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

Authors: #Teruyuki Matsunaga¹*, #Carolina E. Reisenman¹², #Benjamin Goldman-Huertas¹³, Philipp Brand⁴, Kevin Miao¹, Hiromu C. Suzuki¹, Kirsten I. Verster¹, Santiago R. Ramírez⁴, Noah K. Whiteman¹²*

#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: teru.matsu0208@berkeley.edu (T.M.) and whiteman@berkeley.edu (N.K.W.)

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

ABSTRACT:
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 when herbivores co-opt them as cues for host-plant finding, but the evolutionary mechanisms underlying this phenomenon are poorly understood. We focused on Scaptomyza flava, an herbivorous drosophilid specialized on isothiocyanate (ITC)-producing (Brassicaceae) plants, and identified Or67b paralogs that were triplicated as mustard-specific herbivory evolved. Using in vivo 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 detected different ITC subsets. Consistent with this, *S. flava* was attracted to ITCs, as was *Drosophila melanogaster* expressing *S. flava* Or67b3 in the homologous Or67b olfactory circuit. ITCs were likely co-opted as olfactory attractants through gene duplication and functional specialization (neofunctionalization and subfunctionalization) in *S. flava*, a recently-derived herbivore.

**INTRODUCTION**

Many plant molecules used in food, agriculture and medicine first evolved as defenses against natural enemies (Frankel 1959). Among the most familiar are reactive electrophiles that produce pain when eaten, including diallyl disulfide and thiosulfinates in Alliaceae (*e.g.*, garlic), α, β-unsaturated aldehydes in Lauraceae (*e.g.*, cinnamon), and isothiocyanates (ITCs) in Brassicales (*e.g.*, arugula, radish, and wasabi). These electrophiles activate the ‘wasabi taste receptor’ TrpA1, which is conserved in flies and humans (Kang et al. 2010). Although ITCs are potent natural insecticides aversive to most insects, including *D. melanogaster* (Lichtenstein et al. 1962), some insect species are specialized on ITC-bearing Brassicales. This insect-plant interaction has been important in advancing the field of co-evolution (Berenbaum and Zangerl 2008).

Brassicales (or mustards) 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 ancestrally aversive electrophiles are co-opted as olfactory cues for host-plant finding (Kergunteuil et al. 2012). Identifying the mechanisms that allow insects to exploit plants containing toxic chemicals can help us to understand the evolution of herbivorous species, 90% of which are specialized on a limited set of host plants (Bernays and Graham 1988).

*Scaptomyza* represents a compelling genus to investigate how specialist herbivores evolved to co-opt plant defenses as host finding cues. Phylogenetically nested within *Drosophila, Scaptomyza* is a sister group to the Hawaiian *Drosophila* radiation (Lapoint et al. 2013), and contains many herbivorous species with varying degrees of specialization on Brassicales and Caryophyllaceae (Gloss et al. 2019) (Figure 1A). Adult females create feeding punctures on the undersides of living leaves using sclerotized dentate ovipositors (Pelaez et al. 2020), and larvae mine the mesophyll of leaf tissue (Whiteman et al. 2011), an unusual drosophilid life history. *S. flava* specializes on Brassicales (Mitchell-Olids 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, genomic and genetic tools of both Arabidopsis and Drosophila can be utilized to dissect both sides of the plant-herbivore equation. Herbivory evolved ca. 10-15 million years ago in Scaptomyza (Figure 1A) and so it provides an unusually useful context to understand the evolutionary and functional mechanisms underlying chemosensory specialization, because the major herbivorous insect radiations (e.g. Lepidoptera; Kawahara et al. 2019, Phytophaga; McKenna et al. 2019) are far more ancient in origin. S. flava lost three olfactory receptor (Or) genes encoding canonical yeast-associated receptors compared to microbe-feeding Drosophila (Goldman-Huertas et al. 2015). However, the loss of function does not explain how herbivorous Scaptomyza species evolved attraction to Brassicales plants.

Gain of function through gene duplication and subsequent divergence plays an important role in evolutionary innovation and is often 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 Brassicales-specialist pierid butterflies, enzymes that divert hydrolysis of aliphatic glucosinolates away from ITC production to less toxic nitriles evolved in tandem, underpinning their diversification and highlighting the importance of gene family evolution in this co-evolutionary arms race (Wheat et al. 2007). While our understanding of detoxification in plant-herbivore systems has grown rapidly over the past decade (Heckel 2018), the evolutionary mechanisms underlying the chemosensory basis of hostplant orientation and finding are less clear. Here we investigated how the evolution of olfactory receptor genes in the S. flava lineage contributes to attraction to host-plant volatiles, including ITCs.

We took a genes-first approach to study olfactory host-plant specialization in S. flava. Three chemoreceptor protein families (olfactory receptors -Ors-, ionotropic receptors -Irs-, and nociceptive receptors -Trps-) 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 odors important for feeding, reproduction and survival, including aversive chemicals (e.g. Stensmyr et al. 2012), pheromones (e.g. Butenandt 1959; Kurtovic et al. 2007; Sakurai et al. 2004), and host-, oviposition-, and food-related attractive compounds (Dekker et al. 2006a; Dweck et al. 2013; Linz et al. 2013). Accordingly, we scanned the genome sequence of S. flava to identify rapidly evolving Ors and found a lineage-specific triplication of the olfactory receptor gene Or67b. We named these paralogs Or67b1, Or67b2, and Or67b3. The coding sequences of each S. flava paralog exhibit signatures of positive selection. In contrast, Or67b is present as a single copy under strong evolutionary constraint.
across the Drosophilidae (Goldman-Huertas et al. 2015; Guo and Kim 2007; Robertson et al. 2003),
including the microbe-feeding species *S. pallida* and *D. melanogaster*. Our *in vivo* functional
characterizations of all Or67b proteins from *S. flava*, *S. pallida* and *D. melanogaster* show that the three
*S. flava* Or67b paralogs, but not the Or67b orthologs from the microbe-feeding relatives, responded
selectively and sensitively to volatile ITCs. Among these species only *S. flava* is attracted to single ITC
compounds. *Sfla* Or67b3 can confer odor-oriented responses towards ITCs when this Or is expressed in
the *D. melanogaster* homologous Or67b olfactory circuit. In summary, our results suggest a relatively
simple sensory mechanism by which mustard-specialist herbivores evolved olfactory attraction towards
ITCs, via gene duplication followed by neofunctionalization and subfunctionalization of a specific clade
of otherwise conserved Ors.

RESULTS

Phylogenetic analysis and antennal transcriptome identify *S. flava* Or67b paralogs as candidates
mediating the detection of ITCs

To search for candidate ITC-detecting 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* - and *S. flava*-specific gene losses and gains (Figure Supplement 1), as well as those
that were rapidly evolving at the protein level, regardless of duplication history. The Or topology was
largely congruent with previous *Drosophila* Or gene trees (Goldman-Huertas et al. 2015), except for
some deeper nodes with low bootstrap support. We fit foreground/background branch models, as
implemented in PAML, to detect *S. flava* Ors whose coding sequences were evolving under different
selection regimes compared to the *Drosophila* background (Yang 2007). Out of seventy-five *S. flava*
branches tested, seven branch models, related to *Or63a*, *Or67b* and *Or98a* in *D. melanogaster*, inferred
a foreground rate larger than one, consistent with a high rate of fixation of nonsynonymous mutations
during positive selection. Albeit the known functions and expression patterns of individual Ors in
*Scaptomyza* species cannot be directly extrapolated to those of *Scaptomyza* spp homologs, we noted that
*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 have not been characterized functionally. In
contrast, *Or67b* modulates oviposition behavior in adult *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 single branches on the *Or67b* phylogeny, we found significantly elevated $d_{N}/d_{S}$ exclusively in *S. flava* paralogs and *D. mojavensis* *Or67b* (Figure 1B). Furthermore, synteny analysis confirmed that homologous regions upstream and downstream of the two non-syntenic *Or67b* copies in *S. flava* are present in both *S. pallida* and *D. melanogaster*. Thus, the absence of these paralogs in *S. pallida* and *D. melanogaster* is likely not a genome assembly or sequencing artefact but rather, an actual absence (Figure Supplement 2A, B, Supplementary file 1, 2).

