-hexenyl butyrate (compounds 9 and 10), while *Sfla* Or67b paralogs are activated by ITCs only (bottom, paralog-specific activation by compounds 11-13). Note that Or22a, Ir75a and Ir75b orthologs are all divergent but activated by ligands belonging to a single chemical class (whether esters or acids). On the other hand, the ligands of orthologs Or67b from *Dmel* and *Spal* are responsive to a variety of chemical classes which include alcohols, aldehydes and ketones, whereas *Sfla* Or67b orthologs are responsive to ITCs, an entirely different compound chemical class.

range of concentrations (Figure 4B). *Sfla* Or67b2 responded to fewer compounds, particularly at low-to-medium concentrations (Figure 4B). Interestingly, both BITC and 3MIPTC (iberverin) activated all *S. flava* paralogs at all concentrations (with the exception of *Sfla* Or67b2 at the lowest concentration of BITC; Figure 4B). Of particular significance is that in the specialist moth *Plutella xyllostella*, 3MPITC was also recently reported to activate ITC-selective Ors and to stimulate oviposition (Liu et al., 2020). PITC has similar physiological and behavioral effects in this moth species, and in *S. flava* it selectively activates Or67b3 across the range of concentrations tested (Fig. 3A, B). Remarkably, these two compounds have been identified in various Brassicales plants, including *A. thaliana* (Liu et al., 2020). Thus, these results suggest that these two ITC compounds may also be important for host-plant selection in *S. flava*, and in Brassicales specialists from other orders.

Each paralog is most responsive to different ITC subsets (Figure 3B) and each compound differentially activates, with a few exceptions, more than one paralog in a concentration-dependent manner (Figure 4). Accordingly, different ITCs are best discriminated at different concentrations (*e.g.* PITC and SBITC have overlapping coordinates in the odor space at 1:100 vol/vol but are well separated at 1:1,000; Figure 4C). Thus, not only ITC identity, but also compound ratios and concentrations are likely to be used for host plant detection by Brassicales specialists, a strategy that resembles that reported in *Ostrinia* spp. moths, where several species use varying ratios of two unique pheromone components for identifying con-specifics (*e.g.* Miura et al., 2009).

Importantly, our electrophysiological recordings show that 17% of recorded *S. flava* antennal sensilla (n=60, basiconic and trichoid sensilla; Figure 4 –figure supplement 5) house OSNs that respond strongly to stimulation with ITCs (Figure 4 – figure supplement 5). As observed for the Or67b paralogs expressed in the two heterologous systems (Figures 3 and 4), these OSNs can be categorized in three functional classes due to their distinct ITC selectivities (Fig. 4C). The odor tuning of these OSNs does

not completely match that of the three Or67b paralogs expressed heterologously, which may be due to the presence of divergent cellular elements (*e.g.* OBPs), cellular properties, expression of chemosensory proteins, etc., in the native *S. flava* system. A common feature is that all these three OSN types are responsive to stimulation with BITC (Figure 4 – figure supplement 5), as are all three *S. flava* Or67b paralogs (Figures 3 and 4). Thus, these results indicate that the three Or67b paralogs are likely expressed and functional in the antenna of *S. flava*, albeit their sensilla and cellular localization remains to be characterized (Or67bs are expressed in adult flies; Figure 1-figure supplement 3).

An olfactory receptor sensitive to a new niche-specific chemical class of odorant compounds

Across *Drosophila*, orthologous chemoreceptors respond in a species-specific manner to ecologically relevant ligands. The best studied of these is *Or22a*, which has been evolutionarily lost in *S. flava* but not in *S. pallida (Goldman-Huertas et al., 2015)*. In several drosophilids Or22a is tuned to different ecologically relevant esters: ethyl-hexanoate in *D. melanogaster* (Mansourian et al., 2018), methyl-hexanoate in *D. sechellia* (Dekker et al., 2006b), 3-methyl-2-butenyl acetate in *D. erecta* (Linz et al., 2013), and isobutyl acetate in *D. suzukii* (Keesey et al., 2019). Similarly, Ir75a is tuned to acetic acid in *D. melanogaster* and to butyric acid in *D. sechellia* (Prieto-Godino et al., 2016), while Ir75b is respectively tuned to butyric acid and to hexanoic acid in these two species (Prieto-Godino et al., 2017). The finding that each of these chemoreceptors respond to different odorants within a unique chemical class (*i.e.* esters in the case of Or22a and acids in the case of Ir75a and Ir75b) suggests that the chemical niche space and underlying sensory mechanisms of odorant detection are relatively conserved across species. In contrast, while *Spal* Or67b and *Dmel* Or67b respond to alcohols, aldehydes and ketones, *Sfla* Or67b paralogs respond almost exclusively to volatile ITCs, thus acquiring a sensitivity to an entirely different chemical compound class (Figures 3 and 4). Accordingly, our structure-activity studies indicate

that the presence of the ITC functional group (–N=C=S) is key to activate these olfactory receptors (Figure 3A, B). We hypothesize that the striking difference in odor selectivity among Or67b orthologs results from the niche shift that occurred either during the evolution of herbivory in *Scaptomyza* or during specialization on mustards.

Apart from the ITC-sensitive Ors described here, a recent study in the moth *Plutella xylostella*, also a specialist of Brassicales plants, reported two Ors (Or35 and Or49) that are also selective for ITCs and important for oviposition (Liu et al., 2020). As mentioned above, it is relevant that these Ors are selective to three ITC compounds including two (3MPITC and PITC) that activate *S. flava* Or67b paralogs strongly. The sequence of the moth Ors are highly divergent from those of the *S flava* Or67b paralogs, precluding inferences about the molecular structural features that determine binding specificity. Nevertheless, the putative binding pockets of Ors are mostly unknown, although a recent study identified residues of Or proteins which are involved in ligand binding and are likely “hot-spots" for functional evolution of Or22a. (Auer et al., 2020)

Apart from *S. flava* and *P. xylostella* ITC-sensitive Ors, the only other known chemoreceptor responsive to ITCs is TrpA1, a highly conserved taste receptor that detects electrophiles *via* covalent modification of cysteine residues (Hinman et al., 2006; Al-Anzi et al., 2006; Arenas et al., 2017; Macpherson et al., 2007). *S. flava* Or67bs have a larger number of cysteine residues (14-16) than the *D. melanogaster* and *S. pallida* orthologs (12-13 cysteine residues). This suggests the possibility that these residues are involved in the detection of volatile ITCs, although ITC-selective Ors in *P. xylostella* have a much lower proportion of cysteine amino acids.

An olfactory receptor mediating attraction to mustard host plants

At the behavioral level, we found that ITCs and mustard plant volatiles, but not non-mustard plants which nevertheless are very fragrant (such as tomato), attract *S. flava* (Figure 5 and Figure 5 – figure supplement 7). These results strongly suggest that ITCs are key compounds mediating host-plant olfactory attraction. At the same time, we found that mustard plant odors and ITCs activate *Sfla Or67b* paralogs strongly (Figure 2, 3). The simplest interpretation of these results is that Or67b paralogs mediate olfactory attraction of *S. flava* towards mustard host-plant volatiles. This interpretation is further supported by our "gain of function" experiments (Figure 6), in which we found that expression of *Sfla* Or67b3 in either of two olfactory circuits in *D. melanogaster* flies is sufficient to mediate olfactory attraction to BITC.

Our findings, along with previous reports, suggest that the Or67b circuit is involved in mediating host attraction in drosophilid flies. We found that expression of *Sfla* Or67b3 in the *D. melanogaster* Or67b circuit can confer behavioral attraction to ITCs (Figure 6B). Furthermore, we found that silencing Or67b OSNs eliminates olfactory-mediated attraction to its cognate ligand (acetophenone; Figure 6C). OSNs expressing *Dmel* Or67b respond strongly to stimulation with apple cider vinegar (Figure 2), a potent *Drosophila* attractant (Liu et al., 2020; Figure 5), which collectively suggests that the Or67b circuit mediates behavioral attraction. Finally, *D. mojavensis* OSNs expressing Or67b detect aromatic volatile compounds from its cacti hosts (Crowley-Gall et al., 2016; Date et al., 2013). Thus, a plausible hypothesis is that the Or67b neuronal circuit mediates attraction in diverse drosophilids including *D. melanogaster, D. mojavensis* and *S. flava*.

