Classification: Biological Sciences, Evolution

Title: Olfactory receptors tuned to volatile mustard oils in drosophilid flies

Authors: *Teruyuki Matsunaga[1], *Carolina E. Reisenman[2], *Benjamin Goldman-Huertas[3], Philipp Brand[4], Kevin Miao[1], Hiromu Suzuki[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.)

Keywords:

Scaptomyza flava, *Drosophila melanogaster*, herbivory, evolution, olfaction, isothiocyanate, chemoreceptor, olfactory receptor, TrpA1, wasabi, Or67b

Abstract (248):

Plant toxins are effective defenses because they are aversive to enemies. The same molecules, however, are co-opted as host-finding cues by specialist herbivores. Although such behavioral shifts are central to our understanding of herbivorous insect diversification, it is not well understood how they evolve. We addressed this in *Scaptomyza flava,* a herbivorous drosophilid fly within a lineage that shifted to feeding on toxic mustard plants (Brassicales) <10 million years ago. *S flava* lost the ancestral attraction to yeast volatiles and the attendant chemoreceptors that detect these odors. Here we report that *S. flava*, but not its close microbe-feeding relatives *Drosophila melanogaster* and *S. pallida*, is attracted to mustard host-plant odors, including volatile mustard oils (isothiocyanates or ITCs). Our genomic analysis uncovered three *S. flava* paralogs of an olfactory receptor gene (*Or67b*) that likely experienced positive selection. We then tested whether these chemoreceptors could underlie the observed attraction to volatile ITCs. Our *in vivo* recordings revealed that two of the *S. flava* Or67b proteins (Or67b1 and Or67b3) – but not the homologous Ors from microbe-feeding relatives – responded selectively and sensitively to volatile ITCs. These Ors are the only ITC chemoreceptors other than TRP channel family members (*e.g.*, the TrpA1 'wasabi' receptor) known from animals. Remarkably, *S. flava Or67b3* was sufficient to drive olfactory attraction toward butyl ITC when expressed in an attractive olfactory circuit. Our study illuminates that ancestrally aversive chemicals can be co-opted as attractants through gene duplication, leading to the origin of hedonic valence shifts in herbivorous insects.

Significance Statement (120)

Plant toxins trigger aversive olfactory (volatile-mediated) and gustatory (contact-mediated) responses in animals. Paradoxically, toxic plants are colonized by specialist insects that co-opt toxins as host-plant finding cues. The mechanisms underlying these behavioral shifts, from indifference or repulsion, to attraction, remain unclear. To address this, we used a fly lineage, *Scaptomyza flava*, that switched from yeast-feeding to feeding on mustard plants less than 10 million years ago. We found that *S. flava* is attracted to mustard-plant odors and volatile mustard oils (isothiocyanates or ITCs) such as 'wasabi', a behavior enabled by the evolution of genes encoding odorant receptors highly sensitive to ITCs. Our study illuminates how insects colonize toxic host plants through duplication and ecological repurposing of genes encoding pre-existing chemoreceptors.

INTRODUCTION

Many plant compounds used in food, agriculture and medicine originally evolved as toxins that deter and repel enemies (1). Among the most well-known compounds are those reactive electrophiles that induce the sensation of pain, such as diallyl disulfide and thiosulfinates produced by Alliaceae plants (*e.g.*, onions and relatives), α,β-unsaturated aldehydes found in Lauraceae (*e.g.*, cinnamon and olives), and isothiocyanates (ITCs) from mustards and their relatives in the Brassicales. Paradoxically, herbivores specializing on Brassicales evolved electrophile diversion or resistance mechanisms and use these plants as a food source and habitat (2, 3). Mustard specialists can be trapped with ITC baits, indicating that these insects use "aversive" electrophiles for host finding (4). However, the mechanisms mediating the detection of species-specific electrophiles and underlying shifts in hedonic valence, are unknown.

In animals, the only known and broadly conserved electrophile detecting sensors are members of the Transient Receptor Potential (TRP) nociceptive channel family (5–7), and the most studied among them is the Transient Receptor Potential Ankyrin-1 (TrpA1) (8–10). Reactive electrophiles form covalent adducts with cysteine and lysine residues in TrpA1, activating this non-selective cation channel and resulting in the perception of pain (9). A reasonable hypothesis explaining the attraction of specialist mustard-plant insects to host-plant-derived electrophiles is that the TrpA1 circuit has shifted from an aversive to an attractive one. However, TrpA1 is a gustatory, non-selective electrophile sensor activated by myriad electrophiles produced by plants in different orders, including diallyl disulfide, cinnamaldehyde, and ITCs, rendering this chemoreceptor ineffective in signaling host-plant identity (7). Alternatively, chemoreceptor proteins (*e.g.*, olfactory receptors -Ors- and/or ionotropic receptors -Irs-) that commonly mediate responses to volatile compounds in insects (11, 12) are better

81 suited to evoke responses that signal volatile, host-specific electrophiles. Specificity for
82 ecologically relevant volatiles repeatedly arose through evolution of these olfactory
83 chemoreceptors (13, 14). In *D. sechellia*, a specialist of the toxic "noni fruit", Ir75b evolved high
84 sensitivity to a signature noni volatile through a single amino acid residue substitution (15).
85 Similarly, in other drosophilid species, Or22a (and Or22b) evolved to be narrowly tuned to host-
86 plant volatile esters (16, 17), facilitating host-plant specialization.

87 Here, we investigate the evolutionary and functional mechanisms underlying a major
88 shift in host plant use: olfactory specialization on mustard plants by *S. flava*. *S. flava* is a relative
89 of *D. melanogaster* nested phylogenetically in the paraphyletic *Drosophila* subgenus and is a
90 specialist herbivore of ITCs-producing Brassicales plants (18, 19). In behavioral olfactory
91 assays, we found that *S. flava*, but not its microbe-feeding relatives *S. pallida* or *D.*
92 *melanogaster*, is attracted to volatile ITCs and odors from Brassicales plants. We scanned the
93 genome sequence of *S. flava* to identify rapidly evolving chemoreceptors as candidates
94 underlying this behavioral shift, and identified a lineage-specific expansion of three paralogous
95 copies of the olfactory receptor gene *Or67b* that exhibited signatures of positive selection in the
96 *S. flava* lineage. This *Or*, in contrast, is present as a single copy and is under strong evolutionary
97 constraint in *S. pallida* and *D. melanogaster*. *In vivo* functional testing of Or67b copies from
98 these species shows that two *S. flava* Or67b receptors, but not the conserved Or67b proteins from
99 its microbe-feeding relatives, responded sensitively and selectively to volatile ITCs.
100 Furthermore, *D. melanogaster* flies expressing *S. flava Or67b3* were attracted to butyl ITC in
101 olfactory assays, indicating that this receptor is sufficient for attraction to volatile ITCs in this
102 context. We also found that ITCs are not detected through covalent modification of Or67bs, as
103 opposed to TrpA1 (*e.g.* (7)). Or67b *S. flava* proteins selectively shifted their odor-receptive range

to ITCs, an entirely new chemical class, from an ancestral olfactome of ketones, alcohols, and aldehydes. In summary, we discovered an evolutionary mechanism by which specialist herbivorous insects can shift hedonic valence from aversion to attraction towards host plant-specific electrophiles. A novel ITC-sensitive olfactory receptor system supplements the generalist TRP sensory system of a mustard specialist, an adaptation that evolved through gene duplication and neofunctionalization.

RESULTS

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

Because *S. flava* is a Brassicales specialist, we first determined whether this species is attracted to mustard plant volatiles using a dual-choice olfactory assay (based on (20)). We found that both adult male and female *S. flava* are attracted to arugula (*Eruca vesicaria*) leaf odors (two-tailed Binomial tests corrected for multiple testing using false discover rate (FDR), $p<0.05$; Fig. 1). In contrast, *D. melanogaster* was neither attracted nor deterred by arugula odors but instead was strongly attracted to apple cider vinegar volatiles (FDR, $p=0.0001$), in agreement with previous results (*e.g.* (21)). In tests with apple cider vinegar, male *S. flava* preferred the odorless arm of the maze (FDR, $p<0.05$) and females showed a similar tendency (FDR, $p=0.0541$). *S. pallida* exhibited an intermediate behavior, distributing equally between the vinegar-laden and the odorless arms of the maze. Overall, these results demonstrate that *S. flava*, but not *D. melanogaster* or *S. pallida*, is attracted to mustard plant volatiles, and avoids odor sources that strongly attract *D. melanogaster*, characterized by the presence of acetic acid and a variety of ester, carbonyl, and hydroxyl-containing compounds (22, 23).

126 Brassicales plants produce ITCs upon tissue damage, which are highly toxic to non-
127 adapted insects (24). Yet, volatile ITCs are effective in capturing a variety of distantly related
128 specialist insects that have evolved mechanisms for coping with or preventing the formation of
129 ITCs during feeding (4). We investigated whether the sole presence of ITCs can mediate
130 olfactory attraction in *S. flava*. Adult female *S. flava* were attracted to both butyl ITC and sec-
131 butyl ITC (Fig. 1). Males, in contrast, were only attracted to butyl ITC, which suggests that at
132 particular ITCs might be involved in mediating sex-specific behaviors. Interestingly, *S. pallida*
133 oriented against butyl ITC (FDR, p-value<0.05), which may be a strategy to avoid ITCs as this
134 species sometimes feed on decomposing – but not intact – mustard plant leaves. In contrast, *D.*
135 *melanogaster* flies distributed at random between the arms of the maze regardless of the ITC
136 compound tested (FDR, p>0.5), indicating that ITCs do not elicit olfactory attraction in this
137 species. In summary, these experiments demonstrate that *S. flava*, but not its microbe-feeding
138 relatives, is attracted to mustard plant volatiles and ITCs.
139
140 **Phylogenetic analysis identified *Or67b* paralogs as candidates mediating the detection of**
141 **ecologically relevant odorants**
142 To identify potential *S. flava* chemosensory receptors involved in the attraction of
143 mustard volatiles, we conducted a phylogenetic analysis of the Or protein sequences of *S. flava*
144 and four other *Drosophila* species to identify *Scaptomyza*-specific losses and gains (Fig. 2*A*).
145 The Or topology was largely congruent with previous *Drosophila* Or trees (25) except for some
146 deeper nodes, which had weak to moderate bootstrap support. In order to identify *Ors* evolving
147 under positive selection in *S. flava*, we inferred d*N*/d*S*, the ratio of nonsynonymous to
148 synonymous substitution rates, using a foreground/background branch test as implemented in

149 PAML (26). Of 75 *S. flava* branches tested, seven branch models (for paralogs of *Or63a, Or67b*
150 and *Or98a*, and a homolog of *Or67d*) indicated an *S. flava* foreground rate >1, consistent with
151 positive selection (Table S1). Of these receptors, homologs of *Or67d* detect the *Drosophila*
152 pheromone 11-cis-vaccenyl acetate (27) and likely function in intraspecific communication in
153 *Scaptomyza*, *Or63a* is only expressed in larvae (28), and the *Or98a*-like genes found in
154 *Scaptomyza* have no *D. melanogaster* homologs and are functionally uncharacterized. In
155 contrast, *Or67b* is expressed in adults and modulates oviposition behavior (29), making it a good
156 candidate. In a previous study, branch-site tests were also consistent with episodic positive
157 selection acting on *Or67b* paralogs from *S. flava* (19). After expanding the representation of
158 *Or67b* homologs in a phylogenetic analysis, and conducting branch tests in PAML on all
159 branches in the tree, we found support for elevated d*N*/d*S* exclusively in *S. flava* and *D.*
160 *mojavensis* branches (Fig. 2*B*; TableS1). Thus, we focused on *Or67b* paralogs for functional
161 studies.