Additionally, we analyzed the antennal transcriptomes of *S. flava* and *S. pallida*. This confirmed expression of each annotated *Or67b* gene copy in the two species. Relative expression of *Or67b* paralogs was higher in *S. flava* antennae than in *S. pallida* in all replicates compared to other Ors. This was not the case for *Or98a* (Figure Supplement 3, Supplementary file 3), which further supports our choice of *Or67b* as a candidate for functional studies.

In summary, *Or67b* is uniquely triplicated in *S. flava* among the species analyzed, the coding sequences of these paralogs show signatures of positive selection, and the three paralogs are expressed at higher levels in *S. flava* antennae compared to the orthologous copy in the outgroup.

*S. flava* *Or67bs** respond specifically to mustard plant odors

Next, we investigated whether *Sfla* *Or67bs* acquired novel ligand-binding sensitivity towards odorants characteristic of mustard host plants. To study the odor-response profile of Or67b across the three species, we heterologously expressed their Or67b proteins in *D. melanogaster* antennal trichoid (at1) or antennal basiconic 3A (ab3A) olfactory sensory neurons (OSNs) lacking endogenous Ors (Dobritsa et al., 2003; Kurtovic et al. 2007). We then recorded electrophysiological responses to an array of mustard and non-mustard plant volatile organic compounds (VOCs), other odors and odorants known to be activators of *D. melanogaster* Or67b (Figure 2-3 and Figure Supplement 3). We first expressed each Or67b transgene in ab3A OSNs because Or67b is expressed in basiconic sensilla in *D. melanogaster*. Ab3A neurons expressing *Sfla Or67b1, Sfla Or67b3, Spal Or67b,* and *Dmel Or67b* showed spontaneous bursts of action potentials, as described for other Ors expressed in these neurons (Dobritsa et al. 2003), whereas neurons expressing *Sfla Or67b2* showed spontaneous activity only when we expressed this protein in at1 OSNs (Figure 2A and Figure Supplement 5A). Thus, we used the “at1 empty neuron” system for expression of *Sfla Or67b2*, and the “ab3A empty neuron” system for all other
A

S. graminum
S. pallida
S. hsui
D. grimshawi
D. hydei
D. mojavensis
D. virilis
D. willistoni
D. persimilis
D. pseudoobscura
D. ananassae
D. erecta
D. yakuba
D. melanogaster
D. sechellia
D. simulans

Hosts
Fresh Brassicaceae leaves
Fresh Caryophyllaceae leaves
decaying plant tissues

B

S. flava copy 2 (Sfla Or67b2)
S. flava copy 1 (Sfla Or67b1)
S. flava copy 3 (Sfla Or67b3)
S. graminum
S. pallida (Spal Or67b)
S. hsui
D. grimshawi
D. mojavensis
D. hydei
D. virilis
D. persimilis
D. pseudoobscura
D. ananassae
D. simulans
D. sechellia
D. erecta
D. melanogaster (Dmel Or67b)
Or67b proteins. Note that comparisons of odor-evoked response intensity between OSNs expressing Sfla Or67b2 and those of OSNs expressing the other Or67b proteins should be interpreted cautiously (but see Supplemental Figure 5), as the response magnitude is smaller for Ors expressed in the at1 empty neuron (e.g. Syed et al., 2010).

To test which odors specifically activate Sfla Or67b paralogs, we measured the responses of all Or67b copies to stimulation with apple cider vinegar, a potent D. melanogaster attractant (Faucher et al. 2013), crushed arugula leaves (Eruca vesicaria, a mustard), and crushed tomato leaves (Solanum lycopersicum), a non-mustard plant that releases large quantities of VOCs (Schauer et al. 2005). OSNs expressing Dmel Or67b and Spal Or67b, but not those expressing any of the Sfla Or67b paralogs, responded strongly to stimulation with apple cider vinegar (one-sample signed rank tests, p<0.05; Figure 2), indicating that the Sfla Or67b paralogs may have lost the ancestral odorant sensitivity. VOCs from the non-host tomato activated both Dmel Or67b and Spal Or67b (p<0.05), but not any of the Sfla Or67b paralogs (p>0.05; Figure 2C). In contrast, all Sfla Or67b proteins responded to arugula leaf volatiles (p<0.05), suggesting that Sfla Or67bs respond specifically to VOCs characteristic of the hosts S. flava attacks.

In contrast to tomato plants, Brassicales plants, including arugula, produce glucosinolates, some of which are hydrolyzed into ITCs upon tissue damage (Fahey et al. 2001). To test if responses of Sfla Or67bs to mustard leaf odors are mediated by ITCs and/or by other mustard VOCs, we used grated leaves of wild-type four to six weeks old A. thaliana (Col-0) plants and two loss of function A. thaliana mutants in the Col-0 background. One of these mutants (myb28 myb29 cyp79b2 cyp79b3; Glucosinolate Knock Out) is deficient in the production of ITC-precursors derived from aliphatic and indolic glucosinolates as well as camalexin (Beekwilder et al. 2008; Glazebrook and Ausubel 1994), and therefore does not release ITCs upon wounding (Zhao et al. 2002; Zhao et al. 2002; Schuhegger et al. 2006; Sønderby et al. 2007; Beekwilder et al. 2008), whereas the control mutant line phytoalexin deficient 3 (PAD3) produces these glucosinolates but not camalexin (Glazebrook and Ausubel 1994; Schuhegger et al. 2006).

Stimulation with wild-type leaves of Col-0 and PAD3, both of which produce ITCs, produced responses from OSNs expressing Sfla Or67b1, b2 and b3 (one-sample signed rank tests, p>0.05), whereas stimulation with GKO leaves did not (p>0.05; Figure 2C). Furthermore, the responses of all S. flava Or67b paralogs to VOCs from Col-0 and PAD3 were different from the responses to VOCs from...
GKO plants (Kruskal-Wallis ANOVA followed by Tukey-tests, p<0.05). Thus, ITCs, but not other *A. thaliana* leaf volatiles in our experiments, likely activated the three *S. flava* Or67b copies. In contrast, OSNs expressing *Dmel* Or67b or *Spal* Or67b showed similar responses to all three *A. thaliana* genotypes (Figure 2C, Kruskal-Wallis ANOVAs, p>0.05), in agreement with the reported responses of this Or to green leaf volatiles (GLVs) in *D. melanogaster* (Münch and Galizia 2016). Because GKO plants differed from both Col-0 and *PAD3* plants only in their capacity to produce aliphatic and indolic glucosinolates and their hydrolysis products, including ITCs (Beekwilder et al. 2008; Glazebrook and Ausubel 1994), these results demonstrate that *Sfla* Or67b paralogs are selectively activated by these signature Brassicales VOCs.

Because mustard plant roots release a variety of VOCs (including ITCs) but not GLVs (Xue et al. 2020; Sultana et al. 2003; Ishii et al., 1989), we also used these sources as stimuli to further explore natural, ITC-rich volatile blends on Or67b responses. We prepared root homogenates of four mustard plant species, including wasabi (*Eutrema japonicum*; Sultana et al. 2003), horseradish (*Armoracia rusticana*; Sultana et al. 2003), turnip (*Brassica rapa*; Xue et al. 2020), daikon (*Raphanus sativus*; Ishii et al. 1989), and beet (*Beta vulgaris*; Richardson 2013), a non-mustard root (control) vegetable species. OSNs expressing *D. melanogaster* or *S. pallida* Or67b paralogs did not respond to any root stimuli (one-sample signed rank tests, p>0.05, Figure 2C), while OSNs expressing any of the three *S. flava* Or67b paralogs responded to wasabi root volatiles (p<0.05 in all three cases). Finally, *Sfla* Or67b3 additionally responded to horseradish and turnip (p<0.05 in both cases). OSNs expressing *Sfla* Or67b3 showed small (but significant) responses to mustard plant (vegetative or root) VOCs (<40 spikes/second, median). Given the species diversity within Brassicales, it is possible that the natural activator of *Sfla* Or67b3 was not included in this study.