Our results do not preclude that other chemoreceptors and sensory neurons also contribute to mediate olfactory attraction to mustard plant volatiles and ITC compounds. The necessity of the Or67b circuit for orientation towards ITCs and mustard host plants could be probed by testing whether *S. flava* Or67b knock-out mutants can orient towards these odorants. Finally, it remains to be investigated

whether evolution acted on central nervous system cells and circuits to strength and further refine responses to ITC compounds and mustard plant volatiles (Auer et al., 2020; Zhao and McBride, 2020).

In general, taste detection of ITCs in both vertebrates and insects involves the contact chemoreceptor TrpA1 (Hinman et al., 2006; Kang et al., 2010; Macpherson et al., 2007). However, volatile ITCs are widely used to trap agricultural pests of Brassicales (Kergunteuil et al., 2012), and in agreement with this, the antennae of some of these insects, including *S. flava* (Goldman-Huertas et al., 2015; Figure 4-figure supplement 5), respond to volatile ITCs (Piersanti et al., 2020; Renwick et al., 2006; Liu et al., 2020).

Our study further advances our understanding about the detection of volatile ITCs in insects, by identifying and characterizing single chemoreceptor proteins using a heterologous expression system for expression of Ors (Figures 3 and 4), whose odor-response profile is similar to OSN responses in the intact native system (Figure 4-figure supplement 5). Importantly, we uncovered an evolutionary mechanism, gene duplication followed by specialization, by which ITC-responsive olfactory receptors might have evolved in Brassicales specialists. The relevance of these duplications for mustard-plant host specialization is also consistent with the finding that Or67b is duplicated only in *Scaptomyza* spp. mustard specialists and not in other herbivorous *Scaptomyza* species specialized on non-mustard hosts (Figure 1C), and that the number of Or67b paralogs is associated with host-breadth (*e.g*. the mustard specialist *S. montana* uses a smaller number of host species than *S. flava* and has four Or67b copies; G. Wang personal communication). More generally, our results suggest that chemosensory specialization in herbivorous insects can result from relatively simple genetic modifications in the peripheral nervous system that change olfactory receptor tuning, contributing to drastic shifts in niche use.

MATERIALS AND METHODS

Molecular phylogeny of drosophilid olfactory Olfactory Receptors (Or)

Translations of *Ors* from *D. grimshawi*, *D. mojavensis*, *D. virilis* and *D. melanogaster* (builds dgri r1.3, dmoj r1.3, dvir r1.07 and dmel r6.28, respectively) were downloaded from Flybase (www.flybase.org; Thurmond et al., 2019). *S. flava Or* sequences were previously published (Goldman-Huertas et al., 2015). A total of 309 sequences were aligned in MAFFT v7.017 with the E-INS-I algorithm and then manually adjusted (Katoh, 2002). Models were fitted to the alignment using IQ-Tree and tested using the cAIC criterion (Nguyen et al., 2015). A maximum likelihood (ML) phylogeny was generated using the Or protein alignment (in RAxML v8.2.10) with the CAT model of rate heterogeneity with seven distinct categories, the JTT substitution matrix, empirical amino acid frequencies, and 1,000 rapid bootstraps (Stamatakis, 2014). Orco sequences were designated as the outgroup.

Time calibrated molecular phylogeny of *Scaptomyza* and *Drosophila* spp.

A time-calibrated species tree was inferred using the loci: *16S*, *COI*, *COII*, *ND2*, *28S*, *Cad-r*, *Gpdh1*, *Gstd1*, *Marf*, *l(2)tid* (also known as *Alg3* or *Nltdi*) and *AdhR* from 13 spp of *Drosophila* and four *Scaptomyza* spp. *Scaptomyza* sequences were accessed from the genomes in (Kim et al., 2020) and (Gloss et al., 2019) using tblastn searches for protein coding genes and blastn searches for the ribosomal RNA genes. The *Adh Related* gene appears to be deleted in *S. flava* and was coded as missing data. Two uniform priors on the age of *Scaptomyza* and the age of the combined virilis-repleta radiation, Hawaiian *Drosophila* and *Scaptomyza* clade were set as in (Katoh et al., 2017). Protein coding genes were portioned into first and second combined and third position partitions, except *COI and Gpdh1*, which were divided into first, second and third partitions. The partitioning scheme was chosen based on

Partition Finder 2.1.1 (Lanfear et al., 2017) and nucleotide substitution models chosen with IQ-Tree (Nguyen et al., 2015). Model parameters, posterior probabilities, accession numbers and genome coordinates can be found in table 1. Five independent runs of BEAST v2.6.2 (Bouckaert et al., 2019) each with 25 million generations were run logging after every 2500th generation to infer the chronogram with 10% burn-in. Phylogenies were also inferred in RAxML v8.2.12 (Stamatakis, 2014) with the GTR+GAMMA+I model and 1000 rapid bootstraps and with parsimony in PAUP* 4.0a (Swofford, 2002) with TBR branch swapping and 1000 bootstraps. Phylogeny parameters, sequence accession numbers and likelihood scores available in supplementary dataset 1.

Molecular phylogeny of drosophilid *Or67b* genes

Or67b coding sequences (CDS) from *D. grimshawi, D. mojavensis, D. virilis D. sechellia, D. simulans, D. erecta, D. yakuba, D. pseudoobscura, D. persimilis, D. ananassae* and *D. melanogaster* (builds dgri r1.3, dmoj r1.3, dvir r1.07, dsec r1.3, dsim r1.4, dere r1.3, dyak r1.3, dpse r3.2, dper r1.3, dana r1.3 and dmel r6.28, respectively) were downloaded from Flybase (www.flybase.org; Thurmond et al., 2019) and *D. hydei* from Genbank (accession number XM_023314350.2). The *S. pallida* DNA sequence was obtained through PCR and Sanger sequencing as described below. *S. flava* DNA sequences were previously published (Goldman-Huertas et al., 2015). Two more *Scaptomyza* sequences were obtained from (Kim et al., 2020) and (Gloss et al., 2019) including the non-leaf-mining species *S. hsui* and *S. graminum* (a leaf-miner on Caryophyllaceae spp). DNA sequences were aligned, partitioned by codon position, models fitted to all three partitions and chosen according to AICc (GTR+I+G) in IQ-Tree (Nguyen et al., 2015). Trees were inferred using RAxML (v8.2.10) with the GTRGAMMA+I model and 1000 rapid bootstraps, and MrBayes (v3.2.6) setting Nst to 6, nucmodel to 4by4, rates to Invgamma, number of generations to 125,000, burnin equal to 20% of generations, heating to 0.2,

number of chains to 4, runs to 2 and priors set to default setting (Ronquist et al., 2012). An additional parsimony analysis was performed in Paup 4.0 (Swofford, 2002) with TBR branchswapping and 1000 bootstraps. Model parameters, accession numbers and likelihood scores available in Supplementary file 1.

Analysis of molecular evolutionary rates

CDS of homologs of every *Or* gene in *S. flava* found in the twelve *Drosophila* genome builds were aligned to *S. flava Or* CDS. Homology was assessed according to inclusion in well-supported clades in the *Or* translation phylogeny from the twelve species; *S. flava* sequences were previously published (Goldman-Huertas et al., 2015). Sequences were aligned in MAFFT (v7.017) (Katoh, 2002) and adjusted manually to preserve codon alignments. *Or98a*-like genes found in subgenus *Drosophila* species were split into three separate clades, as were a group of *Or83c* paralogs not found in *D. melanogaster*, and a group of *Or85a*-like genes. All examined sequences of *Or46a* contain two alternatively spliced exons, so this gene was analyzed with all gene exon sequences in a single alignment. *Or69a*, however, contains alternatively spliced exons only in species within the subgenus *Sophophora*, and therefore these alternative splice forms were analyzed as separate taxa. Phylogenies were generated for every alignment using PhyML (Guindon et al., 2010) with the GTR+G substitution models. If >70% bootstrap support was found for a topology contrary to the known species topology, or if the *Or* homology group contained duplicates, these trees were used in PAML analyses instead of the species tree.

Our next goal was to identify Or genes experiencing rapid rates of protein evolution. Branch models of sequence evolution were fit using PAML 4.9h (Yang, 2007). A foreground/background branch model was fit for every *S. flava* tip branch and every ancestral branch in a *Scaptomyza*-specific

Or gene duplication clade, and compared in a likelihood ratio test to a null model with one d_N/d_S rate (ratio of non-synonymous to synonymous substitution rate) for every unique phylogeny (75 tests in total). After focusing on *Or67b*, patterns of molecular evolution among the drosophilid *Or67b* homologs were explored using the expanded *Or67b* CDS phylogeny above. Foreground/background branch models were fit for every branch in the *Or67b* phylogeny and the *S. flava* Or67b paralogs clade with likely ratio tests performed as above (34 tests total, table 2). P-values were adjusted for multiple comparisons using the false-discovery rate (FDR) method (Benjamini and Hochberg, 1995). Branch test results and model parameters in Supplementary file 2.