162

163 **_S. flava_ Or67b proteins respond specifically to mustard plant volatiles**

164 We wondered whether any of the three *S. flava* Or67b proteins could mediate the odor-
165 specific attraction of *S. flava* to mustard plant volatiles. First, we confirmed that all three
166 paralogs are expressed in adult *S. flava* (Fig. S2*B*). To study the odor-response profile of Or67b
167 across species, we used a *D. melanogaster* mutant that lacks its endogenous olfactory receptor
168 (Or22a) in the ab3A olfactory receptor neuron (known as the "empty neuron system" (30, 31).
169 We conducted single sensillum recordings from the ab3 sensilla in transgenic *D. melanogaster*
170 flies expressing each of the *S. flava* (*Sfla*) *Or67b* paralogs, as well as the *Or67b* orthologs from
171 *S. pallida* (*Spal*) and *D. melanogaster* (*Dmel*) (Fig. 3*A*). *Or22a* deletion in ab3A was confirmed

by the lack of responses in ab3A to ethyl hexanoate, a strong Or22a activator (Fig. S4*A*). We found that ab3A neurons expressing any of the three *Sfla Or67b* paralogs responded to all mustard plant volatiles tested (Fig. 3*B*; Mann-Whitney U tests adjusted for multiple comparisons using FDR, $p<0.05$ in all cases). In contrast, neurons expressing *Dmel Or67b* responded strongly to apple cider vinegar (FDR, $p<0.05$), but not to any of the mustard plant volatiles tested (FDR, $p>0.1$). Neurons expressing *Spal Or67b* also showed strong responses when stimulated with apple cider vinegar (FDR, $p<0.05$), but displayed only modest responses to arugula volatiles (FDR, $p<0.05$). Interestingly, stimulation with apple cider vinegar decreased spiking activity of neurons expressing *Sfla Or67b1* (FDR, $p<0.05$).

Although neurons expressing any of the *Sfla Or67b* paralogs responded to mustard plant volatiles, they differed in their responsiveness. Stimulation with arugula volatiles, for instance, elicited stronger responses in neurons expressing *Sfla Or67b1* (64 ± 7.8 spikes/sec; average ± SE) than in neurons expressing *Sfla Or67b2* (24 ± 8.7 spikes/sec) or *Sfla Or67b3* (22.3 ± 2.3 spikes/sec) (Fig. 3*B*; Kruskal Wallis ANOVA, $p<0.005$, followed by post-hoc Tukey tests, $p<0.05$; $n=6$ in each group). In contrast, stimulation with horseradish volatiles produced stronger responses in neurons expressing *Sfla Or67b3* (32 ± 4.4 spikes/sec) than in *Sfla Or67b1* (8.75 ± 1.3 spikes/sec) or *Sfla Or67b2* (12.8 ± 2.3 spikes/sec) (Fig. 3*B*; Kruskal Wallis ANOVA, $p<0.005$, followed by post-hoc Dunn's tests, $p<0.05$; $n=6$-9 in each group). Taken together, these results demonstrate that *Dmel Or67b* and *Spal Or67b* are functionally similar and may represent the ancestral phenotype, while *Sfla Or67b* paralogs are divergent and tuned to mustard plant volatiles.

194 **_S. flava Or67b_ paralogs respond selectively to ITCs and have different molecular receptive**
195 **ranges**

196 Upon herbivore damage, mustard plants release ITCs and nitriles in addition to green leaf
197 volatiles (GLVs) (32). To elucidate which odorants mediate activation of *Sfla* Or67b proteins, we
198 tested a panel of 44 odorants which include ITCs, nitriles, GLVs, *Dmel* Or67b activators bearing
199 various functional groups (33), and four ecologically relevant odorants which do not fall within
200 these categories. Because many compounds were tested, for statistical comparisons we selected
201 the odorants that produced the strongest average response across all Ors within each of the six
202 chemical categories tested (those odorants were butyl ITC, mandelonitrile, cis-3-hexenyl
203 butyrate, trans-2-pentenal, and geranyl acetate (Fig. 3*A* and *C*). Neurons expressing any of the
204 three *S. flava* paralogs responded to stimulation with butyl ITC (FDR, $p<0.05$), but not to
205 stimulation with acetophenone (FDR, $p>0.1$). Conversely, acetophenone evoked the strongest
206 excitatory responses from neurons expressing *Dmel Or67b* and *Spal Or67b* (respectively 175.7 ±
207 17.6 and 78 ± 15.2 spikes/second; average ± SE), while butyl ITC evoked no responses from
208 neurons expressing *Dmel Or67b* (FDR, $p>0.5$), and only produced a slight increase in spiking
209 activity (15 ± 4.2 spikes/second; average ± SE; FDR, $p<0.05$) from neurons expressing *Spal*
210 *Or67b*.

211

212 **_S. flava_ Or67b1 and Or67b3 differ in their odor selectivity**

213 Although all *S. flava* paralogs are narrowly tuned to ITCs, they differ in their degree of
214 selectivity and sensitivity. Butyl ITC evoked strong responses in neurons expressing *Sfla Or67b1*
215 and *Or67b3* (196.3 ± 15.2 and 170.6 ± 11.2 spikes/second; average ± SE) that were not different
216 from each other (Kruskal-Wallis ANOVA followed by Dunn's tests; $p>0.05$), while stimulation

with this odorant produced only a modest response from neurons expressing *Sfla Or67b2* (26.7 ± 2.8 spikes/second; Dunn's tests, in both cases $p<0.05$). None of the 44 odorants tested evoked strong responses from neurons expressing *Sfla Or67b2*, but it is possible that other compounds not tested in this study, and in particular mustard plant volatiles (Fig. 3*C*), activate this receptor.

Next, because ITCs are highly diverse in structure (34), we hypothesized that a particular subset of ITCs differentially activate *Sfla* Or67b1 and *Sfla* Or67b3. In order to test this, we stimulated neurons expressing *Sfla Or67b1* and *Or67b3* with various concentrations of six ITCs (3-methyl-thio-propyl ITC, butyl ITC, isobutyl ITC, sec-butyl ITC, benzyl ITC, and phenethyl ITC) that evoked very strong responses (>123 spikes/second) from neurons expressing *Or67b3* at 1:100 vol/vol (Fig. S4). At that concentration, only half of these odorants (3-methyl-thio-propyl ITC, butyl ITC, isobutyl ITC, and benzyl ITC) produced moderate to strong responses (41-147 spikes/second, depending on the odorant) from neurons expressing *Sfla Or67b1*. At 1:1,000 vol/vol, the responses of *Sfla* Or67b3 to stimulation with phenethyl ITC, sec-butyl ITC and butyl ITC were different from those of *Sfla* Or67b1 (Mann-Whitney U tests adjusted for multiple comparisons using FDR, $p<0.05$ in all cases). Thus, although both *Sfla* Or67b1 and Or67b3 are highly selective and sensitive to ITCs, they have distinct odorant receptive ranges: *Sfla* Or67b3 is broadly tuned and strongly responds to all ITCs tested, while *Sfla* Or67b1 has a narrower receptive range and responds strongly only to a smaller subset of ITCs.

Or67b response relatedness, specificities, and functional evolution

We next compared the responses of all the Or67b proteins to all 44 odorants tested in a multi-dimensional space using principal component analysis (Fig. 4*A*). Each dot in Figure 4A shows the vectors quantifying the control-subtracted odor responses of each Or67b copy plotted

along the dimensions of the first three principal components, which captured 63.1%, 26.0%, and 9.0 % of the variance of the original high-dimensional data set. This analysis shows that: 1) most of the ITCs cluster separately from the other odorants tested, contributing to most of the variance (PC1) of the data set; 2) some *Dmel* activators and GLVs such as acetophenone and cis-3-hexenyl-butyrate also produced distinct responses, contributing largely to the variance of PC2; and 3) ITCs that activate *Sfla* Or67b1, such as sec-butyl ITC (but not butyl ITC) contribute to the variance of PC3. We also constructed tuning curves for each Or67b protein (Fig. 4*B*). The strongest responses (center of the distribution) from *Dmel* Or67b and *Spal* Or67b were evoked by *Dmel* activators and GLVs, while the strongest responses from all *Sfla* Or67b paralogs were evoked by ITCs. *Sfla* Or67b1 is particularly narrowly tuned to a few ITCs, as revealed by the high kurtosis value of the distribution (Fig. 4*B*). Then we conducted a neighbor joining cluster analysis based on the Euclidean distances between the odor-evoked responses from all the Ors. Among the *S. flava* paralogs, Or67b1 and Or67b3 have the most similar (shorter Euclidean distance) odorant responses (Fig. 4*C*).

Activation of *S. flava* Or67b1 and Or67b3 requires the presence of the ITC functional group

The finding that *Sfla* Or67b1 and *Sfla* Or67b3 respond to ITCs suggest that both the presence of the ITC functional group (–N=C=S) and the size of the molecule determine binding specificity. These two factors are indeed important to evoke responses from TrpA1, a non-selective electrophile chemoreceptor that relies on ligand-directed covalent modification (10). We addressed whether this is also the case for *Sfla* Or67b1 and *Sfla* Or67b3, by testing their responses to eleven electrophiles (non-ITCs) that activate TrpA1 through covalent linkages with

sulfhydryl side chains. Interestingly, none of the non-ITC electrophiles activated neurons expressing *Sfla Or67b1*, and among these, only trans-2-pentenal produced a very small increase in spike activity (4 ± 1.3 spikes/second) from neurons expressing *Sfla Or67b3* (Fig. 3*C*). Additionally, we found that neurons expressing *Sfla Or67b* can respond to odor pulses that are seconds apart (Fig. S5). Because recovery from ligand-induced covalent modifications takes at least several minutes (and often hours) (7), these results suggest that activation of *Sfla* Or67b1 and Or67b3 by ITCs does not rely on the TrpA1-like covalent modification strategy.

We next investigated whether the presence of the ITC functional group (–N=C=S) is necessary for determining the responses of *Sfla* Or67b1 and *Sfla* Or67b3. In order to test this, we compared the responses of these Ors to benzyl ITC (BITC) and benzyl thiocyanate (-S≡C-N: BTC) (Fig.4*B*), linkage isomers of the same molecular weight. We found that BITC produced robust activity in neurons expressing *Sfla Or67b1* or *Sfla Or67b3* (Mann-Whitney U tests, in both cases $p<0.002$), while BTC had little or no effect (Fig. 3*C* and 4*B*). This differential activation pattern is likely due to the differences in the functional group because BITC and BTC are not only isomers, but also have similar volatilities. These findings are consistent with the hypothesis that *Sfla* Or67b1 and *Sfla* Or67b3 detect ITCs by interacting with the ITC (–N=C=S) functional group selectively.