**S. flava** Or67b paralogs have different ITC selectivity

We next compared the odor-tuning profiles of Or67b copies from all species using a panel of 42 individual odorants (all tested at 1:100 vol/vol). This included ITCs, nitriles, GLVs, and odorants that activate *Dmel* Or67b, including ketones, esters, and alcohols (Münch and Galizia 2016; Figure 3). Testing this varied array of VOCs covered a diverse range of known secondary plant chemicals and other appetitive odors (Liu et al. 2020; Allmann et al. 2013).
We first verified the previously reported *Dmel* Or67b odor-response profile (Münch and Galizia 2016). Most of the odorants categorized as *Dmel* activators and GLVs evoked significant responses (13/17 odorants, one-sample signed rank tests, p<0.05; Figure 3). *Spal* Or67b responded to a smaller number of odorants within those categories (7/17 odorants; p<0.05). Strikingly, only one of the compounds that elicited responses from *Dmel*Or67b and *Spal* Or67b (trans-2-hexen-1-ol) evoked responses from *Sfla* Or67b1 (median=22 net spikes/second; p<0.05), but none of the 17 odorants categorized as *Dmel* activators and GLVs evoked responses from *Sfla* Or67b2 or *Sfla* Or67b3 (Figure 3; p>0.05). While various ITCs evoked responses from each of the *S. flava* paralogs (6-7/10 ITCs depending on the paralog; p<0.05), none evoked responses from *Dmel* Or67b or *Spal* Or67b (p>0.05; Figure 3A, B). Tuning curves (Figure 3C) of all Or67b proteins show that *Sfla* Or67bs have response profiles distinct from *Dmel* Or67b and *Spal* Or67b, whereas non-ITC compounds evoked the strongest responses (acetophenone and *cis*-3-hexenyl-butyrate, respectively, center of the distribution, yellow bars; Figure 3C). All *Sfla* Or67b paralogs had the strongest responses to ITC compounds (center of the distribution, green bars; Figure 3C). *Sfla* Or67b1 had the narrowest odorant-receptive range, responding most strongly to a smaller subset of ITC compounds tested, indicated by the high kurtosis value and sharp peak of the tuning curve. In contrast, *Sfla* Or67b3 had the widest odorant-receptive range, responding similarly to many ITCs, indicated by the low kurtosis and broad peak of the tuning curve (Figure 3C). In sum, *Dmel* Or67b and *Spal* Or67b do not respond to ITCs and have similar odor-response profiles, while each *Sfla* Or67b paralog is differentially responsive to ITCs found in diverse mustard plant species (Fahey et al. 2001).

We confirmed that the differences in ITC selectivity across Or67b proteins were not an artifact resulting from the use of two different empty neurons. We expressed *Spal* Or67b, *Sfla* Or67b1, and *Sfla* Or67b3 in at1 empty OSNs and compared their responses to acetophenone and ITCs (which serve to discriminate Or67bs, Figure 3) with those of ab3A OSNs expressing each of these proteins. In both cases, *Spal* Or67b responded only to acetophenone, whereas *Sfla* Or67b1 and *Sfla* Or67b3 selectively responded to ITCs (Figure 3 and Figure Supplement 5). Although the response intensity was smaller for at1 OSNs than for ab3A OSNs (Figure Supplement 5B), the response profiles of *Sfla* paralogs were similar for both OSN types. For example, BITC activated all paralogs but SBITC only activated Or67b3 (Figure 3 and Figure Supplement 5).

We next tested whether the presence of the ITC functional group (–N=C=S) is necessary for evoking responses from OSNs expressing *Sfla* proteins. We stimulated OSNs with two linkage isomers,
BITC (which bears the ITC functional group) and benzyl thiocyanate BTC (which bears the thiocyanate functional group, -S≡C-N; Figure 3B, C). Stimulation with BITC evoked robust responses in OSNs expressing any of the three paralogs (one-sample signed rank tests, p<0.05), while stimulation with BTC had no effect (p>0.05; Figure 3B, C). This differential activation pattern is therefore likely due to the ITC functional group, as these compounds are not only linkage isomers but also have similar volatilities.

We also stimulated OSNs expressing Sfla paralogs with serial dilutions of eight ITCs that evoked the strongest responses at 1:100 vol/vol (Figure 3B) to further assess paralog ITC-selectivity, which may allow detection and discrimination of different mustard plant accessions or species. Because the magnitude of the responses may be reduced when an Or is expressed in at1 OSNs (versus responses when expressed in ab3A OSNs, Figure Supplemental 5, Syed et al. 2010), we included both net and normalized median responses (Figure 4A-B). In general, OSN responses increased with increasing odorant concentration (Figure 4A-B). All paralogs have similar dose-responses to BITC stimulation, while other ITCs elicited Or-specific dose-responses (Figure 4). For example, Sfla Or67b1 responded sensitively to BITC and 4MBIT down to 10⁻⁴ vol/vol, and Sfla Or67b3 was the only paralog that responded to PITC at all concentrations (Fig. 4B). We also used principal component analysis to analyze the responses of OSNs expressing each of the Or67b paralogs in odor space. Most ITC responses clustered in an odorant-concentration dependent manner: responses separated best at 1:100 vol/vol, except for BITC, 4MBITC and 3MPITC, which separated best at 1:1,000 vol/vol (Figure Supplement 6, Supplementary file 4).

A population of S. flava antennal OSNs are responsive to BITC

Because Sfla Or67b paralogs responded sensitively to diverse ITCs when expressed in the heterologous empty neuron systems of D. melanogaster (Figure 4), we next investigated whether ITC-sensitive Ors are present in the antennae of S. flava by recording responses of OSNs in antennal basiconic-like (n=36) and trichoid-like sensilla (n=36) to stimulation with BITC 1:1,000 vol/vol (Figure Supplement 7). We used BITC at this low, biologically-relevant concentration (Jacobsson et al. 2004; Rohloff and Bones 2005; Gruber et al. 2009), because it evoked responses from all S. flava paralogs (Figure 4B; middle panel). Nearly one-fifth (19%) of recorded OSNs in basiconic-like and in trichoid-like sensilla showed medium to strong responses to stimulation with BITC (median: 130 and 166 net spikes/second, range: 74-252 and 65-200 net spikes/second respectively in each sensilla type; Figure Supplement 7A, B). These ITC-sensitive OSNs are located proximally in basiconic-like sensilla, and
A

**linear ITC**

**organosulfur ITCs**

**unsaturated ITC**

**branched ITCs**

**aromatic ITCs**

B

| Odor          | Or67b1 | Or67b2 | Or67b3 |
|---------------|--------|--------|--------|
| **linear ITC**| BITC   | *      | *      |
| **organosulfur ITCs**| 4MBITC | *      | *      |
| **unsaturated ITC**| 3MPITC | *      | *      |
| **branched ITCs**| AITC   | *      | *      |
| **aromatic ITCs**| IBITC  | *      | *      |
| benzyl ITC    | SBITC  | *      | *      |
| PITC          | *      | *      | *      |

Normalized response

odor concentration (vol./vol.)

Control-subtracted spikes/sec

odor concentration (vol./vol.)
The finding that *S. flava* has BITC-sensitive OSNs in at least two different morphological types of antennal olfactory sensilla is consistent with our finding that *Sfla* Or67b2 is functional only when expressed in the at1 empty neuron system, while *Sfla* Or67b1 and *Sfla* Or67b3 are functional in both at1 and ab3 basiconic sensilla of *D. melanogaster* (Figures 2, 3 and Figure Supplement 5). Future work is needed to confirm both expression of Or67b paralogs in *S. flava* ITC-sensitive sensilla and necessity of that expression for ITC detection.

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

The Or67b paralogs of *S. flava* – but not those of its non-herbivore relatives – responded selectively to ITCs. Thus, we hypothesized that *S. flava* evolved attraction to these odorants, like most other members of the mustard-feeding guild of herbivorous insects. We addressed this using a dual-choice olfactory assay (based on Reisenman et al. 2013; Figure Supplement 8) in which flies are allowed to choose between an odor-laden and an odorless arm of a “y-maze” olfactometer (e.g. Auer et al. 2020). We first tested whether flies are differentially attracted to mustard and non-mustard leaf VOCs under our experimental conditions.