Synteny analysis between *S. flava, S. pallida* and *D. melanogaster Or67b* scaffolds

Five genes up and downstream of each *Or67b* ortholog in *S. flava* were extracted using annotations from a current GenBank assembly (GenBank Assembly ID GCA_003952975.1), shown in Supplementary file 3. These genes are respectively known as pGOIs (genes proximal to the gene of interest, or GOI). To identify pGOIs, we used tBLAST to the *D. grimshawi* genome, the species closest to *S. flava* with a published annotated genome (GenBank Assembly ID GCA_000005155.1). By identifying the pGOI scaffolds, we determined that there was only one copy of *Or67b* in *D. grimshawi*, which is syntenic with the *S. flava Or67b2* ortholog. We determined that this copy is also syntenic with the single *Or67b* copy of *S. pallida* (J. Peláez personal communication) and *D. melanogaster*. In both *S. pallida* and *D. melanogaster*, many of the pGOIs are still present and syntenic with those of *S. flava*, suggesting that the absence of the other *Or67b* orthologs is not a consequence of mis-assembly but instead a real absence (Figure 1 - figure supplement 2).

Fly husbandry and lines

D. melanogaster (Canton-S) were reared using standard cornmeal media, yeast, and agar medium. Isofemale lines of *S. pallida* (collected in Berkeley, California, US) were maintained on Nutri-Fly medium (Genesee Scientific). *S. flava* (collected in New Hampshire, US) were maintained on fresh *A. thaliana* and 10% honey water solution. All flies were cultured at 23°C and 60% relative humidity under a 12-h light/12-h dark cycle. *S. flava* and *S. pallida* were ca. 7-10 days old at the time of experiments; *D. melanogaster* (wild-type or transgenic) were ca. 3-10 days old at the time of the experiments.

M2-MD refers to a CRISPR-Cas9 mediated deletion of *Or22a/b* and a knock-in of *Gal4* and *DsRed* (Gross et al., 2000) by homology directed repair (HDR); in these flies *Gal4* is not functional but *DsRed* is expressed in the eye. The functional absence of *Or22a* and *Or22b* genes in M2-MD flies, or functional absence of *Or67d* in the *Or67d*^{GAL4} line (Kurtovic et al., 2007), were respectively confirmed by electrophysiological analysis on ab3A neurons or at1 neurons (Figure 2-figure supplement 4). The M2-MD line was used to generate flies expressing *Dmel Or67b*, *Spal Or67b*, *Sfla Or67b1* and *Sfla Or67b3* under the control of *Gal4* in the ab3A "empty neuron"(Dobritsa et al., 2003). Similarly, the *Or67d*^{GAL4} line was used to generate flies expressing *Sfla Or67b2* under the control of Gal4 in the at1 "empty neuron"(Kurtovic et al., 2007). The progeny of those crosses was then used for single sensillum recordings and in some cases for behavioral assays. The *UAS-SflaOr67b1*, *UAS-Sfla Or67b2*, *UAS-Sfla Or67b3*, and *UAS-Spal Or67b* strains were generated during this study. The *Or67b-Gal4* fly line (BDSC# 9996) was crossed with *UAS-Or67b* lines or *UAS-Kir2.1* (Baines et al., 2001) for the behavioral assays.

Scaptomyza Or67b gene cloning, UAS line generation, and verification of _S. flava Or67b_ transcription

The *UAS-Or67b* transgene lines were constructed as follows: RNA was extracted from 20–25 days post-emergence adults of both sexes from laboratory-reared *S. pallida* (collected from the White Mountains, New Mexico) and *S. flava* (collected from near Portsmouth, New Hampshire). RNA was extracted using Trizol (Thermo-Fisher, Waltham, MA) and precipitated with isopropanol. Extracted RNA was treated with DNaseI, cDNA was generated using qScript cDNA Supermix (Quantabio, Beverly, MA). Absence of genomic DNA in cDNA preparations was verified by attempting to PCR-amplify fragments of the *Marf* gene from reactions lacking reverse transcriptase (Figure 1-figure supplement 3). PCR conditions and primers are detailed in (Lapoint et al., 2013). CDS plus 7–9 bp of untranslated sequence were amplified using High Fidelity Phusion Taq (New England BioLabs, NEB), 3% DMSO vol/vol, and the PCR primers (Supplementary file 4) with the following program: initial denaturing at 98°C during 30 sec; 35 cycles at 98°C during 10 sec, 58°C during 30 sec, 72°C during 45 sec, and extension at 72°C during 7 min. PCR fragments of the expected size were purified using the Qiaquick Gel purification kit protocol (Qiagen). An overhang was added to purified *Or67b* amplicons with Taq polymerase (Fermentas) and cloned using the pGEM-T Easy cloning kit protocol (Promega). Plasmids were extracted and purified using the GenElute plasmid miniprep kit (Sigma-Aldrich, St. Louis, MO). EcoRI and KpnI cut sites were introduced using restriction enzyme cut-site primers (Supplementary file 3) with 10 ng/µL diluted plasmids (as template) with 3% DMSO vol/vol and the following program: initial denaturing at 98°C during 30 sec; 35 repetitions of 98°C during 10 sec, 55°C during 50 sec; 72°C during 45 sec; and final extension at 72°C during 7 min. The pUAST attB plasmid (Bischof et al., 2007) and the four *S. pallida* and *S. flava Or67b* PCR amplicons with RE flanking sites were individually double-digested with KpnI and EcoRI high-fidelity enzyme in cut smart buffer for 3

hours, according to the manufacturer's protocol (NEB). Cut fragments were gel-purified using the Qiaquick Gel Cleanup Kit (Qiagen) and ligated in a 1:3 vector:insert molar ratio using T4 ligase (Promega). Ligations were transformed into JM109 cells. Some cells were preserved as glycerol stocks and a portion were used for injection into the y^1 w^{67c23}; P{CaryP}attP2 *D. melanogaster* line for generating *Sfla Or67b1, Sfla Or67b2, Sfla Or67b3, Spal Or67b* or into the y^1 w^{67c23}; P{CaryP}attP40 for *Sfla Or67b2* (BestGene Inc., Houston, Texas, USA). Transformants were selected from individually injected flies with eye color rescue phenotypes.

Single sensillum recordings (SSR)

All *Or67b* constructs were expressed in ab3A, except *Sfla Or67b2*, which was expressed in at1, since *Sfla* Or67b2 exhibited spontaneous activity only in at1 trichoid olfactory sensory neurons in transgenic *D. melanogaster* flies (Figure 2 – figure supplemental 4).

Fed, adult female flies were prepared for SSR as previously described (Hallem and Carlson, 2004). We identified the antennal sensilla housing the OSN/s of interest using an Olympus BX51WI upright microscope with 10x and 50x objectives (Olympus, UPlanFL N 10x, UPlanFL N 50x). Extracellular activity was recorded by inserting a tungsten electrode into the base of either ab3 or at1 sensilla. Signals were amplified 100 x (A-M systems, Differential AC Amplifier model 1700), digitized using a 16-bit analog-digital converter, filtered (low cut-off: 300 Hz, high cut off: 500 Hz), and analyzed off-line using WinEDR (v3.9.1; University of Strathclyde, Glasgow). A tube delivering a constant flow of charcoal-filtered air (16 ml/min, using a flowmeter; Gilmont instruments, USA) was placed near the fly's head, and the tip of the stimulation pipette (50 ml) was inserted into the constant air stream. The stimulation pipette contained a piece of filter paper loaded with 20 µl of odorant solution or the solvent control. One second pulse of clean air was delivered to the stimulus pipette using a membrane pump

operated by a Stimulus Controller CS 55 (Syntech, Germany). Ab3 sensilla were identified by using three standard diagnostic odorants (Gonzalez et al., 2016; all odorants were obtained from Sigma-Aldrich, US, purity > 95%): ethyl hexanoate (CAS # 123-66-0), ethyl acetate (CAS # 141-78-6) and 2-heptanone (CAS # 110-43-0). We were able to distinguish at1 sensilla from at2 and at3 because the former houses a single olfactory sensory neuron, whereas at2, at4 and at3 respectively house two, three and three neurons (Gonzalez et al., 2016).