S. flava Or67b3 is sufficient to induce olfactory attraction to ITCs

Because *Sfla* Or67b paralogs, but not the orthologs from *Spal* and *Dmel*, specifically respond to ITC compounds, we asked whether these receptors can mediate olfactory attraction to these odor compounds. We focused on *Sfla* Or67b3, because this Or has a broader sensitivity to ITCs than *Sfla* Or67b1 (Fig. 4*B*). We investigated this by testing the behavioral responses of *D.*

melanogaster "empty neuron" mutants that expressed *Sfla Or67b3* or *Dmel Or67b* in the ab3A neuron. As before, we used a dual-choice "Y-shaped" olfactometer offering an odorant stimulus (butyl ITC 1:100 or 1:1,000 vol/vol) in one of the arms, and the solvent control in the other arm. We found that at the 1:100 vol/vol concentration of butyl ITC the odor responses of flies expressing *Sfla Or67b3* or *Dmel Or67b3* were different from the 50% random expectation (two-tailed Binomial tests, $p<0.005$ and $p<0.01$ respectively), with flies orienting towards the odor in ca. 70-71% of tests (Fig. 5). In contrast, at the lowest odor concentration (1:1,000 vol/vol), the odor responses of flies expressing *Sfla Or67b3* – but not that of those expressing *Dmel Or67b* – were different from the 50% random expectation (two-tailed Binomial tests, $p<0.00005$), with flies choosing the odorous side of the maze in 83% of tests. Moreover, at the lower concentration, the odor-responses of flies expressing the *S. flava* transgene were different from those of flies expressing the *Dmel* transgene (Fisher Exact test, $p<0.05$). None of the three genetic control *Dmel* lines showed a preference for either side of the maze (Binomial tests, $p>0.1$ in all cases), even at the 1:100 vol/vol concentration. These results demonstrate that *Sfla Or67b3* can confer odor sensitivity to ITCs that persists at low (1:1,000 vol/vol) concentrations in a fly that is otherwise not attracted to these odor compounds.

Discussion

Plants have evolved a myriad of toxic compounds to deter enemies which are generally perceived as noxious by animals (1). In spite of this, the vast majority of herbivorous insect species are specialized on one or a few host-plant lineages that contain aversive compounds such as ITCs – the chemicals that give wasabi its characteristic pungent taste (35, 36). Through the course of evolution, specialist insects co-opted these chemicals as cues for host-finding, feeding,

and egg laying (37). Here, we investigated how attraction to these ancestrally aversive chemical cues evolved, using *S. flava*, a recently-derived herbivorous fly of mustard plants and their relatives (Brassicales). In behavioral experiments we found that *S. flava*, but neither of its microbe-feeding close relatives *S. pallida* or *D. melanogaster*, is attracted to mustard-plant volatiles, including volatile ITCs (Fig. 1). We next searched for changes in chemoreceptor proteins that could underlie this host plant-specific olfactory attraction. We found a candidate olfactory receptor gene (*Or67b*), which was triplicated in a recent ancestor of *S. flava*, resulting in three divergent and paralogous odorant receptor genes (*Sfla Or67b1-b3*; Fig. 2). These *Sfla* Or67b paralogs evolved with rapid rates of amino-acid changing mutations, consistent with episodic positive selection (Fig. 2*B*; TableS1). Next, using the "empty neuron" system of *D. melanogaster*, we investigated the odor-response profiles of these receptors. Consistent with our behavioral results, we found that two of the *Sfla* Or67b paralogs (*Sfla* Or67b1 and *Sfla* Or67b3) specifically respond to stimulation with mustard-plant odors and volatile ITCs (Fig. 3 and Fig S5). In contrast, *S. pallida* and *D. melanogaster* paralogs responded strongly to stimulation with apple cider vinegar and a broad range of aldehydes, alcohols and ketones (Fig. 3 and Fig. 4), consistent with their microbe-feeding habits. We also found that the presence of the ITC functional group is necessary for activation of *Sfla* Or67b and furthermore, that activation of the receptor likely does not involve a TrpA1-like covalent modification (Fig. 3, Fig. 4 and Fig. S5 (7)). Remarkably, when *Sfla Or67b3* was expressed in ab3A "empty neuron" system of *D. melanogaster*, adult female flies were attracted to butyl ITC (Fig. 5). This suggests that *Sfla Or67b3* may be sufficient for the behavioral valence shift observed in the evolution of mustard-feeding *Scaptomyza*, when expressed in novel neural circuit that ancestrally mediates attraction. Our results reveal the first Ors characterized from any animal that are tuned to mustard oils, and

332 illustrate an evolutionary mechanism by which mustard specialists can evolve attraction towards
333 otherwise aversive chemical compounds.

334 The gene triplication event that occurred in the *S. flava* lineage raises several interesting
335 molecular evolutionary questions. Gene duplication plays an important role in evolutionary
336 innovation (38) and several outcomes have been proposed regarding to how gene duplication
337 contributes to acquisition of novel gene functions (38, 39): (A) neofunctionalization occurs when
338 one of the two duplicated genes (paralogs) acquires a new function after accumulating *de novo*
339 mutations, while the other copy retains its ancestral function; (B) subfunctionalization is largely
340 considered a degenerative process whereby mutations accumulate in both copies, leading to
341 partitioning of the ancestral function between the two gene copies; (C) specialization occurs
342 when subfunctionalization and neofunctionalization evolve simultaneously, yielding gene copies
343 that are "functionally distinct from each other and from the ancestral gene" (40, 41). The
344 specialization model is supported by our data, as neither of the two outgroup Or67b proteins
345 (*Spal* Or67b and *Dmel* Or67b) responded to ITCs (neofunctionalization), coupled with the fact
346 that *Sfla* Or67b1 has a narrower breadth of tuning to ITCs than *Sfla* Or67b3
347 (subfunctionalization). Our conceptual model for this evolutionary hypothesis (Fig. 5) is that the
348 ancestral Or67b ("1") was broadly tuned to acetophenone, alcohols and ketones, and diverged
349 into a mustard oil-specific Or67b ("2") that lost sensitivity to these compounds and gained
350 responsiveness to ITCs in the *S. flava* lineage. This *Or* triplicated during the course of evolution,
351 ultimately giving rise to three Ors with different odor selectivity ("3"): (a) *Sfla* Or67b3, which
352 retained broad sensitivity to ITCs; (b) *Sfla* Or67b1, which become narrowly tuned to a subset of
353 *Sfla* Or67b3 activators; and (c) *Sfla* Or67b2, which lost sensitivity to a broad range of ITCs (at

354 least all the nine ITC compounds tested here) and likely retained or gained sensitivity to untested
355 compounds.

356 Across *Drosophila*, orthologous chemoreceptors respond in a species-specific manner to
357 ecologically relevant ligands. The best-studied of these is the *Or22a* gene, which has been
358 evolutionarily deleted in *S. flava* but not in *S. pallida* (19). In several drosophilids, Or22a is
359 tuned to different ecologically relevant esters: ethyl-hexanoate in *D. melanogaster* (16), methyl-
360 hexanoate in *D. sechellia* (42), 3-methyl-2-butenyl acetate in *D. erecta* (17), and isobutyl acetate
361 in *D. suzukii* (43). Similarly, Ir75b is tuned to butyric acid in *D. melanogaster* and to hexanoic
362 acid in *D. sechellia* (15). The fact that the chemical classes are still conserved for each
363 chemoreceptor (i.e. esters in the case of Or22a and acids in the case of Ir75b), suggests that the
364 chemical niche space and underlying sensory mechanisms are relatively conserved across
365 species. In contrast, while *Spal* Or67b and *Dmel* Or67b respond to alcohols, aldehydes and
366 ketones, *Sfla* Or67b1 and *Sfla* Or67b3 respond selectively to volatile ITCs, thus acquiring a
367 previously undescribed sensitivity to an entirely different chemical compound class (Fig. 6*b*).
368 We hypothesize that the striking difference between Or67b-activating ligands among these
369 species resulted from the dramatic shift from feeding on microbes to feeding on live mustard
370 plant tissue that occurred in the evolution of herbivorous *Scaptomyza*.

371 TrpA1, which mediates taste-aversion to ITCs, is one of the most well-studied contact
372 chemoreceptors (7, 8, 10). However, volatile ITCs are widely used to trap agricultural pests of
373 Brassicales (4), and in agreement with this, the antennae of some of these insects respond to
374 volatile ITCs (44). Our study further advances the understanding of how volatile ITCs may be
375 detected by these insects. Furthermore, we uncovered a potential evolutionary mechanism, gene
376 duplication followed by neofunctionalization and subfunctionalization, by which ITC-responsive

377 Ors evolved in Brassicales specialists. More generally, our results suggest that major host shifts
378 and specialization events in herbivorous insects can result from simple genetic changes of large
379 effect in the peripheral nervous system, similar to the apple maggot fly *Rhagoletis pomonella*
380 (45). This in turn may explain why such events are so common across herbivorous insect
381 lineages, resulting in rapid rates of diversification and the most diverse guild of macroscopic life
382 to have evolved (46, 47).
383
384
385 **Material and Methods:**
386 Material and methods are summarized below. For complete details, see Supplemental Methods.
387 **Fly lines**
388 M2-MD refers to a CRISPR/Cas9-mediated deletion of *Or22a/b* and a knock-in of *Gal4*
389 and *DsRed* by homologous repair. *Gal4* is not functional but *DsRed* is expressed in the eye. The
390 functional absence of *Or22a* and *Or22b* genes in M2-MD flies was confirmed by
391 electrophysiological analysis on *Or22a/b* expressing neurons in wild-type flies (Fig. S3). The
392 four *UAS-Or67b* lines created in this study, the *UAS-Dmel Or67b* line, and the *UAS-dTrpA1*
393 (stock no. 26264) line was each crossed into the M2-MD line. The progeny were then used for
394 SSR recordings and behavioral experiments.
395
396 **Behavioral tests of olfactory attraction**
397 The olfactory responses of mated, non-starved *D. melanogaster* (Canton-S or transgenic),
398 *S. pallida*, and *S. flava* were tested using a dual-choice "Y-shaped" olfactometer based on one
399 previously published (20). 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. Four insects were released at once (to increase experimental efficacy), but only the first choice was considered.

The odorants (20 µl of 1:100 vol/vol mineral oil solution) used in experiments were butyl ITC (Sigma-Aldrich, CAS # 592-82-5, USA) and sec-butyl ITC (Sigma-Aldrich, CAS # 15585-98-5, USA). For tests with transgenic flies, responses were tested at 1:100 vol/vol and 1:1,000 vol/vol of butyl ITC. We also used apple cider vinegar (40 µl, O Organics, USA; 40 µl of distilled water was as control stimulus in these tests). For tests of host orientation, two-four leaves from young arugula plants grown in an insect and insecticide/pesticide free chamber or greenhouse were excised just before tests.

For each odor/odorant, species, sex and genotype, the number of tests with insects orienting towards the odorous arm and the odorless arm of the Y-maze was used to conduct two-tailed Binomial tests (48). When appropriate, p-values were adjusted for multiple testing/comparisons using the false discovery rate (FDR) method of Benjamin-Hochberg (49); results were considered statistically significant if the Benjamini-Hochberg adjusted p-value was <0.05. For all tests, we verified that the power was >0.8.