In the y-maze, *S. flava* was attracted to arugula and *A. thaliana* Col-0 leaf VOCs (two-tailed Binomial tests, p<0.05), while *S. pallida* was not (p>0.05; Figure 5 and Figure Supplement 9). Leaf VOCs from the non-host plant tomato did not attract or repel *S. flava* (p>0.05), but attracted *S. pallida* (p<0.01; Figure 5), which has been reported from rotting tomato (Maca 1972). *S. flava* tested with leaf VOCs from *A. thaliana* PAD3 and ITC-lacking GKO mutants similarly preferred VOCs from either of these genotypes over clean air (Figure Supplement 9). This is not surprising because GKO plants release many VOCs other than ITCs that may mediate long-distance olfactory attraction (Glazebrook and Ausubel 1994; Sønderby et al. 2007; Beekwilder et al. 2008; Zhao et al. 2002). As expected (Faucher et al. 2013), *D. melanogaster* was strongly attracted to apple cider vinegar odors (p<0.05; Figure 5). Both *S. flava* and *S. pallida*, in contrast, distributed at random between the apple cider vinegar odor-laden and the odorless arm of the maze (p>0.05; Figure 5). In summary, *S. flava* is specifically attracted to mustard plant odors, but not to non-host plant odors or those that attract *D. melanogaster*, including acetic acid and ester-, carbonyl-, and hydroxyl-containing compounds (Aurand et al. 1966; Cousin et al. 2017).

We next investigated whether ITCs alone can mediate olfactory attraction in *S. flava*. 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). *S.
BITC

SBITC

Apple cider vinegar

Aruagula

Tomato

100 1000 5050
% to control % to odor % to control % to odor % to control % to odor

Dmel

Spal

Sfla

p=0.054

Dmel Spal Sfla

(4) (27) ****

(20) (23)

(14) (17)

(3) (25) *****

(20) (23)

(6) (20) **

(20) (20)

(20) (20)

(20) (20)

100 1000 5050
% to control % to odor % to control % to odor % to control % to odor

Arugula
tomato

(18) (18)

(17) (14)

(6) (20)

(25) (5)

(20) (26)

(25) ****

(24) (7)

(16) (23)

(7) (23)

(16) (23)

(16) (23)

(16) (23)

(16) (23)

(16) (23)

(16) (23)

(16) (23)

(16) (23)

(16) (23)
flava was attracted to both BITC and SBITC (p<0.05; Figure 5B). Interestingly, S. pallida strongly avoided BITC (p<0.005; Figure 5), which must occur via an Or67b-independent olfactory pathway because Spal Or67b does not respond to any of the ITCs tested, even at high concentrations (Figure 3). We also conducted dual-choice trapping experiments with D. melanogaster widely used for microbe-feeding Drosophila spp. (e.g. Prieto-Godino et al. 2017; Figure Supplement 10A). As previously reported (e.g. Auer et al., 2020), flies were captured in traps baited with apple cider vinegar over solvent control (Wilcoxon-matched pairs tests, p<0.005, n=9; Figure Supplement 10A), while flies distributed at random between the control and the BITC-baited test traps (p>0.9, n=24; Figure Supplement 10A). Thus, BITC does not attract D. melanogaster in a short-distance (trap) olfactory assay either. We could not use this trap assay with S. flava, because neither trap captured flies, likely due to the particular biology of this fly species, females of which only attack and oviposit on the underside of leaves. It is possible that using intact leaves as an oviposition substrate has selected for differences in search behavior. Overall, single ITCs compounds that evoke strong responses from Sfla Or67b paralogs expressed in D. melanogaster and from S. flava antennal OSNs (Figures 3-4; Figure Supplement 7) can mediate olfactory orientation in S. flava, but do not attract (and can even repulse) its microbe-feeding relatives.

Expression of Sfla Or67b in the homologous olfactory circuit of D. melanogaster confers behavioral responses to ITCs

Because Sfla Or67b paralogs selectively respond to ITCs, we tested if ectopic expression of these receptors in the homologous olfactory circuit of the microbe-feeder species is sufficient to mediate behavioral responses to these compounds. To test this we first used the Or22a olfactory circuit because it mediates detection of ecologically relevant chemicals in diverse drosophilids (Dekker et al. 2006a; Linz et al. 2013; Mansourian et al. 2018). We focused on Sfla Or67b3, as this paralog has the broadest ITC receptive range (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 was under the control of Gal4 in the ab3A neuron. We used BITC at 1:1,000 vol/vol because Sfla Or67b3 responds reliably to this odorant concentration (Figure 4).

While genetic control lines carrying Gal4 or UAS transgenes showed no preference for either arm of the maze (one-sample signed rank tests, p>0.05, Figure 6A), flies expressing Sfla Or67b3, but not Dmel Or67b, preferred the BITC-bearing arm (p<0.05, Figure 6A). Importantly, the genetic parental
**A**

| Or22a-Gal4  | UAS-Dmel Or67b | UAS-Sfla Or67b3 |
|-------------|----------------|-----------------|
| +           | -              | -              |
| -           | +              | -              |
| -           | -              | +              |
| +           | +              | -              |
| +           | -              | +              |

100 1000 5050 % to control

(19) (18) (20)(16) (27) (28)

**B**

| Or67b-Gal4  | UAS-Dmel Or67b | UAS-Sfla Or67b3 |
|-------------|----------------|-----------------|
| +           | -              | -              |
| +           | +              | -              |
| +           | -              | +              |

100 1000 5050 % to control

(19) (23) (21) (19)(14) (26)

**C**

| Or67b-Gal4  | UAS-Kir2.1 |
|-------------|------------|
| +           | -          |
| -           | +          |
| +           | +          |

p=0.052

100 1000 5050 % to acetophenone

(17) (29) (14) (25) (27) (15)
control flies showed normal odor responses towards apple cider vinegar, a *D. melanogaster* attractant, in the dual-choice trap assay (Figure Supplement 10B). We next tested whether *Sfla* Or67b3 can confer behavioral responses to BITC when expressed in *D. melanogaster* Or67b-OSNs. In this experiment, flies did not lack expression of the endogenous Or67b. *Gal4* control flies and flies expressing *Dmel* Or67b under the control of *Gal4* distributed at random between the two arms of the y-maze (p>0.05, Figure 6B), while flies expressing *Sfla* Or67b3 preferred the ITC-bearing arm of the maze (p<0.05; Figure 6B, gray bar). The attraction to BITC in these transgenic flies could be due to selective activation of OSNs expressing *Sfla* Or67b3, alone or concurrent with activation of other unknown attractive olfactory pathways by BITC moieties other than the ITC functional group.

Or67b-mediated host attraction has been proposed for other drosophilid species besides *S. flava*, including the distantly related *D. mojavensis* (Crowley-Gall et al. 2016). However, even in *D. melanogaster* it remains unclear if the Or67b circuit mediates olfactory attraction to host odors. Addressing this question could facilitate interpretation of the results of our gain of function experiments in which *Sfla* Or67b3 was expressed in the native *Dmel* Or67b circuit (Figure 6B). Our electrophysiological recordings showed that *Dmel* Or67b was activated by esters, alcohol, and ketones (Figure 3), which are released from fruits (Maldonado-Celis et al. 2019). Of these chemicals, we used acetophenone, because this compound is a naturally-occurring odorant released from guava pulp (Idstein and Scherier 1985) that strongly activates *Dmel* Or67b (Figure 3; Münch and Galizia 2016).

Importantly, acetophenone attracts *D. melanogaster* at low concentrations, but repels these flies at high concentrations (Strutz et al. 2014). We then tested if *Dmel* Or67b mediates olfactory attraction to host odors by silencing *Dmel* Or67b OSNs using UAS-Kir2.1, an inwardly-rectifying potassium channel that suppresses neuronal activity (Baines et al. 2001) under the control of Or67b-Gal4. Thus, we expect these flies to no longer be attracted to low concentrations of acetophenone.