The following odor sources (purchased from Berkeley Bowl in Berkeley, California, USA, unless otherwise mentioned; 20 µl of material were loaded on filter paper unless noted) were used: apple cider vinegar (40 µl, O Organics, USA), grated roots of the four Brassicaceae species *Wasabia japonica* (wasabi), *Armoracia rusticana* (horseradish), *Brassica rapa* (turnip), *Raphanus sativus* (daikon), and the Amaranthaceae species *Beta vulgaris* (beet). Brassicaceae species *Eruca vesicaria* (arugula), *Arabidopsis thaliana* and Solanaceae species *Solanum lycopersicum* (tomato) were grown from seeds at 23°C and 60% relative humidity under a 12-hours light: 12-hours dark cycle, and leaves from 3-8 weeks old plants were used for odor stimulation. The following *A. thaliana* genotypes were used: wild-type (Col-0), glucosinolate knockout (GKO) mutant in *myb28 myb29 cyp79b2 cyp79b3*, which has no detectable aliphatic and indolic glucosinolates or camalexin (Beekwilder et al., 2008; Sønderby et al., 2007; Zhao et al., 2002), and camalexin-deficient *phytoalexin deficient 3* (pad3) mutants that have wildtype levels of aliphatic and indolic glucosinolates, and are more appropriate controls for comparisons with GKO plants than Col-0 (Schuhegger et al., 2006). Roots and leaves were grated and homogenized to ~500 µl. All synthetic odorants were diluted in mineral oil (1:100 vol/vol) unless otherwise noted. The following odorants (all from Sigma-Aldrich, US, purity >95%) were diluted in dimethyl sulfoxide (DMSO): mandelonitrile (CAS # 532-28-5), iodoacetamide (CAS # 144-48-9), N-methylmaleimide (CAS # 930-88-1), N-hydroxysuccinimide (CAS # 6066-82-6), and benzyl thiocyanate

(CAS # 3012-37-1). 11-*cis* vaccenyl acetate (CAS # 6186-98-7) and 4MBITC (CAS # 4430-36-8) were diluted in ethanol. All chemicals used in this study are listed in supplementary file 6.

The "net number of spikes" were obtained by counting the number of spikes originating from the desired olfactory sensory neuron of interest during a 1-second window which started 0.2 seconds after the onset of stimulation, and subtracting from this number the background spiking activity (obtained by counting the number of spikes in a 1-second window prior to the onset of the odor stimulation). In all figures (unless otherwise noted) the net response to each odor or odorant stimulation was subtracted from the median net response to the solvent control.

In SSR experiments it is common to observe unspecific slight increases/decreases in spiking activity which are not considered meaningful, including small responses to solvent controls. Accordingly, it is not uncommon to find reports, including investigations using the empty neuron system of *D. melanogaster*, in which net spiking responses smaller than 10-25 spikes/second are not considered true odor-evoked responses (*e.g.* Olsson et al., 2006; Stensmyr et al., 2003). Therefore, we asked if the net responses to a given combination of Or and odorant compound are statistically significant ($p<0.05$) using one-sample signed Rank tests under the following null and alternative hypotheses:

Ho: net number of spikes >-10 and <10

Ha: net number of spikes <-10 or >10

Tuning curves and kurtosis values were generated and calculated in excel. Similarly, a matrix of median responses (control-subtracted) was produced and used for Principal Component Analysis in R statistical software. For generation of the heatmap in R, the median responses of OR-odorant pairs were normalized to the maximum median response (BITC at 1:100 vol/vol) for each Or across all odorants.

This normalization served to adjust for the potential differential responsiveness between the ab3A and at1 empty neuron systems (Kurtovic et al., 2007).

Behavioral tests

The olfactory responses of mated, fed adult female *D. melanogaster* (Canton-S), *S. pallida*, *S. flava*, and transgenic flies were tested using a custom-made dual-choice "Y-shaped" olfactometer (Reisenman et al., 2013; Figure 5-figure supplement 6). Flies were starved 24 hours for experiments shown in Fig. 5C. The "Y piece" of the olfactometer was a propylene connector, and the arms of the "Y" were each connected to a 1-ml syringe containing a piece of filter paper (6 x 50 mm) loaded with the odor or control stimuli. Charcoal-filtered air was delivered to each of the two stimulus syringes using silicon tubing at 250 ml/min; thus, at the base of the maze the air flow was approximately 500 ml/min. Two hours (in the case of *D. melanogaster* and *S. pallida*) or *ca*. 20 hours before tests (in the case of *S. flava*) insects were gently anesthetized under CO_2 and placed in groups of thee-four in open-top and mesh-bottom cylindrical release containers (20 mm long x 10 mm diameter) constructed using silicon tubing. The open top of the containers was capped with a piece of cotton soaked in distilled water (in the case of *D. melanogaster* and *S. pallida*) or with a piece of cotton soaked in 10% vol/vol aqueous honey solution (in the case of *S. flava*). Before tests, each release tube was placed on ice during 45-60 seconds to slow down insect activity; the cotton cap was then removed and the open-top of the tube was carefully slid into the open end of the Y maze. Thus, upon being released, insects could walk upwind towards the "decision point" (intersection of the short and long arms of the "Y") and turn towards either the odor-laden or the odorless arm of the maze. Although three-four insects were released at once (to increase experimental efficacy), only the first choice (and the time of the choice) was recorded; a choice was considered as such only if the insect walked past at least 10 mm into one of the arms, orienting upwind.

The test was discarded if two or more insects choose different arms of the maze within a five-second window of each other. Each test lasted a maximum of five minutes, and each group of insects was used only once. As much as possible, insects from the same cohort were tested in the same day with different odors/odorants; insects from the three different species were also, as much as possible, tested in the same day with a given odor/odorant. Similarly, experimental and control transgenic flies were tested in parallel as much as possible. Tests with each species/fly lines/stimulus were conducted in at least five different days with different progeny to compensate for possible day-to-day variations. In general, results from an individual test session with a given odor/odorant were discarded if insects did not make a choice in more than 50% of tests (this happened in less than 5-10% of experimental sessions except for *S. pallida*, which often had low activity levels). The position of the odor and odorless arms was switched every 1-2 tests to control for positional asymmetries; the mazes and odor sources were changed and replaced for clean/new ones every 4 tests or 10 minutes, whichever occurred first.

The odor/odorant was loaded in a piece of filter paper and inserted into the 1-ml syringe right before tests; control syringes had a piece of filter paper loaded with the mineral oil solvent (for odorant solutions) or water (in tests apple cider vinegar). Experimental and control filter papers were replaced by fresh ones every 4 tests or about 10-11 minutes, whichever came first. The odorants (20 µl of 1:100 vol/vol mineral oil solution unless noted) used in experiments were BITC (Sigma-Aldrich, CAS # 592-82-5, USA) and SBITC (Sigma-Aldrich, CAS # 15585-98-5, USA), and acetophenone (Sigma-Aldrich, CAS # 98-86-2, USA) in one experiment (Fig. 5C). We also used apple cider vinegar (40 µl, O Organics, USA; 40 µl of distilled water was a control stimulus in these tests). For tests of host-orientation, leaves from young arugula, *A. thaliana*, or tomato plants, grown in an insect and insecticide/pesticide free chamber or greenhouse, were excised just before tests and placed in 5-ml syringes connected to the Y-maze; control syringes had two pieces of wet tissue paper. In all cases the

Y-mazes, tubing and syringes were washed with 70% ethanol and allowed to air-dry before reusing. Experiments were conducted during the 2nd-5th hour of the insects' photophase at 24 ºC under white light (Feit electric, 100 Watts; in the case of *S. pallida* and *D. melanogaster*) or green light (Sunlite green, 100 Watts; in the case of experiments with *S. flava*). In all cases the total number of tests in which at least one insect chooses one or the other arm of the maze is indicated in the figures for each species/fly line/odor.

Because *S. flava* females lay eggs in leaves, generation of mutants using CRISPR-Cas9 has been not possible at the time of these experiments. Therefore, we conducted tests of "gain of function" in which *D. melanogaster* flies expressed *Sfla* Or67b3 in the basiconic ab3A "empty neuron", or in ab9b neuron (which express Or67b in *D. melanogaster*; Couto et al., 2005). Controls and experimental flies were tested with butyl-ITC 1:1,000 vol/vol. To investigate the role of the Or67b circuit in mediating olfactory attraction, we expressed UAS-Kir2.1, a genetically encoded inwardly rectifying potassium channel that by keeping cells hyperpolarized prevents neuronal excitation (Baines et al., 2001), under the control of Or67b-Gal4. Experimental and control fly lines were water-starved during 24 hs and tested with acetophenone 1:50,000 vol/vol. To verify that previously reported behavioral responses to acetophenone occurred under our experimental conditions, *D. melanogaster* flies (Canton-S) were also tested with various concentrations of this odorant.