Molecular phylogeny of drosophilid odorant receptors:

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, (50)) and *S. flava* sequences were obtained from ref. (19). Three hundred and nine sequences were aligned in MAFFT v7.017 (51). Models were fitted to the alignment using IQ-Tree and tested using the AIC criterion (52). A maximum likelihood (ML)

423 phylogeny was generated using the Or protein alignment in RAxML v8.2.10 (53). Orco
424 sequences were designated as the outgroup.

425

426 **Molecular phylogeny of drosophilid *Or67b* genes:**

427 *Or67b* CDS from *D. grimshawi*, *D. mojavensis, D. virilis D. sechellia, D. simulans*, *D. erecta, D.*
428 *yakuba, D. pseudoobscura, D. persimilis, D. ananassae, D. willistoni* and *D. melanogaster*
429 (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
430 r1.3, dana r1.3, dwil r1.3 and dmel r6.28, respectively) were downloaded from Flybase
431 (www.flybase.org, (50)). The *S. pallida* DNA sequence was obtained through PCR. Sequences
432 were aligned, models fitted and chosen according to AIC (GTR+I+G) in IQ-Tree (52). Trees
433 were inferred using RAxML (v8.2.10) with the GTRCATI model and 1000 rapid bootstraps, and
434 MrBayes (v3.2.6) (54).

435

436 **Analysis of molecular evolution:**

437 CDS of homologs of every *Or* gene in *S. flava* found in the 12 *Drosophila* genome builds were
438 aligned to *S. flava Or* CDS. Homology was assessed according to inclusion in well supported
439 clades in the *Or* translation phylogeny above. Sequences were aligned in MAFFT (v7.017) (51).
440 Phylogenies were generated for every alignment using PhyML (55). If these trees showed >70%
441 bootstrap support for a topology contrary to the known species topology, or if the Or homology
442 group contained duplicates, these trees were used in PAML analyses instead of the species tree.
443 Branch models of sequence evolution were fit using PAML 4.9h (26). A foreground/background
444 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 for every unique phylogeny (total: 75 tests). 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 with identical likely ratio tests performed as above (29 tests total Fig. 2*B*; Table S1). P-values were adjusted for multiple testing using the FDR method (49).

Distance tree based on SSR recordings from neurons expressing *Or67b* transgenes

A matrix of average responses from five *Or67b* transgene receptors to 44 odorant compounds (1:100 vol/vol concentration) was produced (Table S3). Net responses were obtained by subtracting the response to mineral oil or dimethyl sulfoxide solvents. A Euclidean distance matrix was generated using the *dist* function from the R stats package (56). Receptor responses were clustered by using the neighbor joining (NJ) algorithm on this distance matrix. Support for clusters was assessed using 1,000 bootstraps of the original response matrix by generating distance matrices and NJ trees on the pseudo-datasets in ape (v5.3) (57).

Single sensillum recordings (SSR)

Adult female flies were fed and then prepared for SSR as previously described (58). Extracellular activity was recorded by inserting a tungsten electrode into the base of the ab3 sensillum. The following odor sources (all purchased in Berkeley, California, US unless otherwise mentioned; 20 µl of material loaded on filter paper) were used: apple cider vinegar (O Organics, USA), grated roots of *Wasabia japonica* (wasabi), organic roots of *Armoracia rusticana* (horseradish), *Brassica rapa* (turnip), *Raphanus sativus* (daikon), and *Beta vulgaris*

(beet). *Eruca vesicaria* (arugula) was grown from seeds in the lab, and leaves of 3-8 weeks old plants were used for odor stimulation. Roots or leaves were grated and homogenized using a vegetable grater to a volume equivalent to ~500 µl. The "net number of spikes" were obtained by counting the number of spikes during a 1-second window 0.2 seconds after the onset of stimulation, and subtracting from this number the background spiking activity (number of spikes during 1-second previous to the onset of odor stimulation). Data were analyzed using Mann-Whitney Rank sum tests (for comparisons involving two means) and p-values were adjusted for multiple comparisons using the FDR method, or by Kruskal-Wallis ANOVAs followed by Dunn's or Dunnet's tests (for comparisons involving more than two means).

Data Analysis and figure generation:

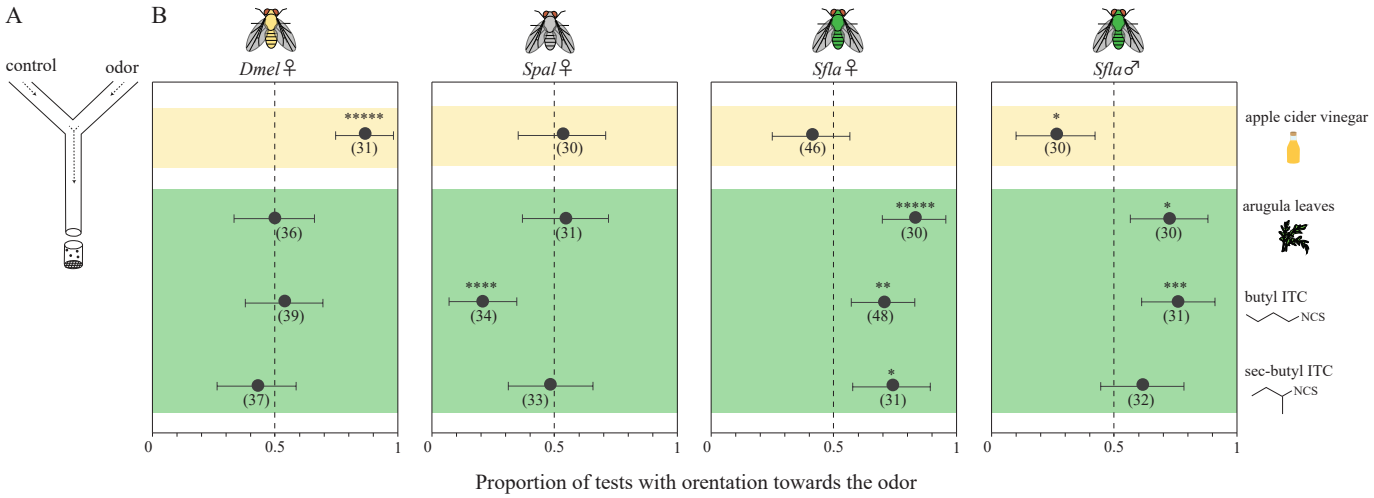
All images and drawings are originals prepared by the authors. Figures were prepared via a combination of WinEDR (v3.9.1), R Studio (v1.2.1335), Microsoft Excel (2016), Adobe Illustrator (2019), Python, and Geneious (10.0.9).

Acknowledgments:

We are grateful to Drs. Chauda Sebastian, Dennis Mathew, John Carlson and Bloomington Drosophila Stock Center (NIH P40OD018537) for sharing *M2-MD*, *UAS-Dmel Or67b,* and *UAS-dTrpA1* (stock no. 26264), 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. We thank members of the Whiteman and Scott Laboratories for discussions and comments on the manuscript. This work was supported by the National Institute of General Medical Sciences of

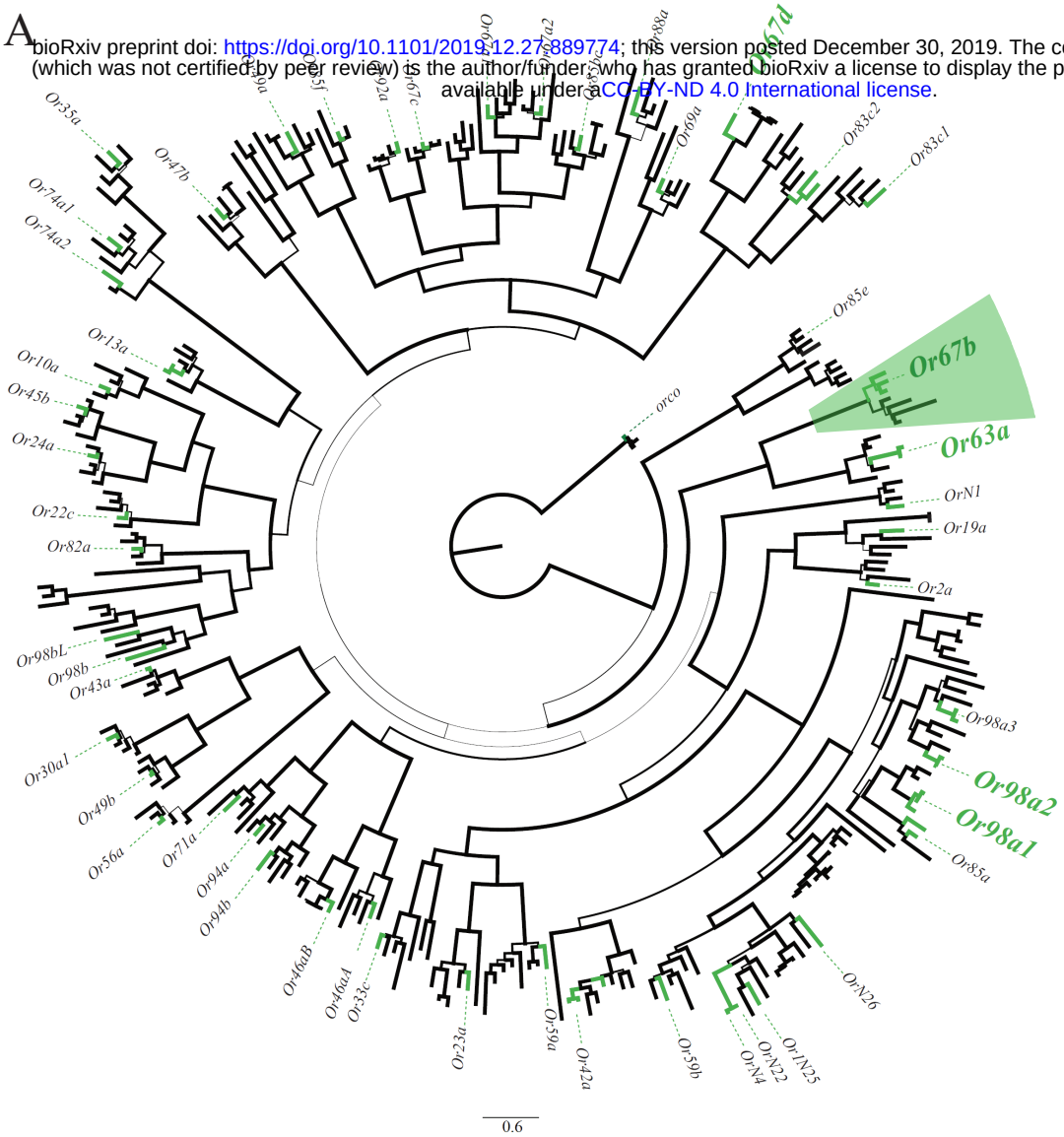
the National Institutes of Health (award number R35GM119816 to N.K.W.) and the National Science Foundation (award number IOS 1755188 supporting B.G.-H.).