Our experimental conditions faithfully reproduced previously reported results: wild-type *D. melanogaster* (Canton-S) was attracted and repelled by low and high concentrations of acetophenone, respectively (Figure Supplement 11). Flies with Or67b OSNs silenced were repelled by low concentration of acetophenone (Figure 6C; p<0.05), whereas *Gal4* and UAS genetic controls tended to be attracted to the same concentrations of this odorant compound (63% in both cases; binomial tests, p=0.052 and 0.077; Figure 6C). Altogether, these results suggest that the *D. melanogaster* Or67b olfactory circuit is necessary for mediating attraction toward low concentrations of the Or67b cognate ligand acetophenone.
Thus, as was proposed for *D. mojavensis* (Crowley-Gall et al. 2016), the Or67b circuit is a reasonable candidate for mediating olfactory attraction to host odors in drosophilids widely, including *S. flava*. Importantly, it remains to be investigated whether *S. flava* Or67b is necessary for host finding, which can be tested by developing knockout mutants in the future. Finally, it is possible that the *Dmel* Or67b circuit does not reflect the ancestral character state for *S. flava* or *D. mojavensis*, given that these lineages may have split over 60 million years ago (Figure 1A).

**DISCUSSION**

Most herbivorous insect species specialize on a narrow range of host plant species that each synthesize similar secondary chemicals (Jaenike 1990). While these toxins serve to defend plants against herbivory, ancestrally aversive molecules are co-opted as attractants by herbivores specializing on a new host plant lineage. We investigated the evolution of attraction to toxic host-plants using *S. flava* as a model. A candidate Or lineage (*Or67b*) was triplicated in a recent ancestor of *S. flava* and experienced rapid protein evolution, resulting in three divergent paralogs (*Sfla Or67b1-b3*; Figure 1) expressed in the antennae (Figure Supplement 3). Each *Sfla Or67b* paralog specifically responded to stimulation with mustard-plant odors and volatile ITCs in heterologous expression systems (Figures 2, 3, and Supplemental Figure 5), with each paralog responding to a specific subset of ITCs (Figure 4). In contrast, *S. pallida* and *D. melanogaster* Or67b orthologs did not respond to ITCs but showed strong responses to stimulation with apple cider vinegar and a broad range of aldehydes, alcohols and ketones (Figures 2, 3 and Supplemental Figure 5), consistent with their microbe-feeding niche (Galizia and Sachse 2009). Recordings from *S. flava* antennal sensilla revealed OSNs sensitive to ITCs (Figure Supplement 7). We found that *S. flava*, but not *S. pallida* or *D. melanogaster*, is attracted to volatile ITC compounds (Figure 5). Ectopic expression of *S. flava* Or67b3 in the *D. melanogaster* homologous Or67b olfactory circuit conferred odor-oriented behavioral responses to ITCs (Figure 6B). Finally, suppression of activity in Or67b OSNs in *D. melanogaster* decreased preference to its Or cognate ligand (Figure 6C). These results suggest that the ancestral Or67b olfactory circuit likely mediates olfactory attraction. Thus, gene duplication followed by specialization is a mechanism by which specialist herbivores evolve Ors that may confer olfactory attraction towards ancestrally aversive chemical compounds.
Evolutionary path of olfactory receptor specialization

The *S. flava* Or67b triplication reported here raises several molecular evolutionary questions, including how the expansion contributed to the generation of novel gene functions. In this regard, different scenarios have been proposed (Ohno 1970; Heidel-Fischer et al. 2019): (A) neofunctionalization, when one of the duplicated genes (paralogs) acquires a new function after accumulating *de novo* mutations, while the other copy retains ancestral function; (B) subfunctionalization *sensu stricto*, where mutations accumulate in both copies leading to partitioning of ancestral function; and (C) specialization, when subfunctionalization and neofunctionalization evolve simultaneously, yielding gene copies that are functionally different from one another and the ancestral copy (He and Zhang 2005; Assis and Bachtrog 2013).

Rapidly evolving *Sfla* Or67b paralogs strongly and selectively respond to ITCs, while the more conserved orthologous proteins from both a close and a more distant microbe-feeding relative did not respond to ITCs at all (Figures 3, 4 and Figure Supplement 5). *Sfla* Or67b paralogs thus evolved entirely new ligand-binding affinities (to ITCS) in the *S. flava* lineage at least since divergence from the common ancestor with *S. pallida* Or67b. This is consistent with a neofunctionalization event that pre-dates the duplications of *Sfla* Or67b. Further, each *Sfla* Or67b paralog selectively responds to different subsets of ITCs across a range of odorant concentrations, which comports with subfunctionalization. We therefore conclude that our data is most consistent with a specialization scenario to explain the evolution and function of *Sfla* Or67b paralogs.

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

Across Drosophilidae, several orthologous chemoreceptors respond in a species-specific manner to ecologically relevant ligands, suggesting that adaptative evolution has shaped their ligand-binding capabilities. Or22a is a good example of a hot-spot for sensory evolution in *Drosophila* because shifts in ligand-binding are associated with different odor-preference behaviors (Mansourian et al. 2018; Auer et al. 2020). Or22a and other highly variable receptors such as Ir75a/b have typically evolved new specificities towards different odorants within one chemical class (Prieto-Godino et al. 2017; Figure 7). However, this is not the case for Or67b, which responds to alcohols, aldehydes and ketones in non-herbivorous *D. melanogaster* and *S. pallida*, but in *S. flava*, to volatile ITCs, an entirely different chemical class (Figures 3 and 4). By comparing the responses of BITC and its linkage isomer BTC, we found that the presence of the ITC functional group (–N=C=S) is key for activation of these olfactory
A

1. ethyl hexanoate
2. methyl hexanoate
3. 3-methyl-2-butenyl acetate
4. isobutyl acetate
5. acetic acid
6. butyric acid
7. propionic acid
8. hexanoic acid

(C) Sfla Or67b1
Sfla Or67b2
Sfla Or67b3
Spal Or67b

B

1. ethyl hexanoate
2. methyl hexanoate
3. 3-methyl-2-butenyl acetate
4. isobutyl acetate
5. acetic acid
6. butyric acid
7. propionic acid
8. hexanoic acid

9. acetophenone
10. cis-3-hexenyl butyrate
11. AITC
12. 4MBITC
13. SBITC

ketones, esters, alcohols

ITCs

Downloaded from https://academic.oup.com/mbe/advance-article/doi/10.1093/molbev/msab362/6486429 by guest on 29 December 2021
receptors, highlighting a functional shift in ligand binding specificity (Figure 3B, C). The striking
difference in odor selectivity among Or67b orthologs may result from the evolution of herbivory in
Scaptomyza or subsequent specialization on mustard plants.

An olfactory receptor likely mediating attraction to mustard host plants

Through stimulation with VOCs from leaves of A. thaliana genotypes with and without ITCs, S.
flava has at least two different olfactory pathways for mustard-plant attraction: ITC-dependent and ITC-
independent (Figure 5 and Figure Supplement 9). This is consistent with the fact that plant VOCs are
complex bouquets of diverse, widely distributed and lineage-specific molecules that are variably capable
of releasing attraction behaviors (Bruce and Pickett 2011). Our results do not preclude that other
chemoreceptors and OSNs also contribute to mediate olfactory attraction to mustard plant VOCs and
ITC compounds in S. flava. The necessity of Or67bs for orientation towards ITCs requires generating
loss of function mutants and testing whether flies can still orient towards these odorants. Because S.
flava females lay eggs in the leaf mesophyll and cannot be reared on available media, generation of
Or67b mutants using CRISPR-Cas9 is not yet feasible. Newer technologies such as ReMOT Control
may enable such experiments (Chaverra-Rodriguez et al. 2018). It also remains to be explored if
evolution also sculpted neural wiring to modulate responses to ITC compounds (Auer et al. 2020; Zhao
and McBride 2020).