For each odor/odorant, species, and fly line, the number of tests in which an insect made a choice for one or the other arm of the Y-maze (with the criteria described above) were tested against a 50% expected random distribution using two-tailed Binomial tests (Zar, 1999). Results were considered statistically significant if $p<0.05$. Because binomial tests require a large number of tests to evince relatively small differences in deviations from an expected proportion (for instance, the smallest sample size to detect a 10% deviation from the expected 50% random choice with $p<0.05$ is 78), we specifically

note whenever $p>0.05$ but <0.09 (*e.g.* with n=55, a proportion of odor choice=60% gives a p-value=0.088).

We also conducted odor-trapping experiments with Canton-S flies. In these experiments a group of 48 hours-starved flies (n=18-20) was released in a container that has an odor-baited trap (2000 ul of 1:50 vol/vol of BITC in mineral oil with triton-x added) and a control trap (2000 ul of mineral oil with triton-x added). After 48 hours the number of flies in each trap, as well as the number of flies not trapped, were counted. Tests with <60% of total capture were discarded. For each test we calculated a preference index as: (number of flies in odorant trap - number of flies in control trap/total number of flies captured). The number of flies captured in each trap was compared and Wilcoxon-matched pair test was performed.

Data Analysis and figure generation

All images and drawings are originals prepared by the authors. Figures were prepared via a combination of WinEDR (v3.9.1), R Studio (v1.2.1335), Microsoft Excel (2016), Adobe Illustrator (2019), ChemDraw (19.0), Python, and Geneious (10.0.9).

Acknowledgments

We are grateful to Drs. Chauda Sebastian, Dennis Mathew, John Carlson, Barry J. Dickson and Bloomington Drosophila Stock Center (NIH P40OD018537) for sharing *M2-MD*, *UAS-Dmel Or67b*, and *Or67d*^{Gal4}, and to Drs. Johannes Bischof and Konrad Basler for donation of the pUASTattB plasmid. C.E.R. thanks Dr. Kristin Scott for support and encouragement. T.M. thanks Dr. Makoto Hiroi for advice on SSR experiments. We thank Mr. George Wang for synteny analysis of *S. montana* and members of the Whiteman and Scott Laboratories for discussions and comments on the manuscript. This

work was supported by the Uehara Memorial Foundation (award number 201931028 to T.M.), the National Institute of General Medical Sciences of the National Institutes of Health (award number R35GM119816 to N.K.W.) and the National Science Foundation (award number IOS 1755188 and DEB 1601355 supporting B.G.-H.).

References:

Al-Anzi B, Tracey WD Jr, Benzer S. 2006. Response of *Drosophila* to wasabi is mediated by painless, the fly homolog of mammalian TRPA1/ANKTM1. *Curr Biol* **16**:1034–1040. doi:10.1016/j.cub.2006.04.002

Anderson P, Anton S. 2014. Experience-based modulation of behavioural responses to plant volatiles and other sensory cues in insect herbivores. *Plant Cell Environ* **37**:1826–1835.

Anderson P, Sadek MM, Larsson M, Hansson BS, Thöming G. 2013. Larval host plant experience modulates both mate finding and oviposition choice in a moth. *Anim Behav* **85**:1169–1175. doi:10.1016/j.anbehav.2013.03.002

Arenas OM, Zaharieva EE, Para A, Vásquez-Doorman C, Petersen CP, Gallio M. 2017. Activation of planarian TRPA1 by reactive oxygen species reveals a conserved mechanism for animal nociception. *Nat Neurosci* **20**:1686–1693. doi:10.1038/s41593-017-0005-0

Assis R, Bachtrog D. 2013. Neofunctionalization of young duplicate genes in *Drosophila*. *Proc Natl Acad Sci U S A* **110**:17409–17414. doi:10.1073/pnas.1313759110

Auer TO, Khallaf MA, Silbering AF, Zappia G, Ellis K, Álvarez-Ocaña R, Roman Arguello J, Hansson BS, Gregory S X, Caron SJC, Knaden M, Benton R. 2020. Olfactory receptor and circuit evolution

promote host specialization. *Nature* **579**(7799):402-408.doi:10.1038/s41586-020-2073-7

Aurand LW, Singleton JA, Bell TA, Etchells JL. 1966. Volatile Components in the Vapors of Natural and Distilled Vinegars. *J Food Sci* **31**:172–177. doi:10.1111/j.1365-2621.1966.tb00474.x

Baines RA, Uhler JP, Thompson A, Sweeney ST, Bate M. 2001. Altered electrical properties in *Drosophila* neurons developing without synaptic transmission. *J Neurosci* **21**:1523–1531.

Barron AB. 2001. The life and death of Hopkins' host-selection principle. *J Insect Behav* **14**:725–737.

Beekwilder J, Van Leeuwen W, Van Dam NM, Bertossi M, Grandi V, Mizzi L, Soloviev M, Szabados L, Molthoff JW, Schipper B, Others. 2008. The impact of the absence of aliphatic glucosinolates on insect herbivory in *Arabidopsis*. *PLoS One* **3**.

Benjamini Y, Hochberg Y. 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)*. doi:10.1111/j.2517-6161.1995.tb02031.x

Berenbaum MR, Zangerl AR. 2008. Facing the future of plant-insect interaction research: le retour à la "raison d'être." *Plant Physiol* **146**:804.

Bischof J, Maeda RK, Hediger M, Karch F, Basler K. 2007. An optimized transgenesis system for *Drosophila* using germ-line-specific φC31 integrases. *Proc Natl Acad Sci U S A* **104**:3312–3317. doi:10.1073/pnas.0611511104

Bouckaert R, Vaughan TG, Barido-Sottani J, Duchêne S, Fourment M, Gavryushkina A, Heled J, Jones G, Kühnert D, De Maio N, Matschiner M, Mendes FK, Müller NF, Ogilvie HA, du Plessis L, Popinga A, Rambaut A, Rasmussen D, Siveroni I, Suchard MA, Wu C-H, Xie D, Zhang C, Stadler T, Drummond AJ. 2019. BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLoS Comput Biol* **15**:e1006650. doi:10.1371/journal.pcbi.1006650

Butenandt A. 1959. Über den Sexual-Lockstoff des Seidenspinners Bombyx mori: Reindarstellung und

Konstitution. Verlag d. Zeitschr. f. Naturforschung.

Chin SG, Maguire SE, Huoviala P, Jefferis GSXE, Potter CJ. 2018. Olfactory Neurons and Brain Centers Directing Oviposition Decisions in *Drosophila. Cell Rep* **24**:1667–1678. doi:10.1016/j.celrep.2018.07.018

Chittenden HM. 1902. The American Fur Trade of the Far West: A History of the Pioneer Trading Posts and Early Fur Companies of the Missouri Valley and the Rocky Mountains and the Overland Commerce with Santa Fe. F.P. Harper.

Choo Y-M, Xu P, Hwang JK, Zeng F, Tan K, Bhagavathy G, Chauhan KR, Leal WS. 2018. Reverse chemical ecology approach for the identification of an oviposition attractant for *Culex quinquefasciatus*. *Proc Natl Acad Sci U S A* **115**:714–719. doi:10.1073/pnas.1718284115

Cousin FJ, Le Guellec R, Schlusselhuber M, Dalmasso M, Laplace J-M, Cretenet M. 2017. Microorganisms in Fermented Apple Beverages: Current Knowledge and Future Directions. *Microorganisms* **5**. doi:10.3390/microorganisms5030039

Couto A, Alenius M, Dickson BJ. 2005. Molecular, anatomical, and functional organization of the *Drosophila* olfactory system. *Curr Biol* **15**:1535–1547. doi:10.1016/j.cub.2005.07.034

Crowley-Gall A, Date P, Han C, Rhodes N, Andolfatto P, Layne JE, Rollmann SM. 2016. Population differences in olfaction accompany host shift in *Drosophila mojavensis*. *Proc Biol Sci* **283**. doi:10.1098/rspb.2016.1562

Date P, Dweck HKM, Stensmyr MC, Shann J, Hansson BS, Rollmann SM. 2013. Divergence in olfactory host plant preference in *D. mojavensis* in response to cactus host use. *PLoS One* **8**:e70027. doi:10.1371/journal.pone.0070027