Figure 1



Proportion of tests with orentation towards the odor

A



B

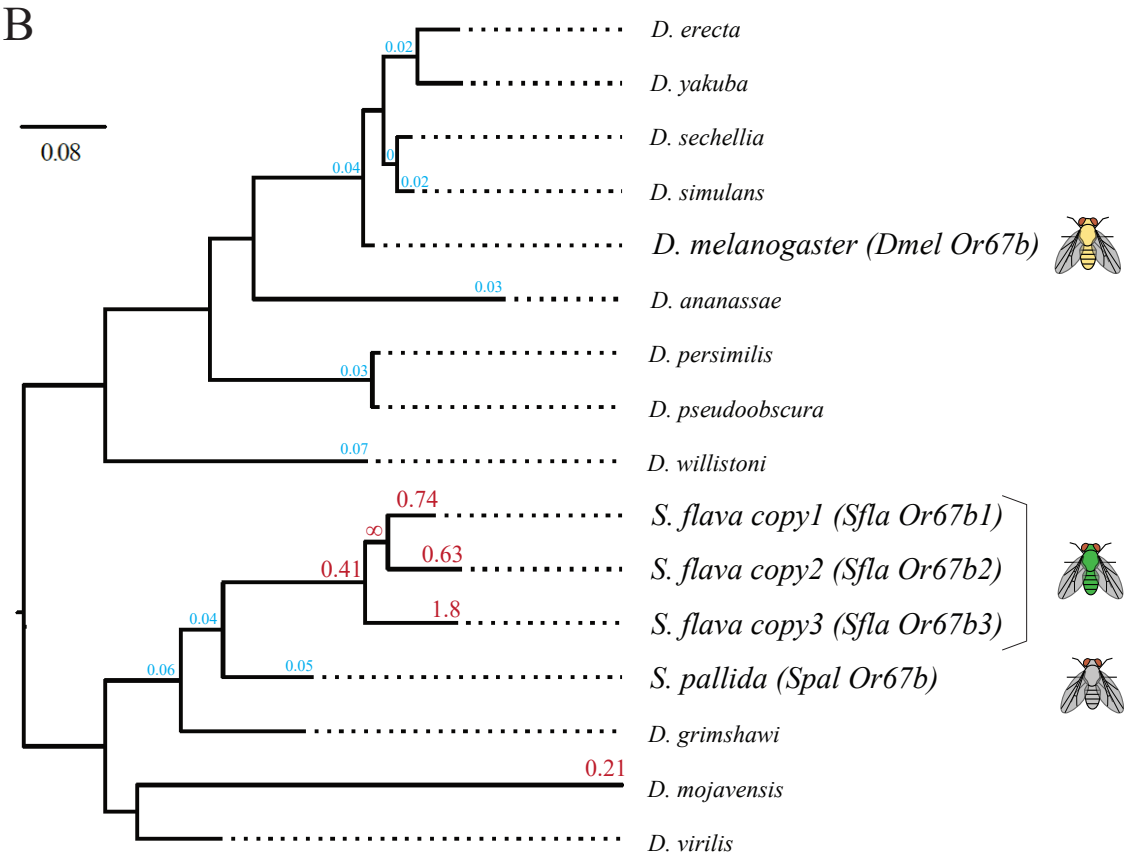
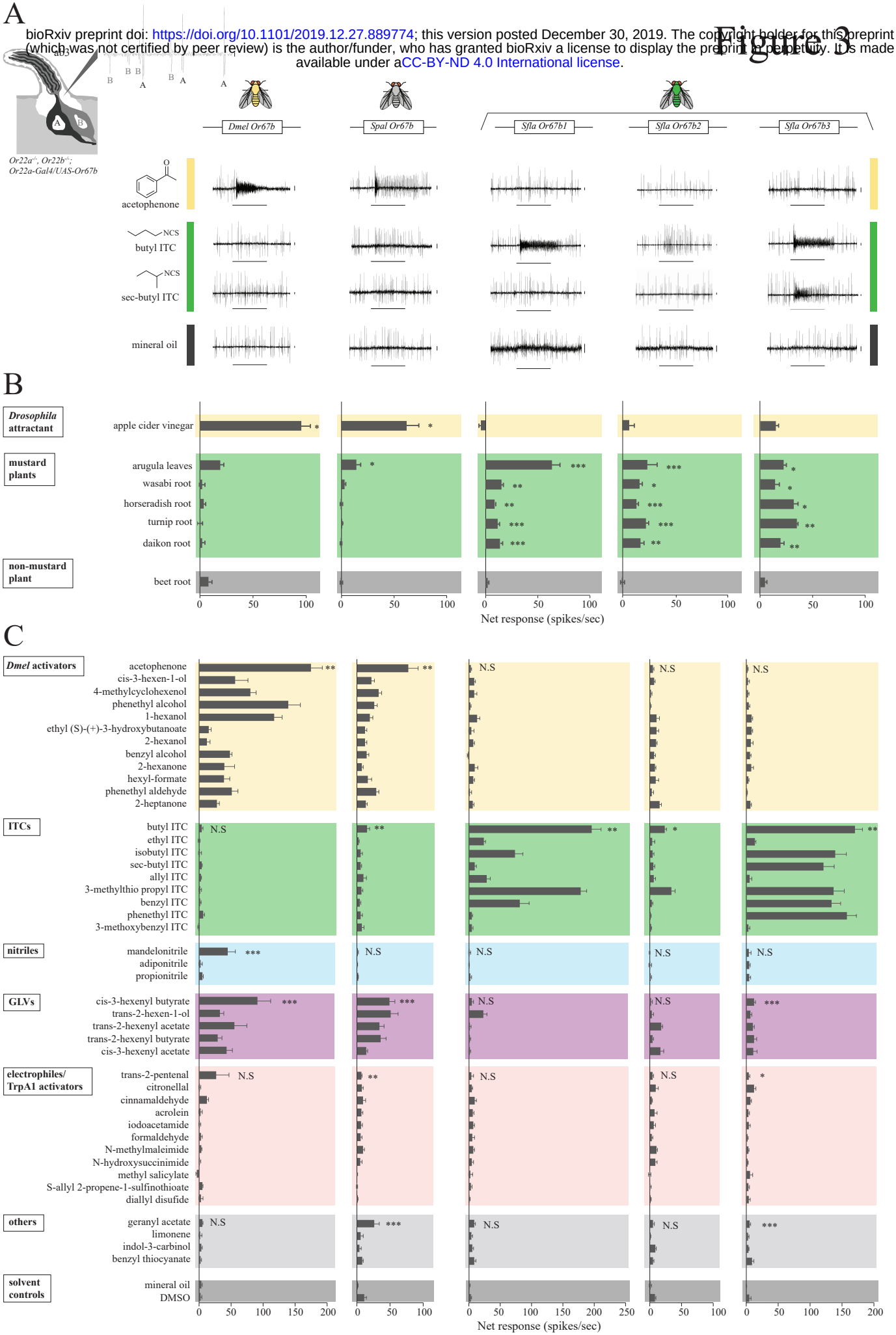
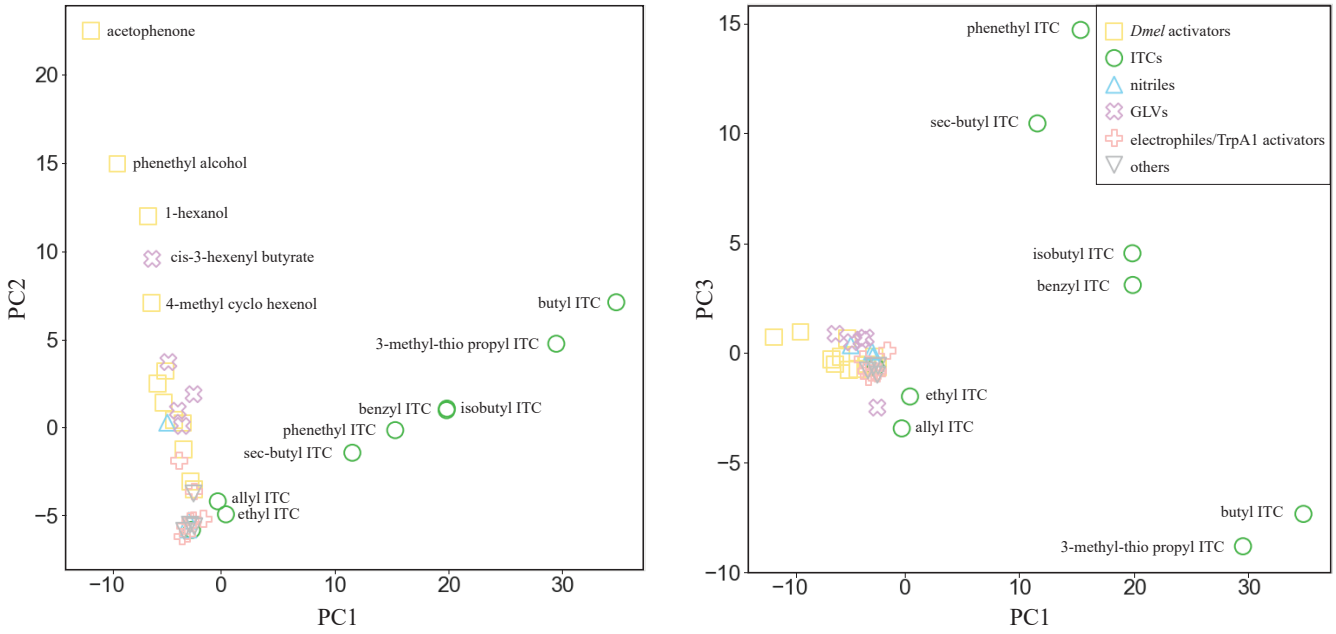