ITC taste detection in vertebrates and insects leads to aversive behaviors mediated by the contact
chemoreceptor TrpA1 (Hinman et al. 2006; Macpherson et al. 2007; Kang et al. 2010). However,
volatile ITCs are widely used to trap pests of Brassicales (Kergunteuil et al. 2012) and the antennae of
some of these insects, including S. flava (Goldman-Huertas et al. 2015; Figure Supplement 7), respond
to volatile ITCs (Liu et al. 2020; Piersanti et al. 2020; Renwick et al. 2006). Similarly, two Ors in the
mustard specialist diamondback moth Plutella xylostella respond selectively to ITCs (Liu et al. 2020),
including 3MPITC and PITC, which strongly activate S. flava Or67b paralogs. Thus, these ITCs are
likely important for host-orientation in distantly related mustard specialists.

Facing highly inducible host plant defenses in vegetative tissues is a distinct challenge for leaf-
mining Scaptomyza, compared to other herbivorous drosophilids like D. sechellia and D. suzukii, which
feed on ripe fruit. Although glucosinolates and jasmonate-inducible defenses reduce growth rates of S.
flava larvae (Whiteman et al., 2011; Whiteman et al. 2012), these flies have evolved highly efficient
glutathione S-transferases through gene duplication events (Gloss et al. 2014), some of which detoxify ITCs more efficiently in vitro than those from any known animal (Gloss et al. 2019). Yet, here we show that volatile ITCs at biologically relevant concentrations are actually attractive to S. flava adult females. The attraction to ITCs by S. flava may in part be explained by the Or67b triplication, which is only known from Brassicales-specialist Scaptomyza species (Kim et al. 2021) and not S. graminum, which attacks leaves of Caryophyllaceae (Figure 1A).

Although paradoxical, these results are entirely consistent with Fraenkel’s (Frankel 1959) overarching hypothesis that plant toxins evolve to dissuade and punish herbivores, but specialists eventually overcome and then co-opt the toxins as faithful signposts of their host plants. Our study adds to the growing body of evidence that gene duplication and subsequent evolution have played an important role in co-evolution (sensu lato) between Brassicales plants and the diverse guild of herbivorous insects that use them as hosts. More generally, chemosensory specialization can result from relatively simple genetic modifications in the peripheral nervous system that changes olfactory receptor tuning, contributing to major niche shifts in insects.

MATERIALS AND METHODS

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 species of Drosophila and four Scaptomyza spp. Scaptomyza sequences were accessed from the genomes in Gloss et al. (2019) and Kim et al. (2020) 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 Supplementary file 5. Five independent runs of BEAST v2.6.2 (Bouckaert et al. 2019) each with 25 million generations were run logging after every 2,500th generation to infer the chronogram with 10% burn-in. Phylogenies were also inferred in RAxML v8.2.12 (Stamatakis 2014) with the
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 Gloss et al. 2019 and Kim et al. 2020 including the non-leaf-mining species S. hsui (subgenus Hemiscaptomyza and a microbe-feeder) and S. graminum (subgenus Scaptomyza and a leaf-miner on Caryophyllaceae). 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 1,000 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 1,000 bootstraps. Model parameters, accession numbers and likelihood scores available in Supplementary file 5.

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 potentially consistent with episodic positive selection. 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 dN/dS rate (ratio of non-synonymous to synonymous substitution rate) for every unique phylogeny (75 tests in total). After focusing on Or67b and selecting this as a candidate (as explained above), 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 before (34 tests total, Supplementary file 6). P-values were adjusted for multiple comparisons using the false-discovery rate (FDR) method (Benjamini and Hochberg 1995). Branch test results and model parameters are included in Supplementary file 6.

**Fly husbandry and lines**

_D. melanogaster_ (wild-type Canton-S and transgenic lines) were reared using cornmeal media, yeast, and agar medium prepared by UC-Berkeley core facilities. 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_ Col-0 plants 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.
Flies used for heterologous expression of Ors were of the following genotypes: for ab3A “empty neuron” flies, \textit{Or}22\textit{ab}\textit{Gal4::3xP3-DsRed} was used. This “M2-MD” line was generated by a CRISPR-Cas9 mediated deletion of \textit{Or22a/b} and a knock-in of \textit{Gal4} and \textit{DsRed} (Gross et al. 2000) by homology directed repair (HDR); in these flies \textit{Gal4} is not functional. Therefore, \textit{Or22ab\textit{Gal4::3xP3-DsRed}; Or22a-/Gal4/UAS-Or67b} flies were used for experiments. For at1 “empty neuron” flies, \textit{Or67d\textit{Gal4}} (Kurtovic et al. 2007) was used. The \textit{Or67b-Gal4} fly line (BDSC# 9996) was crossed with \textit{UAS-Or67b} lines or \textit{UAS-Kir2.1} (Baines et al. 2001) for behavioral assays.

\textbf{Scaptomyza \textit{Or67b} gene cloning, \textit{UAS} line generation}

The \textit{UAS-Or67b} transgene lines were constructed as follows: RNA was extracted from 20–25 days post-emergence adults of both sexes from laboratory-reared \textit{S. pallida} (collected from the White Mountains, New Mexico, USA) and \textit{S. flava} (collected from near Portsmouth, New Hampshire, USA). RNA was extracted using Trizol (Thermo-Fisher, Waltham, MA, USA) and precipitated with isopropanol. Extracted RNA was treated with DNaseI, cDNA was generated using qScript cDNA Supermix (Quantabio, Beverly, MA, USA). PCR conditions and primers are detailed in (Lapoint et al. 2013). EcoRI and KpnI cut sites were introduced using restriction enzyme cut-site primers (Supplementary file 7) with 10 ng/µL diluted plasmids (as template) with 3% DMSO vol/vol. The pUAST attB plasmid (Bischof et al. 2007) and the four \textit{S. pallida} and \textit{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 \textit{y\textsuperscript{1} w\textsuperscript{67c23}; P\{}\text{CaryP}\{}\text{attP2 D. melanogaster} line for generating \textit{Sfla Or67b1}, \textit{Sfla Or67b2}, \textit{Sfla Or67b3}, \textit{Spal Or67b} or into the \textit{y\textsuperscript{1} w\textsuperscript{67c23}; P\{}\text{CaryP}\{}\text{attP40 for Sfla Or67b2} (BestGene Inc., Houston, Texas, USA). Transformants were selected from individually injected flies with compound eye color rescue phenotypes.

\textbf{Single sensillum recordings (SSR)}

Adult female flies that were fed and 3-15 days old were prepared for SSR as previously described (Hallem and Carlson 2004). We recorded the responses of 6-10 sensilla obtained from 2-4
individuals for each experiment/odorant. Signals were amplified 100x (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 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; Figure Supplement 4). We were able to distinguish at1 sensilla from all other at sensilla because at1 is the only antennal trichoid sensilla known to house a single OSN (Gonzalez et al. 2016).

The following odor sources were used (purchased from Berkeley Bowl in Berkeley, California, USA, 20 µl of material were loaded on filter paper unless noted): grated roots of the four Brassicaceae species; *W. japonica* (wasabi), *A. rusticana* (horseradish), *B. rapa* (turnip), *R. sativus* (daikon), and the Amaranthaceae species *B. vulgaris* (beet). Approximately 10 g of roots were grated immediately before experiments. We also used pure apple cider vinegar as a stimulus (40 µl, O Organics, USA). Brassicaceae species *E. vesicaria* (arugula), *A. thaliana*, and *S. lycopersicum* (tomato) were grown from seeds at 23°C and 60% relative humidity under a 12-hour light: 12-hour 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), aliphatic and indolic Glucosinolate Knock Out (GKO) mutant in *myb28 myb29 cyp79b2 cyp79b3*, which has no detectable aliphatic and indolic glucosinolates or camalexin (Zhao et al. 2002; Sønderby et al. 2007; Beekwilder et al. 2008), and camalexin-deficient phytoalexin deficient 3 (PAD3) mutants that have wild-type levels of aliphatic and indolic glucosinolates but no camalexin. Therefore, PAD3 plants are more appropriate controls for comparisons with GKO plants than Col-0 plants for our study (Schuhegger et al. 2006). Three-four leaves were excised from plants and homogenized with a grater immediately before tests; the homogenate was replaced every 30 min since OSN responses were stable at least during this time window. 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 8.