Degen T, Dillmann C, Marion-Poll F, Turlings TCJ. 2004. High genetic variability of herbivore-induced volatile emission within a broad range of maize inbred lines. *Plant Physiol* **135**:1928–1938.

doi:10.1104/pp.104.039891

Dekker T, Ibba I, Siju KP, Stensmyr MC, Hansson BS. 2006a. Olfactory shifts parallel superspecialism for toxic fruit in *Drosophila melanogaster* sibling, *D. sechellia. Curr Biol* **16**:101–109. doi:10.1016/j.cub.2005.11.075

Demirel N, Cranshaw W. 2006. Relative attraction of color traps and plant extracts to the false chinch bug *Nysius raphanus* and its parasitoid, *Phasia occidentis*, on brassica crops in Colorado. *Phytoparasitica* **34**:197–203. doi:10.1007/BF02981320

Dobritsa AA, van der Goes van Naters W, Warr CG, Steinbrecht RA, Carlson JR. 2003. Integrating the molecular and cellular basis of odor coding in the *Drosophila* antenna. *Neuron* **37**:827–841. doi:10.1016/s0896-6273(03)00094-1

Dweck HKM, Ebrahim SAM, Kromann S, Bown D, Hillbur Y, Sachse S, Hansson BS, Stensmyr MC. 2013. Olfactory preference for egg laying on citrus substrates in *Drosophila*. *Curr Biol* **23**:2472–2480. doi:10.1016/j.cub.2013.10.047

Eckenrode CJ, Arn H. 1972. Trapping cabbage maggots with plant bait and allyl isothiocyanate. *J Econ Entomol* **65**:1343–1345. doi:10.1093/jee/65.5.1343

Edger PP, Heidel-Fischer HM, Bekaert M, Rota J, Glöckner G, Platts AE, Heckel DG, Der JP, Wafula EK, Tang M, Hofberger JA, Smithson A, Hall JC, Blanchette M, Bureau TE, Wright SI, dePamphilis CW, Eric Schranz M, Barker MS, Conant GC, Wahlberg N, Vogel H, Pires JC, Wheat CW. 2015. The butterfly plant arms-race escalated by gene and genome duplications. *Proc Natl Acad Sci U S A* **112**:8362–8366. doi:10.1073/pnas.1503926112

Fahey JW, Zalcmann AT, Talalay P. 2001. The chemical diversity and distribution of glucosinolates and isothiocyanates among plants. *Phytochemistry* **56**:5–51. doi:10.1016/s0031-9422(00)00316-2

Faucher CP, Hilker M, de Bruyne M. 2013. Interactions of carbon dioxide and food odours in

Drosophila: olfactory hedonics and sensory neuron properties. *PLoS One* **8**:e56361. doi:10.1371/journal.pone.0056361

Frankel G. 1959. The raison d'etre of secondary plant substances. *Science* **129**:1466–1470.

Galizia C, Sachse S. 2009. Odor Coding in Insects. *Frontiers in Neuroscience*. doi:10.1201/9781420071993-c2

Gloss AD, Nelson Dittrich AC, Lapoint RT, Goldman-Huertas B, Verster KI, Pelaez JL, Nelson ADL, Aguilar J, Armstrong E, Charboneau JLM, Groen SC, Hembry DH, Ochoa CJ, O'Connor TK, Prost S, Suzuki HC, Zaaijer S, Nabity PD, Whiteman NK. 2019. Evolution of herbivory remodels a *Drosophila* genome. *bioRxiv*. doi:10.1101/767160

Gloss AD, Vassão DG, Hailey AL, Nelson Dittrich AC, Schramm K, Reichelt M, Rast TJ, Weichsel A, Cravens MG, Gershenzon J, Montfort WR, Whiteman NK. 2014. Evolution in an ancient detoxification pathway is coupled with a transition to herbivory in the drosophilidae. *Mol Biol Evol* **31**:2441–2456. doi:10.1093/molbev/msu201

Goldman-Huertas B, Mitchell RF, Lapoint RT, Faucher CP, Hildebrand JG, Whiteman NK. 2015. Evolution of herbivory in Drosophilidae linked to loss of behaviors, antennal responses, odorant receptors, and ancestral diet. *Proc Natl Acad Sci U S A* **112**:3026–3031. doi:10.1073/pnas.1424656112

Gonzalez F, Witzgall P, Walker WB. 2016. Protocol for Heterologous Expression of Insect Odourant Receptors in *Drosophila*. *Front Ecol Evol* **4**:189. doi:10.3389/fevo.2016.00024

Gross LA, Baird GS, Hoffman RC, Baldridge KK, Tsien RY. 2000. The structure of the chromophore within DsRed, a red fluorescent protein from coral. *Proc Natl Acad Sci U S A* **97**:11990–11995. doi:10.1073/pnas.97.22.11990

Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New Algorithms and

Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Systematic Biology*. doi:10.1093/sysbio/syq010

Guo S, Kim J. 2007. Molecular Evolution of *Drosophila* Odorant Receptor Genes. *Molecular Biology and Evolution*. **24**(5):1198-207. doi:10.1093/molbev/msm038

Gupta P, Wright SE, Kim S-H, Srivastava SK. 2014. Phenethyl isothiocyanate: a comprehensive review of anti-cancer mechanisms. *Biochim Biophys Acta* **1846**:405–424. doi:10.1016/j.bbcan.2014.08.003

Hallem EA, Carlson JR. 2006. Coding of Odors by a Receptor Repertoire. *Cell*. **125**(1):143-60. doi:10.1016/j.cell.2006.01.050

Hallem EA, Carlson JR. 2004. The odor coding system of *Drosophila. Trends in Genetics*. **20**(9):453-9. doi:10.1016/j.tig.2004.06.015

Heidel-Fischer HM, Kirsch R, Reichelt M, Ahn S-J, Wielsch N, Baxter SW, Heckel DG, Vogel H, Kroymann J. 2019. An Insect Counteradaptation against Host Plant Defenses Evolved through Concerted Neofunctionalization. *Mol Biol Evol* **36**:930–941. doi:10.1093/molbev/msz019

He X, Zhang J. 2005. Rapid subfunctionalization accompanied by prolonged and substantial neofunctionalization in duplicate gene evolution. *Genetics* **169**:1157–1164. doi:10.1534/genetics.104.037051

Hinman A, Chuang H-H, Bautista DM, Julius D. 2006. TRP channel activation by reversible covalent modification. *Proc Natl Acad Sci U S A* **103**:19564–19568. doi:10.1073/pnas.0609598103

Hoare DJ, Humble J, Jin D, Gilding N, Petersen R, Cobb M, McCrohan C. 2011. Modeling peripheral olfactory coding in *Drosophila* larvae. *PLoS One* **6**:e22996. doi:10.1371/journal.pone.0022996

Jaenike J. 1990. Host specialization in phytophagous insects. *Annu Rev Ecol Syst* **21**:243–273. doi:10.1146/annurev.es.21.110190.001331

Jirovetz L, Smith D, Buchbauer G. 2002. Aroma compound analysis of *Eruca sativa (Brassicaceae)*

SPME headspace leaf samples using GC, GC-MS, and olfactometry. *J Agric Food Chem* **50**:4643–4646. doi:10.1021/jf020129n

Joseph RM, Carlson JR. 2015. *Drosophila* Chemoreceptors: A Molecular Interface Between the Chemical World and the Brain. *Trends in Genetics*. **31**(12). doi:10.1016/j.tig.2015.09.005

Kang K, Pulver SR, Panzano VC, Chang EC, Griffith LC, Theobald DL, Garrity PA. 2010. Analysis of *Drosophila* TRPA1 reveals an ancient origin for human chemical nociception. *Nature* **464**:597–600. doi:10.1038/nature08848

Katoh K. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*. doi:10.1093/nar/gkf436

Katoh T, Izumitani HF, Yamashita S, Watada M. 2017. Multiple origins of Hawaiian drosophilids: phylogeography of *Scaptomyza* hardy (Diptera: Drosophilidae). *Entomol Sci* **20**:33–44.