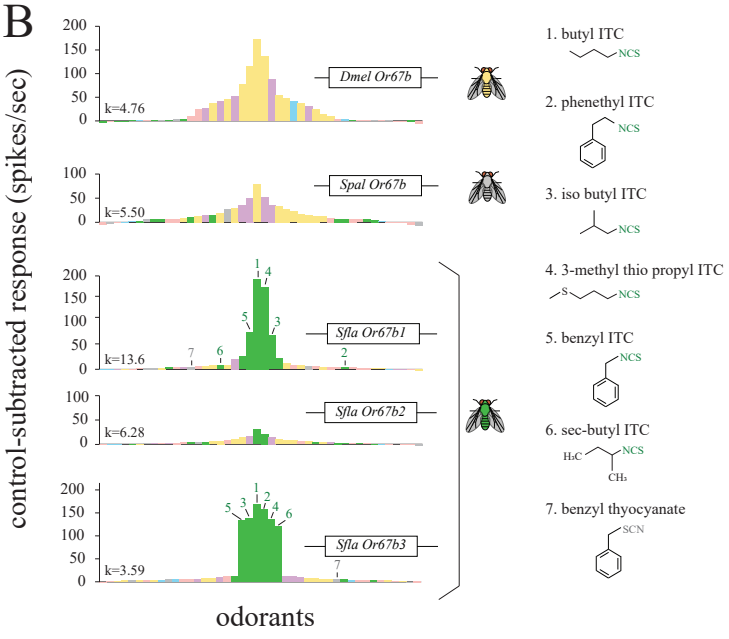
Figure 3



A



B



C

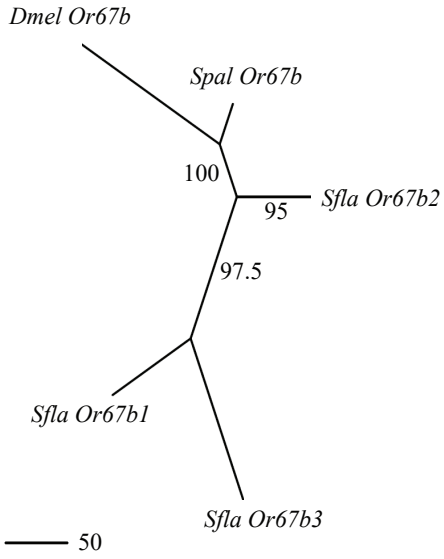
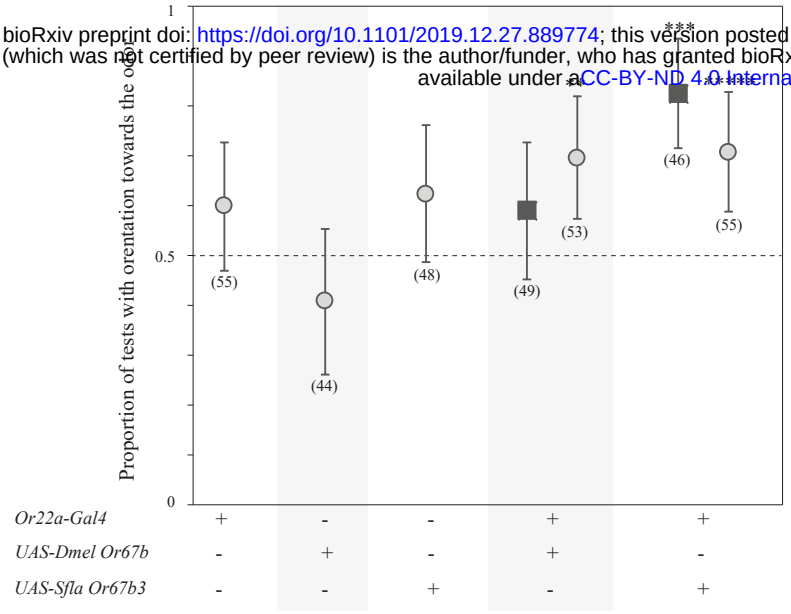
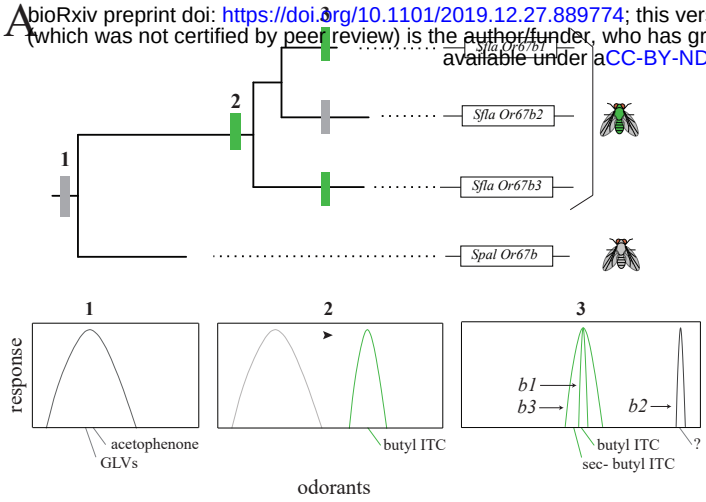


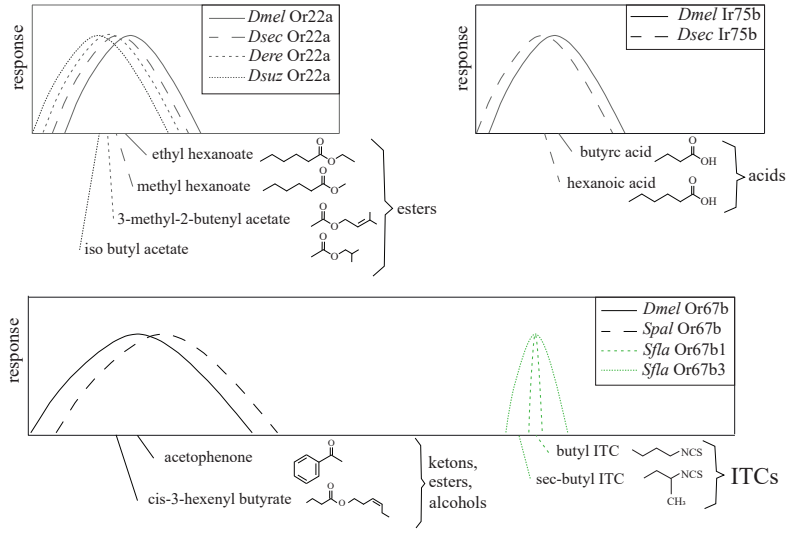
Figure 5



A



B



518 **Figure Legends:**

519

520 **Figure 1: Olfactory behavioral responses of *S. flava* and its microbe-feeding relatives *S.**

521 ***pallida* and *D. melanogaster* to mustard oil volatiles and ITCs.** (A) Schematic representation

522 of the dual-choice maze used to test the olfactory responses of flies. One arm of the maze was

523 connected to a 1-ml syringe containing a piece of filter paper loaded with an odorant (20 µl of a

524 1:100 vol/vol solution diluted in mineral oil) or apple cider vinegar (40 µl), while the other arm

525 was connected to a syringe containing a piece of filter paper loaded with the control stimulus (20

526 µl of mineral oil or 40 µl of water). In the case of arugula, 2-4 leaves were excised from a plant

527 just before tests and placed in a 5-ml syringe (the control syringe contained clean tissue paper).

528 In each test a group of non-starved flies (n=4) was anesthetized in ice during 50-60 seconds and

529 allowed to choose between the odor-laden or the odorless arm of the maze. Each test (maximum

530 duration=5 min) ended when the first insect (out of the 4 released) made a choice for one or the

531 other arm of the maze. (B) Olfactory behavioral responses of *D. melanogaster*, *S. pallida* and *S.*

532 *flava* to the odors and odorants indicated to the right. Data represent the proportion of tests in

533 which insects choose the odorous arm of the maze, and the error bars indicate the 95%

534 confidence interval of the proportion; the dotted reference line at 0.5 indicate no preference for

535 the odorous or odorless arm of the maze. Numbers between parentheses indicate the number of

536 tests with choices, either for the odorless or the odorous arm of the maze (on average, insects

537 made a choice in 82% of the tests). For each fly species and odor, data was analyzed using two-

538 tailed Binomial tests adjusted for multiple testing using FDR (adjusted p-values: * $p<0.05$, **

539 $p<0.01$, *** $p<0.005$, **** $p<0.001$). *S. flava* was attracted to mustard plant volatiles (arugula)

540 and ITCs and avoided apple cider vinegar volatiles, while *S. pallida* avoided butyl ITC. *D.*

541 *melanogaster* was attracted to vinegar but not to arugula volatiles or ITCs.

Figure 2: Maximum likelihood (ML) phylogeny of *Or67b*s in Drosophilidae. (A) 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*. Gene names with bold fonts include branches with estimated d*N*/d*S* >1 from branch tests conducted in PAML, consistent with episodic positive selection. (B) ML phylogeny reconstructed from coding sequence of *Or67b* orthologs of twelve *Drosophila* species, *S. pallida* 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 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* branches are indicated by fly icons colored as in Figure 1. Scale bars are in units of substitutions per site.

Figure 3: Odor responses of orthologous and paralogous Or67b from *D. melanogaster*, *S. pallida* and S. *flava* to mustard-plant volatiles.

(A) *Or67b* paralogs and orthologs were expressed in a *D. melanogaster* mutant that lacks its endogenous *Or* (*Or22a*) in the ab3A olfactory receptor neuron (30). Shown is a representative trace of a single sensillum recording obtained from a *D. melanogaster* fly with *Dmel Or67b* expressed in the ab3A neuron (*Or22a*$^{-/-}$ *Or22b*$^{-/-}$; *Or22a-Gal4/UAS-Dmel Or67b*) (top left). Recordings from the ab3 sensilla produced spike trains from the two olfactory receptor cells housed within this sensilla (ab3A, which expresses *Or67b*, and ab3B, which expresses the native *Or85b*) that had distinct spike amplitudes. The ab3A neuron (labeled A) produced spikes of larger amplitude than the ab3B neuron (labeled B). Scale bar: 10 mV. Bottom: representative electrophysiological recordings obtained from the ab3 sensilla of flies expressing *Dmel Or67b*, *Spal Or67b*, *Sfla Or67b1*, *Sfla Or67b2* or *Or67b3* in the ab3A neuron in response to stimulation with acetophenone, butyl ITC, sec-butyl ITC and mineral oil (solvent control). The bars below records indicate the onset and duration (1 sec) of the odor stimulation. Calibration bars: 10 mV. (B) Responses (net number of spikes/second; average ± SE, n= 6-8 sensilla obtained from 2-4 females) evoked by stimulation with apple cider vinegar, mustard plant volatiles and non-mustard plant volatiles (beet, control). Asterisks indicate significant differences between the responses to stimulation with an odor and the control stimulus (beet) (Mann-Whitney U tests corrected for multiple comparisons using FDR; adjusted p-values: * $p<0.05$; ** $p<0.01$; *** $p<0.005$). Note that neurons expressing *Dmel Or67b* and *Spal Or67b*, but not those expressing any of the *S. flava Or67b* paralogs, respond to apple cider vinegar. Conversely, only neurons expressing *S. flava Or67b* paralogs respond to mustard plant volatiles; neurons expressing *Sfla* Or67b3 responded to all mustard plant volatiles tested. (C) Responses (net number of spikes/sec;

588 average ± SE, n= 6-20 for each odorant, n=6-15 for mineral oil, obtained from 2-7 females)
589 evoked by stimulation with single odorants categorized as follows: *Dmel* Or67b activators
590 (Database of Odor Responses, (33); orange), ITCs (green), nitriles (blue), green leaf volatiles
591 (GLVs, purple), electrophiles/*TrpA1* activators (pink), other odorants not belonging to any of the
592 previous categories (gray), and the two solvent controls (black). The responses of the strongest
593 activator within each category (acetophenone, butyl ITC, mandelonitrile, cis-3-hexenyl butyrate,
594 trans-2-pentenal, and geranyl acetate) were compared against the mineral oil control. Asterisks
595 indicate significant differences between the responses to stimulation with each of these odorants
596 and the control stimulus (statistics as described in B). Note that ITCs strongly activate neurons
597 expressing *Sfla Or67b1* or *Sfla Or67b3*, and that for the most part these Ors do not respond to
598 any of the other odorants tested. Some ITCs, such as sec-butyl ITC and phenethyl ITC, only
599 activated neurons expressing *Sfla Or67b3*. *Spal* Or67b has an odorant response profile similar to
600 that of *Dmel* Or67b, responding mostly to stimulation with *D. melanogaster* activators and
601 GLVs.
602
603
604
605
606
607
608
609