The “net number of spikes” was obtained by counting the number of spikes originating from the OSN of interest during a 1-second window which started 0.2 seconds after the onset of stimulation, and subtracting from this count 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 used. Some odorants were dissolved in solvents other than mineral oil and so the respective control was used (Supplementary file 9).

In SSR experiments it is common to observe unspecific, slight changes in spiking activity which are not considered biologically meaningful, including small responses to solvent controls. It is common to find reports, including investigations using the empty neuron system of D. melanogaster (Hallem and Carlson 2006), in which net spiking responses smaller than 10-25 spikes/second are not considered true odor-evoked responses (e.g. Stensmyr et al. 2003; Olsson et al. 2006). Therefore, we asked whether the net responses of a given Or to a given stimulus are statistically significant (p<0.05) using one-sample signed rank tests under the following null and alternative hypotheses: Ho: net number of spikes $\geq$-10 and $<10$, and Ha: net number of spikes $<10$ or $>10$. The responses of a given paralog to various stimuli were also compared using Kruskal-Wallis ANOVAs (followed by post-hoc tests if significant). Although results were considered significant only if p<0.05, we indicated cases in which p-values were slightly larger (0.05<p<0.08). We report these values because “a result that does not meet the p<0.05 threshold should not be considered meaningless when in fact provides…at least preliminary evidence that requires further attention” (Makin and de Xivry 2019), especially with small sample sizes due to the nature of the experiment, as in this report.

Tuning curves and kurtosis values were generated and calculated in Microsoft Excel (2016). Similarly, a matrix of median responses (control-subtracted) was produced and used for PCA in R statistical software. For generation of the heatmap in R version 1.4.1717 (Figure 4B), the median responses of OR-odorant pairs were normalized to the maximum median response (which was evoked by BITC at 1:100 vol/vol) for each Or.
Y-maze binary choice assay

The olfactory responses of mated, fed adult female D. melanogaster, S. pallida, and S. flava were tested using a dual-choice “Y-shaped” olfactometer (Reisenman et al. 2013; Figure Supplement 8). 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. The odorous stimulus was loaded on a piece of filter paper and inserted into the 1 ml syringe immediately prior to each test; control syringes had a piece of filter paper loaded with the mineral oil solvent (for odorant solutions) or water (in tests apple cider vinegar). The odorants (20 µl of 1:100 or 1:1,000 vol/vol mineral oil solution, depending on the experiment) used in experiments were BITC (CAS # 592-82-5) and SBITC (CAS # 15585-98-5), and acetophenone (CAS # 98-86-2) at various concentrations. We also used apple cider vinegar (40 µl, O Organics, USA). For tests of host-orientation, leaves from four to six weeks old E. sativa, A. thaliana, or S. lycopersicum plants, grown in an insect and insecticide/pesticide free chamber or greenhouse, were excised and broken 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 reuse.

Flies were starved 24 hours for these experiments. Flies were anesthetized under CO2 two hours (in the case of D. melanogaster and S. pallida) or ca. 20 hours before tests (in the case of S. flava), and placed in groups of 3-4 individuals in open-top and mesh-bottom cylindrical release containers (20 mm long x 10 mm diameter) constructed using silicon tubing. The open tops of the containers were capped with tissue paper soaked in distilled water (in the case of D. melanogaster and S. pallida) or with a piece of tissue soaked with 10% vol/vol aqueous honey solution (in the case of S. flava). Before tests, each release tube was placed on ice for 45-60 seconds to slow down insect activity; the cap was then removed and the open-top of the tube was slid into the open end of the Y-maze. Only the first choice (and the time of the choice) was recorded among each group of 3-4 flies per trial. A choice was considered as such only if the fly walked at least 10 mm into one of the arms, orienting upwind. Each test lasted a maximum of five minutes. Experiments were conducted during the 2-5 hour of the insects’ photophase at 24 ºC under white light (Feit electric, 100 Watts; in the case of S. pallida and D. melanogaster). Green light (Sunlite green, 100 Watts) was used in the case of experiments with S. flava, as our observations suggest that this species is more active under this condition. The test was discarded if two or more insects chose different arms of the maze within a five-second window of each other. Results from a test
session with a given species/fly line were discarded if insects did not make a choice in at least 50% of
tests. This happened in less than 5-10% of experimental sessions except for *S. pallida*, which often had
low activity levels. Insects from the same cohort were tested in the same day with different stimuli as
much as possible. In the case of experiments using transgenic flies, we conducted experiments with the
progeny of at least 4-5 independent crosses; genetic control and test flies were tested in parallel as much
as possible. Tests with each combination of fly line or species and stimulus were conducted in at least
five different days with different progeny. The position of the odor and the odorless arm 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.

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. We established “*a priori*” a minimum sample size (*n*≥30), and data
collection for each experimental series/species/condition ended when significance was achieved, or
when *n*=55, regardless of the outcome. *S. pallida* was unusually inactive and so sample sizes were
sometimes lower but averaged *n*=28 across experimental series. This criterion was adopted because
binomial tests require an untenably large sample size to evince small deviations from an expected
proportion (*e.g.* *n*≥78 to detect a 10% deviation from the expected 50% random choice with *p*<0.05). In
cases where *n*<55, we conducted a similar number of tests for control and experimental lines to ensure
that results were comparable. Similarly, we also noted behavioral results when 0.05<*p*<0.08, because
such outcomes indicate that significance at *p*<0.05 could likely be achieved by increasing sample size
(Makin and de Xivry 2019), although increasing sample size is often difficult for this type of behavioral
experiments. For instance, a 60% choice with *n*=55 yields *p*=0.08, and sample size needs to be ≥74 for
achieving *p*<0.05. Statistical power was in most significant cases >0.8; exceptions include cases where
deviations from the expected distribution were small (*e.g.* a significant 10% deviation requires *n*=188 to
achieve 0.8 statistical power; Supplementary file 10).

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 Or67dGal4, 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 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.).
FIGURE CAPTIONS

Figure 1. Phylogeny of *Scaptomyza* and the orthologs of the odorant receptor *Or67b* (A) Time-calibrated Bayesian chronogram of *Scaptomyza* and *Drosophila* spp. inferred from nine protein coding and two ribosomal genes. Bars indicate 95% highest posterior density (HPD) age estimates in millions of years ago (MYA). 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*. Source species for the *Or67b* coding sequences crossed into the empty neuron systems in subsequent figures are indicated. Support values are only indicated for bifurcations with less than 100% support in any analysis, labeled as in (A). Branches with significant support (FDR p-value < 0.05) for dN/dS 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 dN/dS is greater than the background). Only *S. flava* and *D. mojavensis* branches have significantly elevated dN/dS according to branch model tests. *S. flava*, *S. pallida* and *D. melanogaster* labeled identically to (A). Scale bar units are substitutions per site.

Figure 2. Responses of homologous *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 (SSR) 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 (antennal basiconic 3A) OSNs (Hallem and Carlson 2006; left), or *Or67d* in at1 (antennal trichoid 1) OSNs (Kurtovic et al. 2007; right). Note that the at1 empty neuron system was used only for expression of *Sfla* Or67b2, as this Or was not functional in the ab3A empty neuron system. Although comparison between the responses of this Or with other *Or67b* proteins should be interpreted cautiously (Syed et al., 2010), Or tuning (but not response intensity) is independent from the empty neuron system expressed for Or expression (Figure supplement 5). The antennal basiconic sensilla houses the A (which expresses one of the Or67b proteins) and the native intact B neuron; the A neuron has larger amplitude spikes than the B neuron, allowing discrimination of spikes originating from either neuron. The antennal trichoid sensilla houses a single OSN expressing *Sfla* Or67b2; arrows indicate spikes (see also Figure supplement 4). Calibration bars (vertical lines to the right of traces): 10 mV throughout all figures unless noted. (B) Representative
electrophysiological recordings obtained from the targeted sensilla of flies expressing Or67b in OSNs in response to stimulation with apple cider vinegar, crushed rosette leaves of wild-type Col-0 Arabidopsis thaliana, and phytoalexin deficient 3 (PAD3) and aliphatic and indolic Glucosinolate Knock Out (GKO; myb28 myb29 cyp79b2 cyp79b3) A. thaliana mutants. Although all three A. thaliana genotypes have the same genetic background (Beekwilder et al. 2008; Glazebrook and Ausubel 1994), PAD3 plants are a more appropriate control for this study than GKO or Col-0, since PAD3 is deficient (as is GKO) in the production of camalexin but still produces normal levels of aliphatic or indolic glucosinolates. The bars above records indicate the onset and duration (1 sec) of the stimulation throughout all figures unless otherwise noted. (C) Responses (net number of spikes/second, control-subtracted, n=6-9 obtained from 2-4 animals) evoked by stimulation with apple cider vinegar, odors from homogenized mustard leaves (arugula and A. thaliana), grated mustard root odors (wasabi, horseradish, turnip and daikon), homogenized non-mustard leaf odors (tomato), and grated non-mustard root odors (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. Blue, gray, and green rectangles respectively show responses by stimulation with the Drosophila melanogaster attractant apple cider vinegar, non-isothiocyanate (ITC)-bearing plants, and ITC-bearing plants. Asterisks indicate significant differences between the control-subtracted net number of spikes and a threshold value for responsiveness (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, responded to stimulation with apple cider vinegar. Conversely, only neurons expressing S. flava Or67b paralogs responded to arugula odors (which bear ITCs). None of the S. flava paralogs responded to stimulation with the non-host (non-ITC bearing) tomato. 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 odors evoked responses from all Sfla Or67b paralogs but not from the Dmel or the Spal paralogs.