Keesey I, Zhang J, Depetris-Chauvin A, Obiero GF, Knaden M, Hansson BS. 2019. Evolution of a pest: towards the complete neuroethology of *Drosophila suzukii* and the subgenus *Sophophora*. *bioRxiv*. doi:https://doi.org/10.1101/717322

Kergunteuil A, Dugravot S, Mortreuil A, Le Ralec A, Cortesero AM. 2012. Selecting volatiles to protect brassicaceous crops against the cabbage root fly, Delia radicum. *Entomol Exp Appl* **144**:69–77. doi: 10.1111/j.1570-7458.2012.01257.x

Kim BY, Wang JR, Miller DE, Barmina O, Delaney E, Thompson A, Comeault AA, Peede D, D'Agostino ERR, Pelaez J, Aguilar JM, Haji D, Matsunaga T, Armstrong EE, Zych M, Ogawa Y, Stamenković-Radak M, Jelić M, Veselinović MS, Tanasković M, Erić P, Gao J-J, Katoh TK, Toda MJ, Watabe H, Watada M, Davis JS, Moyle LC, Manoli G, Bertolini E, Košťál V, Scott Hawley R, Takahashi A, Jones CD, Price DK, Whiteman N, Kopp A, Matute DR, Petrov DA. 2020. Highly contiguous assemblies of 101 drosophilid genomes. *bioRxiv*

doi:https://doi.org/10.1101/2020.12.14.422775

Kim H-Y, Gornsawun G, Shin I-S. 2015. Antibacterial activities of isothiocyanates (ITCs) extracted from horseradish *(Armoracia rusticana)* root in liquid and vapor phases against 5 dominant bacteria isolated from low-salt Jeotgal, a Korean salted and fermented seafood. *Food Sci Biotechnol* **24**:1405–1412. doi:10.1007/s10068-015-0180-2

Kreher SA, Kwon JY, Carlson JR. 2005. The molecular basis of odor coding in the *Drosophila* larva. *Neuron* **46**:445–456. doi:10.1016/j.neuron.2005.04.007

Kreher SA, Mathew D, Kim J, Carlson JR. 2008. Translation of sensory input into behavioral output via an olfactory system. *Neuron* **59**:110–124. doi:10.1016/j.neuron.2008.06.010

Kurtovic A, Widmer A, Dickson BJ. 2007. A single class of olfactory neurons mediates behavioural responses to a *Drosophila* sex pheromone. *Nature* **446**:542–546. doi:10.1038/nature05672

Lanfear R, Frandsen PB, Wright AM, Senfeld T, Calcott B. 2017. PartitionFinder 2: New Methods for Selecting Partitioned Models of Evolution for Molecular and Morphological Phylogenetic Analyses. *Mol Biol Evol* **34**:772–773. doi:10.1093/molbev/msw260

Lapoint RT, O'Grady PM, Whiteman NK. 2013. Diversification and dispersal of the Hawaiian Drosophilidae: the evolution of *Scaptomyza. Mol Phylogenet Evol* **69**:95–108. doi:10.1016/j.ympev.2013.04.032

Larter NK, Sun JS, Carlson JR. 2016. Organization and function of *Drosophila* odorant binding proteins. *Elife* **5**. doi:10.7554/eLife.20242

Lichtenstein EP, Strong FM, Morgan DG. 1962. Naturally Occurring Insecticides, Identification of 2-Phenylethylisothiocyanate As an Insecticide Occurring Naturally in the Edible Part of Turnips. *J Agric Food Chem* **10**:30–33. doi:10.1021/jf60119a009

Linz J, Baschwitz A, Strutz A, Dweck HKM, Sachse S, Hansson BS, Stensmyr MC. 2013. Host plant-

driven sensory specialization in *Drosophila erecta*. *Proc Biol Sci* **280**:20130626. doi:10.1098/rspb.2013.0626

Liu X-L, Zhang J, Yan Q, Miao C-L, Han W-K, Hou W, Yang K, Hansson BS, Peng Y-C, Guo J-M, Xu H, Wang C-Z, Dong S-L, Knaden M. 2020. The Molecular Basis of Host Selection in a Crucifer-Specialized Moth. *Curr Biol* **30**:4476–4482.e5. doi:10.1016/j.cub.2020.08.047

Maca J. 1972. Czechoslovak species of the genus *Scaptomyza* Hardy (Diptera, Drosophilidae) and their bionomics. *Acta entomologica Bohemoslovaca*.

Macpherson LJ, Dubin AE, Evans MJ, Marr F, Schultz PG, Cravatt BF, Patapoutian A. 2007. Noxious compounds activate TRPA1 ion channels through covalent modification of cysteines. *Nature* **445**:541–545. doi:10.1038/nature05544

Mansourian S, Enjin A, Jirle EV, Ramesh V, Rehermann G, Becher PG, Pool JE, Stensmyr MC. 2018. Wild African *Drosophila melanogaster* Are Seasonal Specialists on Marula Fruit. *Curr Biol* **28**:3960–3968.e3. doi:10.1016/j.cub.2018.10.033

McBride CS, Arguello JR, O'Meara BC. 2007. Five *Drosophila* genomes reveal nonneutral evolution and the signature of host specialization in the chemoreceptor superfamily. *Genetics* **177**:1395–1416. doi:10.1534/genetics.107.078683

Mitchell-Olds T. 2001. *Arabidopsis thaliana* and its wild relatives: a model system for ecology and evolution. *Trends Ecol Evol* **16**:693–700. doi:10.1016/S0169-5347(01)02291-1

Mithen R, Raybould AF, Giamoustaris A. 1995. Divergent selection for secondary metabolites between wild populations of *Brassica oleracea* and its implications for plant-herbivore interactions. *Heredity* **75**:472.

Miura N, Nakagawa T, Tatsuki S, Touhara K, Ishikawa Y. 2009. A male-specific odorant receptor conserved through the evolution of sex pheromones in Ostrinia moth species. *Int J Biol Sci* **5**:319–

330. doi:10.7150/ijbs.5.319

Montague SA, Mathew D, Carlson JR. 2011. Similar odorants elicit different behavioral and physiological responses, some supersustained. *J Neurosci* **31**:7891–7899. doi:10.1523/JNEUROSCI.6254-10.2011

Münch D, Galizia CG. 2016. DoOR 2.0--Comprehensive Mapping of *Drosophila melanogaster* Odorant Responses. *Sci Rep* **6**:21841. doi:10.1038/srep21841

Newton E, Bullock JM, Hodgson D. 2009. Bottom-up effects of glucosinolate variation on aphid colony dynamics in wild cabbage populations. *Ecol Entomol* **34**:614–623. doi:10.1111/j.1365-2311.2009.01111.x

Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Molecular Biology and Evolution*. doi:10.1093/molbev/msu300

Ohno S. 1970. The Enormous Diversity in Genome Sizes of Fish as a Reflection of Nature's Extensive Experiments with Gene Duplication. *Trans Am Fish Soc* **99**:120–130.

Olsson SB, Linn CE Jr, Roelofs WL. 2006. The chemosensory basis for behavioral divergence involved in sympatric host shifts. I. Characterizing olfactory receptor neuron classes responding to key host volatiles. *J Comp Physiol A Neuroethol Sens Neural Behav Physiol* **192**:279–288. doi:10.1007/s00359-005-0069-2

Pelaez JN, Gloss AD, Ray JF, Whiteman NK. 2020. Evolution and genetic basis of the plant-penetrating ovipositor, a key adaptation in the transition to herbivory within the Drosophilidae. *bioRxiv*. doi:10.1101/2020.05.07.083253

Piersanti S, Rebora M, Ederli L, Pasqualini S, Salerno G. 2020. Role of chemical cues in cabbage stink bug host plant selection. *J Insect Physiol* **120**:103994. doi:10.1016/j.jinsphys.2019.103994

Pivnick KA, Jarvis BJ, Slater GP. 1994. Identification of olfactory cues used in host-plant finding by diamondback moth,Plutella xylostella (Lepidoptera: Plutellidae). *J Chem Ecol* **20**:1407–1427. doi:10.1007/BF02059870

Pivnick KA, Reed DW, Millar JG, Underhill EW. 1991. Attraction of northern false chinch bug *Nysius niger* (Heteroptera: *Lygaeidae*) to mustard oils. *Journal of Chemical Ecology*. **17**, 931–941 doi:10.1007/bf01395600

Prieto-Godino LL, Rytz R, Bargeton B, Abuin L, Roman Arguello J, Peraro MD, Benton R. 2016. Olfactory receptor pseudo-pseudogenes. *Nature*. **539**: 93–97 doi:10.1038/nature19824

Prieto-Godino LL, Rytz R, Cruchet S, Bargeton B, Abuin L, Silbering AF, Ruta V, Dal Peraro M, Benton R. 2017. Evolution of Acid-Sensing Olfactory Circuits in Drosophilids. *Neuron* **93**:661–676.e6. doi:10.1016/j.neuron.2016.12.024