Figure 4: Or67b orthologs and paralogs have distinct and non-overlapping odor-response profiles, tuning breadths, and functional relatedness. (A) Vectors quantifying the odor-evoked responses (control-subtracted) were projected onto a two-dimensional subspace determined by the first three principal components of the data-set (left: PC1 vs PC2; right: PC1 vs PC3). PC1-3 captured each 63.1%, 26.0%, and 9.0% of the variances of the original five-dimensional data. Each dot represents one odorant, and the different colors and symbols indicate different chemical categories (as in Figure 3C and in the inset, right graph). (B) Tuning curves of each Or67b copy, showing the distribution of mean responses (n= 6-20) to the 44 odorants tested (color-coded as in A). The odorants are displayed along the horizontal axis according to the responses (net spikes/sec) they elicit from each Or. The odorants eliciting the strongest responses for each Or are located at the center of the distribution, while weaker activators are distributed along the edges of the distribution. Thus, the order of the odorants is different for the five Ors. Note that the strongest responses (center of the distribution) from *Dmel* Or67b and *Spal* Or67b are evoked by *D. melanogaster* activators and GLVs, while the strongest responses from all *Sfla* Or67b paralogs are evoked by ITCs. The tuning breadth of each Or is quantified by the kurtosis value (k) of the distribution with higher values indicating narrower odor-response profiles. The chemical structure of the top six *Sfla* Or67b3 activators and benzyl thiocyanate (numbers in green) are indicated to the right. (C) Neighbor joining cluster analysis for the five Or67b orthologs based on the Euclidean distances between the odor-evoked responses of the receptors. *Dmel* Or67b and *Spal* Or67b are consistently separated from the *S. flava* odor response profiles, and *Sfla* Or67b1 and *Sfla* Or67b3 are consistently joined, having similar response profiles.

Figure 5: *S. flava Or67b3* confers olfactory sensitivity to ITCs

The responses of *D. melanogaster* flies expressing *Dmel Or67b* or *Sfla Or67b3* under the control of *Gal4* in the "empty" neuron (ab3A neuron) and three control parental lines were conducted as described in the caption to Figure 1. Genotypes are indicated at the bottom. One arm of the maze was connected to a 1-ml syringe containing a piece of filter paper loaded with 20 µl of 1:100 vol/vol (gray circles) or 1:1,000 vol/vol (black rectangles) of butyl ITC, while the other arm was connected to a syringe containing a piece of filter paper loaded with 20 µl of mineral oil. Data represent the proportion of tests in which insects choose the odorous arm of the maze; the deviation bars indicate the 95% confidence interval of the proportion. The dotted reference line at 0.5 indicate no preference for either arm of the maze; the numbers between parenthesis indicate the number of tests with choices for one or the other arm. Control flies (first three columns) showed no preference for either arm of the maze, even when tested with the highest concentration of butyl ITC (two-tailed Binomial tests, in all cases $p>0.1$). Note that flies carrying the *UAS-Sfla Or67b3* transgene under the control of *Gal4* in the ab3A neuron were attracted to both concentrations of butyl ITC (two-tailed Binomial tests, *** $p<0.005$; **** $p<0.001$). In contrast, flies carrying the *UAS-Dmel Or67b* transgene under the control of *Gal4* were attracted only to the highest concentration of the odorant (two-tailed Binomial tests, **$p<0.01$). At the lowest concentration, the response of flies carrying the *S. flava* transgene was different from the response of flies carrying the *D. melanogaster* transgene (Fisher Exact test, $p<0.05$). These results indicate that in these flies, *Sfla Or67b3* is sufficient to confer attraction to butyl ITC, a behavior that persists at low concentrations.

Figure 6: A model for the evolution of *Or67b* and comparison of tuning curves of orthologous Or22a, Ir75b and Or67b (A) The evolution of *Scaptomyza Or67b* begins with a shift in the tuning breadth and ligand specificity of an ancestral Or67b, from *Dmel* activators and GLVs to ITCs (neofunctionalization). Subsequent gene triplication of *Sfla Or67b* gave rise to three paralogous *Or67b* genes (*Sfla Or67b1, Sfla Or67b2,* and *Sfla Or67b3*). *Sfla* Or67b1 evolved to become more narrowly tuned, specializing in the detection of a few ITCs (subfunctionalization), while *Sfla* Or67b3 kept its ancestral, more broadly tuned state, responding to a broad range of ITCs. The ligands and tuning breadth of *Sfla* Or67b2 remain unresolved in this model. (B) The ligand specificities of orthologous Or22a, Ir75b, and Or67b. 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 the host-derived esters indicated (top left; ref. 16, 17, 42, 43). The Ir75b orthologs from *D. melanogaster* (*Dmel* Ir75b), and *D. sechellia* (*Dsec* Ir75b) are strongly activated by the two acids indicated (top right; ref. 15). *Dmel* Or67b and *Spal* Or67b are strongly activated by acetophenone, while *Sfla* Or67b copies are activated by ITCs only (bottom). Note that Or22a and Ir75b orthologs are both divergent but activated by ligands belonging to a single chemical class (esters and acids, respectively). In contrast, the ligands of orthologous 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 an entirely different chemical class (ITCs).

679 **References:**

680 1. G. Frankel, The raison d'etre of secondary plant substances. *Science* **129**, 1466–1470
681 (1959).

682 2. I. Winde, U. Wittstock, Insect herbivore counteradaptations to the plant glucosinolate–
683 myrosinase system. *Phytochemistry* **72**, 1566–1575 (2011).

684 3. A. D. Gloss, *et al.*, Evolution in an ancient detoxification pathway is coupled with a
685 transition to herbivory in the Drosophilidae. *Mol. Biol. Evol.* **31**, 2441–2456 (2014).

686 4. A. Kergunteuil, S. Dugravot, A. Mortreuil, A. Le Ralec, A. M. Cortesero, Selecting
687 volatiles to protect brassicaceous crops against the cabbage root fly, Delia radicum.
688 *Entomol. Exp. Appl.* **144**, 69–77 (2012).

689 5. B. Al-Anzi, W. D. Tracey Jr, S. Benzer, Response of Drosophila to wasabi is mediated by
690 painless, the fly homolog of mammalian TRPA1/ANKTM1. *Curr. Biol.* **16**, 1034–1040
691 (2006).

692 6. T. Ohta, T. Imagawa, S. Ito, Novel agonistic action of mustard oil on recombinant and
693 endogenous porcine transient receptor potential V1 (pTRPV1) channels. *Biochemical*
694 *Pharmacology* **73**, 1646–1656 (2007).

695 7. L. J. Macpherson, *et al.*, Noxious compounds activate TRPA1 ion channels through
696 covalent modification of cysteines. *Nature* **445**, 541–545 (2007).

697 8. K. Kang, *et al.*, Analysis of Drosophila TRPA1 reveals an ancient origin for human
698 chemical nociception. *Nature* **464**, 597–600 (2010).

699 9. O. M. Arenas, *et al.*, Activation of planarian TRPA1 by reactive oxygen species reveals a
700 conserved mechanism for animal nociception. *Nat. Neurosci.* **20**, 1686–1693 (2017).

701 10. A. Hinman, H.-H. Chuang, D. M. Bautista, D. Julius, TRP channel activation by reversible
702 covalent modification. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 19564–19568 (2006).

703 11. R. M. Joseph, J. R. Carlson, Drosophila chemoreceptors: A molecular interface between the
704 chemical world and the brain. *Trends in Genetics* **31**, 683–695 (2015).

705 12. C. Gomez-Diaz, F. Martin, J. M. Garcia-Fernandez, E. Alcorta, The two main olfactory
706 receptor families in Drosophila, ORs and IRs: A Comparative Approach. *Front. Cell.*
707 *Neurosci.* **12**, 253 (2018).

708 13. A. Sanchez-Gracia, F. G. Vieira, J. Rozas, "Molecular evolution of the major chemosensory
709 gene families in insects. Heredity (Edinb). 2009; 103 (3): 208-16" (Epub 2009/05/14. doi:
710 10.1038/hdy. 2009.55. PubMed PMID: 19436326).

711

14. P. Brand, S. R. Ramírez, The Evolutionary Dynamics of the Odorant Receptor Gene Family in Corbiculate Bees. *Genome Biol. Evol.* **9**, 2023–2036 (2017).

15. L. L. Prieto-Godino, *et al.*, Evolution of acid-sensing olfactory circuits in drosophilids. *Neuron* **93**, 661–676.e6 (2017).

16. S. Mansourian, *et al.*, Wild African Drosophila melanogaster are seasonal specialists on marula fruit. *Curr. Biol.* **28**, 3960–3968.e3 (2018).

17. J. Linz, *et al.*, Host plant-driven sensory specialization in Drosophila erecta. *Proc. Biol. Sci.* **280**, 20130626 (2013).

18. N. K. Whiteman, *et al.*, Mining the plant-herbivore interface with a leafmining Drosophila of Arabidopsis. *Mol. Ecol.* **20**, 995–1014 (2011).

19. B. Goldman-Huertas, *et al.*, 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 (2015).

20. C. E. Reisenman, Y. Lee, T. Gregory, P. G. Guerenstein, Effects of starvation on the olfactory responses of the blood-sucking bug Rhodnius prolixus. *J. Insect Physiol.* **59**, 717–721 (2013).

21. C. P. Faucher, M. Hilker, M. de Bruyne, Interactions of carbon dioxide and food odours in Drosophila: olfactory hedonics and sensory neuron properties. *PLoS One* **8**, e56361 (2013).

22. L. W. Aurand, J. A. Singleton, T. A. Bell, J. L. Etchells, Volatile components in the vapors of natural and distilled vinegars. *J. Food Sci.* **31**, 172–177 (1966).

23. F. J. Cousin, *et al.*, Microorganisms in fermented apple beverages: Current knowledge and future directions. *Microorganisms* **5** (2017).

24. A. D. Gloss, *et al.*, Evolution of herbivory remodels a Drosophila genome. *bioRxiv*, 767160 (2019).

25. C. S. McBride, J. R. Arguello, B. C. O'Meara, Five Drosophila genomes reveal nonneutral evolution and the signature of host specialization in the chemoreceptor superfamily. *Genetics* **177**, 1395–1416 (2007).

26. Z. Yang, PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).

27. T. S. Ha, D. P. Smith, A pheromone receptor mediates 11-cis-vaccenyl acetate-induced responses in Drosophila. *J. Neurosci.* **26**, 8727–8733 (2006).

28. D. J. Hoare, *et al.*, Modeling peripheral olfactory coding in Drosophila larvae. *PLoS One* **6**, e22996 (2011).

29. S. G. Chin, S. E. Maguire, P. Huoviala, G. S. X. E. Jefferis, C. J. Potter, Olfactory neurons and brain centers directing oviposition decisions in Drosophila. *Cell Rep.* **24**, 1667–1678 (2018).

30. E. A. Hallem, J. R. Carlson, Coding of odors by a receptor repertoire. *Cell* **125**, 143–160 (2006).

31. F. Gonzalez, P. Witzgall, W. B. Walker, Protocol for heterologous expression of insect odourant receptors in Drosophila. *Front. Ecol. Evol.* **4**, 189 (2016).