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. As before, the at1 empty neuron system was only used for expressing Sfla Or67b2 (see Figure Supplement 5 for responses to ITC from at1 OSNs expressing S.
pallida and S. flava proteins expressed in at1). (A) Representative electrophysiological recordings obtained from the targeted sensilla of flies expressing Or67b genes in the empty neuron systems in response to stimulation with acetophenone and BITC at 1:100 vol/vol. (B) 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). OSNs expressing any of the Sfla Or67b paralogs respond strongly and selectively to ITCs (see also Supplement Figure 5); these OSNs do not respond to benzyl thiocyanate stimulation (yellow bars), indicating that the ITC functional group (-N=C=S) is necessary for evoking responses from the S. flava paralogs. Spal Or67b and Dmel Or67b have similar odor-response profiles, responding mostly to stimulation with D. melanogaster activators and GLVs (*p<0.05, **p<0.01, one-sample signed rank tests performed as explained in Figure 2). Most odorants were diluted in mineral oil but some were diluted in other solvents as needed (see material and methods). Responses to the appropriate solvent control were subtracted from odorant-evoked responses. (C) 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 (Hallem and Carlson 2006). 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 in bottom.

Figure 4. Sfla Or67b1-3 have distinct ITC selectivity. (A) Dose responses of Sfla Or67b1, Sfla Or67b2 and Sfla Or67b3 (abbreviated as 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). As before, the at1 empty neuron system was used for expressing Sfla Or67b2 (see also Supplement Figure 5). 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 the three *Sfla* paralog normalized (to allow comparisons across paralogs) by each paralog’s median response to 1:100 vol/vol of BITC (the strongest ITC activator across all 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 ITC concentration, with many compounds evoking responses from all paralogs, particularly in the case of *Sfla* Or67b1 and Or67b3. The number of stimuli that evoked responses decreased with decreasing odorant concentration. Overall, three ITCs evoked responses from two paralogs down to 10⁻⁴ vol/vol. See Figure supplement 6 for analysis of responses in odor space.

**Figure 5.** Olfactory behavioral responses of *S. flava* and its microbe-feeding relatives *S. pallida* and *D. melanogaster* to ecologically relevant odors and ITCs. (A) Schematic representation of the dual choice y-maze used to test the olfactory responses of flies (see details in Figure supplement 8 and materials and methods). One arm of the maze offered constant odor/odorant airflow (apple cider vinegar, arugula, tomato or single ITC compounds at 1:100 vol/vol), while the control arm offered a constant odorless airflow (controls: water for odor sources, and mineral oil for single odorants). 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 apple cider vinegar odors and VOCs from leaves of arugula and tomato plants. Data represent the percentage of tests in which animals choose the odorous or odorless arms of the maze; numbers between parentheses 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**<0.01, ****p**<0.001, *****p**<0.0001; gray shaded bars serve to visualize that the proportion of test with flies orienting towards that arm of the y-maze is significantly different from random). *S. flava* was attracted to mustard (arugula) VOCs but not to non-mustard (tomato) VOCs. *S. flava* tended to avoid apple cider vinegar odors although differences (p=0.054). *D. melanogaster* was strongly attracted to apple cider vinegar but not to arugula or tomato leaf VOCs. *S. pallida* was only attracted to tomato leaf VOCs. See Figure Supplement 9 for responses to *A. thaliana* volatiles. (C) Olfactory behavioral responses of flies from the three species to single ITC compounds (20 µl of a 1:100 vol/vol dilution in mineral oil loaded in filter paper); the control arm had 20 µl of mineral oil loaded in filter paper. Assays were conducted as...
described in B. *S. flava* was strongly attracted to both ITC compounds tested, *S. pallida* was strongly repelled by BITC, and *D. melanogaster* was indifferent to either ITC.

**Figure 6. Ectopic expression of Sfla Or67b3 in Or22a OSNs or Or67b OSNs conferred behavioral responses to BITC in *D. melanogaster*.** The behavioral responses of *D. melanogaster* flies expressing different *Or67b* transgenes were tested in dual-choice assays (odorant vs. mineral oil) and experiments were conducted and analyzed as in Figure 5 but flies were starved 24 hours previous to testing. (A) Responses of flies expressing *Dmel Or67b* or *Sfla Or67b3* in Or22a OSNs lacking its cognate Or and tested with BITC 1:1,000 vol/vol vs. mineral oil. The two parental control lines (first two groups) and flies expressing *Dmel Or67b* (third group) were not attracted or repelled by BITC (binomial tests, p>0.05 in all cases), while flies expressing *Sfla Or67b3* were attracted to the odorant (* p<0.05). Importantly, parental genetic control lines retain olfactory attraction towards apple cider vinegar, a potent *D. melanogaster* olfactory attractant and activator (Figure Supplement 10). These results show that ITCs can evoke olfactory behavioral responses when *Sfla Or67b3* is expressed in the *D. melanogaster* Or22a olfactory pathway, an important circuit for detection of host odorants in many drosophilid species (Dekker et al. 2006a; Linz et al. 2013; Mansourian et al. 2018). (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. As in A, only flies carrying the *Sfla Or 67b3* transgene were attracted to BITC (* p<0.05, third group). (C) Responses of *D. melanogaster* expressing a silencer of synaptic activity (Kir2.1) in Or67b OSNs (third group), along with the responses of the two parental genetic control lines (first and second group) in tests with acetophenone 1:50,000 vol/vol, a strong *Dmel Or67b* activator (Münch and Galizia 2016; Figure 3). Genetic control flies showed a trend for attraction (0.05<p<0.08) to low concentrations of acetophenone, while flies with Or67b OSNs silenced lost attraction and were instead repelled by the odorant. These behaviors mirrored those of wild-type *D. melanogaster* flies tested with low and high concentrations of acetophenone, respectively (Figure supplement 11; Strutz et al. 2014). In A-C transgenes indicated to the left; gray shaded bars serve to visualize that the proportion of tests with flies orienting towards that arm of the y-maze is significantly different (p<0.05) from random.
Figure 7. A model for the evolution of Or67b and comparison with the known evolution of other Or and Ir orthologs in D. melanogaster and specialist drosophilid species. (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 two additional paralogous Or67b genes (Sfla Or67b1, Sfla Or67b2, and Sfla Or67b3; c), each of the paralogs has 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 protein 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; Linz et al. 2013; Mansourian et al. 2018; Keesey et al. 2019). The Ir75a protein 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 protein 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 copies are activated by ITCs only (bottom, paralog-specific activation by compounds 11-13, abbreviations as in Figure 3). 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. Notably, copies of Or22a, Or42b and Or85d have been evolutionarily lost in S. flava. These Ors are important in mediating attraction to canonical microbe (especially yeast)-associated odors (Goldman-Huertas et al. 2015).