Rasmann S, Köllner TG, Degenhardt J, Hiltpold I, Toepfer S, Kuhlmann U, Gershenzon J, Turlings TCJ. 2005. Recruitment of entomopathogenic nematodes by insect-damaged maize roots. *Nature*. **434**: 732–737. doi:10.1038/nature03451

Reisenman CE, Lee Y, Gregory T, Guerenstein PG. 2013. Effects of starvation on the olfactory responses of the blood-sucking bug *Rhodnius prolixus. J Insect Physiol* **59**:717–721. doi:10.1016/j.jinsphys.2013.04.003

Renwick JAA, Haribal M, Gouinguené S, Städler E. 2006. Isothiocyanates stimulating oviposition by the diamondback moth, *Plutella xylostella. J Chem Ecol* **32**:755–766. doi:10.1007/s10886-006-9036-9

Robertson HM, Warr CG, Carlson JR. 2003. Molecular evolution of the insect chemoreceptor gene superfamily in *Drosophila melanogaster. Proc Natl Acad Sci U S A* **100 Suppl 2**:14537–14542. doi:10.1073/pnas.2335847100

Rohloff J, Bones AM. 2005. Volatile profiling of *Arabidopsis thaliana*--Putative olfactory compounds in plant communication. *Phytochemistry* **66**:1941–1955.

Ronderos DS, Lin C-C, Potter CJ, Smith DP. 2014. Farnesol-detecting olfactory neurons in *Drosophila. J Neurosci* **34**:3959–3968. doi:10.1523/JNEUROSCI.4582-13.2014

Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol* **61**:539–542. doi:10.1093/sysbio/sys029

Sakurai T, Nakagawa T, Mitsuno H, Mori H, Endo Y, Tanoue S, Yasukochi Y, Touhara K, Nishioka T. 2004. Identification and functional characterization of a sex pheromone receptor in the silkmoth *Bombyx mori*. *Proc Natl Acad Sci U S A* **101**:16653–16658. doi:10.1073/pnas.0407596101

Schuhegger R, Nafisi M, Mansourova M, Petersen BL, Olsen CE, Svatos A, Halkier BA, Glawischnig E. 2006. CYP71B15 (PAD3) catalyzes the final step in camalexin biosynthesis. *Plant Physiol* **141**:1248–1254. doi:10.1104/pp.106.082024

Semmelhack JL, Wang JW. 2009. Select *Drosophila* glomeruli mediate innate olfactory attraction and aversion. *Nature* **459**:218–223. doi:10.1038/nature07983

Sønderby IE, Hansen BG, Bjarnholt N, Ticconi C, Halkier BA, Kliebenstein DJ. 2007. A systems biology approach identifies a R2R3 MYB gene subfamily with distinct and overlapping functions in regulation of aliphatic glucosinolates. *PLoS One* **2**:e1322. doi:10.1371/journal.pone.0001322

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. doi:10.1093/bioinformatics/btu033

Stensmyr MC, Dweck HKM, Farhan A, Ibba I, Strutz A, Mukunda L, Linz J, Grabe V, Steck K, Lavista-Llanos S, Wicher D, Sachse S, Knaden M, Becher PG, Seki Y, Hansson BS. 2012. A conserved dedicated olfactory circuit for detecting harmful microbes in *Drosophila. Cell* **151**:1345–

1357. doi:10.1016/j.cell.2012.09.046

Stensmyr MC, Giordano E, Balloi A, Angioy A-M, Hansson BS. 2003. Novel natural ligands for *Drosophila* olfactory receptor neurones. *J Exp Biol* **206**:715–724. doi:10.1242/jeb.00143

Strutz A, Soelter J, Baschwitz A, Farhan A, Grabe V, Rybak J, Knaden M, Schmuker M, Hansson BS, Sachse S. 2014. Decoding odor quality and intensity in the *Drosophila* brain. *Elife* **3**:e04147. doi:10.7554/eLife.04147

Sultana T, Savage GP, McNeil DL, Porter NG, Clark B. 2003. Comparison of flavour compounds in wasabi and horseradish. *J Food Agric Environ* **1**:117–121.

Swofford DL. 2002. PAUP: phylogenetic analysis using parsimony, version 4.0 b10.

Thurmond J, Goodman JL, Strelets VB, Attrill H, Gramates LS, Marygold SJ, Matthews BB, Millburn G, Antonazzo G, Trovisco V, Kaufman TC, Calvi BR, FlyBase Consortium. 2019. FlyBase 2.0: the next generation. *Nucleic Acids Res* **47**:D759–D765. doi:10.1093/nar/gky1003

Wei JJ, Fu T, Yang T, Liu Y, Wang GR. 2015. A TRPA1 channel that senses thermal stimulus and irritating chemicals in *Helicoverpa armigera*. *Insect Molecular Biology*. **24**(4):412-21.doi:10.1111/imb.12168

Wheat CW, Vogel H, Wittstock U, Braby MF, Underwood D, Mitchell-Olds T. 2007. The genetic basis of a plant–insect coevolutionary key innovation. *Proc Natl Acad Sci U S A* **104**:20427–20431. doi:10.1073/pnas.0706229104

Whiteman NK, Groen SC, Chevasco D, Bear A, Beckwith N, Gregory TR, Denoux C, Mammarella N, Ausubel FM, Pierce NE. 2011. Mining the plant--herbivore interface with a leafmining *Drosophila* of *Arabidopsis*. *Mol Ecol* **20**:995–1014.

Willmore B, Tolhurst DJ. 2001. Characterizing the sparseness of neural codes. *Network* **12**:255–270.

Winde I, Wittstock U. 2011. Insect herbivore counteradaptations to the plant glucosinolate–myrosinase

system. *Phytochemistry* **72**:1566–1575. doi:10.1016/j.phytochem.2011.01.016

Xue Y-L, Han H-T, Liu C-J, Gao Q, Li J-H, Zhang J-H, Li D-J, Liu C-Q. 2020. Multivariate analyses of the volatile components in fresh and dried turnip (*Brassica rapa L.*) chips via HS-SPME-GC-MS. *J Food Sci Technol* **57**:3390–3399. doi:10.1007/s13197-020-04372-y

Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* **24**:1586–1591. doi:10.1093/molbev/msm088

Zar JH. 1999. Biostatistical Analysis. Prentice Hall.

Zhao Y, Hull AK, Gupta NR, Goss KA, Alonso J, Ecker JR, Normanly J, Chory J, Celenza JL. 2002. Trp-dependent auxin biosynthesis in *Arabidopsis*: involvement of cytochrome P450s CYP79B2 and CYP79B3. *Genes Dev* **16**:3100–3112. doi:10.1101/gad.1035402

Zhao Z, McBride CS. 2020. Correction to: Evolution of olfactory circuits in insects. *J Comp Physiol A Neuroethol Sens Neural Behav Physiol* **206**:663. doi:10.1007/s00359-020-01409-7

Supplementary file 1 Phylogenetic analysis

Phylogenetic dataset summary, sequence accession numbers, genome sequence coordinates and phylogenetic model parameters and results.

Supplementary file 2 Molecular evolution analyses

Selected parameters and results from branch d_N/d_S tests of all *S. flava* Ors cds sequences and the expanded *Or67b* dataset.

Supplementary file 3 Or67b synteny

The three spreadsheets indicate pGOIs for the three *Or67b* orthologs identified in *S. flava.* For each sheet, we have the following columns: “Position from GOI in *S. flava”* (e.g. -1 if one gene upstream of *Or67b*, +2 if two genes downstream of *or67b*); “*S. flava* Annotation ID”; “*D. grimshawi* homolog”

(identified via blastn searches); "*D. grimshawi* homolog scaffold"; "*S. pallida* homolog locations", which include scaffold and coordinates; "*D. melanogaster* homolog", and "*D. melanogaster* homolog coordinates". "NA" is written in the cell if homologs were not found after executing blastn, blastx and tblastx searches.

Supplementary file 4 PCR primers

Nucleotides in lower case are either in untranslated sequence (CDS amplification) or are restriction enzyme cut sites (RE cut site addition). CDS amplification primers were used to amplify full *Or67b* CDS sequence from cDNA. Primers labeled "RE cut site addition" were used to engineer restriction enzyme cut-sites via PCR mutagenesis in order to ligate *Or67b* CDS sequences into the pUASTattB plasmid. All sequences are listed in a 5' to 3' orientation.

Supplementary file 5 PC coordinates

Supplementary file 6 pGOI sequences

Supplementary file 7 list of chemicals