32. A. A. Agrawal, N. S. Kurashige, A role for isothiocyanates in plant resistance against the specialist herbivore Pieris rapae. *J. Chem. Ecol.* **29**, 1403–1415 (2003).

33. D. Münch, C. G. Galizia, DoOR 2.0--Comprehensive mapping of Drosophila melanogaster odorant responses. *Sci. Rep.* **6**, 21841 (2016).

34. S. F. Vaughn, M. A. Berhow, Glucosinolate hydrolysis products from various plant sources: pH effects, isolation, and purification. *Industrial Crops and Products* **21**, 193–202 (2005).

35. P. R. Ehrlich, P. H. Raven, Butterflies and plants: A study in coevolution. *Evolution* **18**, 586–608 (1964).

36. J. A. Depree, T. M. Howard, G. P. Savage, Flavour and pharmaceutical properties of the volatile sulphur compounds of Wasabi (Wasabia japonica). *Food Res. Int.* **31**, 329–337 (1998).

37. E. A. Bernays, Evolution of Feeding Behavior in Insect Herbivores: Success seen as different ways to eat without being eaten. *Bioscience* **48**, 35–44 (1998).

38. S. Ohno, 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 (1970).

39. H. M. Heidel-Fischer, *et al.*, An insect counteradaptation against host plant defenses evolved through concerted neofunctionalization. *Mol. Biol. Evol.* **36**, 930–941 (2019).

40. X. He, J. Zhang, Rapid subfunctionalization accompanied by prolonged and substantial neofunctionalization in duplicate gene evolution. *Genetics* **169**, 1157–1164 (2005).

41. R. Assis, D. Bachtrog, Neofunctionalization of young duplicate genes in Drosophila. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 17409–17414 (2013).

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

43. I. Keesey, *et al.*, Evolution of a pest: towards the complete neuroethology of Drosophila suzukii and the subgenus Sophophora. *BioRxiv* (2019).

778 44. J. A. A. Renwick, M. Haribal, S. Gouinguené, E. Städler, Isothiocyanates stimulating
779 oviposition by the diamondback moth, Plutella xylostella. *J. Chem. Ecol.* **32**, 755–766
780 (2006).

781 45. C. Tait, S. Batra, S. S. Ramaswamy, J. L. Feder, S. B. Olsson, Sensory specificity and
782 speciation: a potential neuronal pathway for host fruit odour discrimination in Rhagoletis
783 pomonella. *Proc. Biol. Sci.* **283** (2016).

784 46. J. J. Wiens, R. T. Lapoint, N. K. Whiteman, Herbivory increases diversification across
785 insect clades. *Nat. Commun.* **6**, 8370 (2015).

786 47. C. Mitter, B. Farrell, B. Wiegmann, The phylogenetic study of adaptive zones: Has
787 phytophagy promoted insect diversification? *Am. Nat.* **132**, 107–128 (1988).

788 48. J. H. Zar, *Biostatistical Analysis* (Prentice Hall, 1999).

789 49. Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: A practical and powerful
790 approach to multiple testing. *Journal of the Royal Statistical Society: Series B*
791 *(Methodological)* **57**, 289–300 (1995).

792 50. J. Thurmond, *et al.*, FlyBase 2.0: the next generation. *Nucleic Acids Res.* **47**, D759–D765
793 (2019).

794 51. K. Katoh, MAFFT: a novel method for rapid multiple sequence alignment based on fast
795 Fourier transform. *Nucleic Acids Research* **30**, 3059–3066 (2002).

796 52. L.-T. Nguyen, H. A. Schmidt, A. von Haeseler, B. Q. Minh, IQ-TREE: A fast and effective
797 stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology*
798 *and Evolution* **32**, 268–274 (2015).

799 53. A. Stamatakis, RAxML version 8: a tool for phylogenetic analysis and post-analysis of
800 large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).

801 54. F. Ronquist, *et al.*, MrBayes 3.2: efficient Bayesian phylogenetic inference and model
802 choice across a large model space. *Syst. Biol.* **61**, 539–542 (2012).

803 55. S. Guindon, *et al.*, New Algorithms and methods to estimate maximum-likelihood
804 phylogenies: Assessing the performance of PhyML 3.0. *Systematic Biology* **59**, 307–321
805 (2010).

806 56. R. C. Team, R: a language and environment for statistical computing [online]. R Foundation
807 for Statistical Computing, Vienna, Austria (2018).

808 57. E. Paradis, K. Schliep, ape 5.0: an environment for modern phylogenetics and evolutionary
809 analyses in R. *Bioinformatics* **35**, 526–528 (2019).

810 58. E. A. Hallem, J. R. Carlson, The odor coding system of Drosophila. *Trends in Genetics* **20**,
811 453–459 (2004).

59. R. T. Lapoint, P. M. O'Grady, N. K. Whiteman, Diversification and dispersal of the Hawaiian Drosophilidae: the evolution of Scaptomyza. *Mol. Phylogenet. Evol.* **69**, 95–108 (2013).

60. J. Bischof, R. K. Maeda, M. Hediger, F. Karch, K. Basler, An optimized transgenesis system for Drosophila using germ-line-specific φC31 integrases. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 3312–3317 (2007).

Table S1: Parameters from Odorant Receptor molecular evolution analyses

Test*	Branch†	background dN/dS	foreground dN/dS	dN – dS	FDR adjusted p-value
OR	Or2a	0.10948	0.08368	-1.1523	0.4031933
OR	Or10a	0.16074	0.08826	-0.5837	**0.03599091**
OR	Or13a1	0.13815	0.95538	-0.0014	**8.36204E-06**
OR	Or13a2	0.14485	0.29386	-0.0114	0.5374215
OR	Or13a ancestor	0.1405	0.22696	-0.2232	0.1062494
OR	Or19a	0.27184	0.256	-0.3699	0.8337806
OR	Or22c	0.15572	0.25996	-0.1496	0.1475187
OR	Or23a	0.16331	0.2829	-0.2438	0.0934421
OR	Or24a	0.08324	0.09358	-0.5121	0.7036877
OR	Or30a	0.07027	0.10685	-0.4864	0.1396827
OR	Or33c	0.20981	0.15322	-0.402	0.2748754
OR	Or35a	0.10058	0.80581	-0.2963	**0.00245945**
OR	Or42a1	0.10806	0.11791	-0.336	0.7716366
OR	Or42a2	0.10578	0.80581	-0.012	**0.000970903**
OR	Or42a3	0.10571	0.80581	-0.0045	**0.000643469**
OR	Or42a2 and 3 ancestor	0.11046	0.08409	-0.5505	0.3637872
OR	Or43a	0.14775	0.08518	-0.3217	0.1557934
OR	Or45b	0.10067	0.10449	-0.3992	0.8803498
OR	Or46a	0.14313	0.20832	-0.2632	0.1151226
OR	Or47b	0.16957	0.15781	-0.3375	0.8025064
OR	Or49a	0.19774	0.17914	-0.5377	0.7036877
OR	Or49b	0.10784	0.05836	-0.3747	0.09592573
OR	Or56a	0.12598	0.05652	-0.5602	**0.004298517**
OR	Or59a	0.1085	0.31529	-0.2475	**0.000415309**
OR	Or59b	0.09936	0.23227	-0.4421	**0.000415309**
OR	Or63a1	0.12346	∞	**0.0172**	**3.51892E-05**
OR	Or63a2	0.12136	∞	**0.0274**	**1.35815E-09**
OR	Or63a ancestor	0.10903	0.56602	-0.1212	**5.0862E-13**
OR	Or67a1	0.20346	0.13116	-0.528	0.0934421
OR	Or67a2	0.2067	**0.06369**	-0.0693	**0.0004729**
OR	Or67b1	0.11719	0.87005	-0.0059	**2.5914E-07**

OR	Or67b2	0.11479	0.69462	-0.0253	**1.96849E-09**
OR	Or67b3	0.10844	**1.95161**	**0.0414**	**9.78045E-20**
OR	Or67b1 and 2 ancestor	0.11932	∞	**0.0282**	**3.90702E-08**
OR	Or67b1, 2 and 3 ancestor	0.11679	0.28275	-0.2591	**0.000585544**
OR	Or67c	0.05491	0.07255	-0.4886	0.4029281
OR	Or67d	0.12277	**7.22359**	**0.0948**	**2.56153E-05**
OR	Or69a	0.21665	0.12519	-0.4274	0.05279388
OR	Or71a	0.15846	0.25285	-0.3464	0.1151226
OR	Or74a1	0.13897	0.16825	-0.3766	0.5728869
OR	Or74a2	0.13723	0.18825	-0.4375	0.2932306
OR	Or82a	0.12569	0.18201	-0.2805	0.1926929
OR	Or83a	0.0967	0.13983	-0.4518	0.148133
OR	Or83c1	0.16267	0.21756	-0.4381	0.3145556
OR	Or83c2	0.22615	0.41682	-0.1561	0.0934421
OR	Or83c3	0.22488	0.39035	-0.2043	0.0934421
OR	Or83c2 and 3 ancestor	0.25561	0.17609	-0.2679	0.5410598
OR	Or85aLike	0.22488	0.39035	-1.8882	**0.001826812**
OR	Or85bc	0.19387	**0.07261**	-0.32	0.2783313
OR	Or85e	0.15034	0.20391	-0.3561	0.6592839
OR	Or85f	0.11206	0.13076	-0.4139	0.08611496
OR	Or88a	0.18031	0.09816	-0.2848	**0.003678095**
OR	Or92a	0.14147	0.32458	-0.3894	**3.51892E-05**
OR	Or94a	0.044	0.13816	-0.4616	0.1933
OR	Or94b	0.09284	0.14353	-0.5473	0.09592573
OR	Or98a1	0.09893	0.17168	**0.0013**	0.1796849
OR	Or98a3	0.18051	1.69406	-0.0093	0.6770641
OR	Or98a7	0.18219	0.28105	**0.0258**	**5.47517E-05**
OR	Or98a1 and 3 ancestor	0.14909	**2.74217**	-0.0113	**0.00014207**
OR	Or98a1, 3 and 7 ancestor	0.15018	0.81658	-0.6623	0.7716366
OR	Or98a2	0.18749	0.17369	-0.024	0.7716366
OR	Or98a6	0.24649	0.30314	-0.044	0.6592839
OR	Or98a2 and 6 ancestor	0.24491	0.30931	-0.33	0.0934421
OR	Or98a4	0.2739	0.14173	**0.0063**	**0.000676304**
OR	Or98a5	0.22865	**1.23866**	**0.01**	0.05279388
OR	Or98a4 and 5 ancestor	0.237	4.11669	-0.4823	**0.01935304**

OR	Or98b1	0.28857	**0.13007**	-0.8492	0.7613693
OR	Or98bLike	0.18605	0.20625	-0.8656	0.5728869
OR	OrN1	0.19137	0.15525	-0.3194	0.1960608
OR	OrN2 copy 1	0.02988	0.07861	-0.0006	0.1278416
OR	OrN2 copy 4	0.14809	0.21155	**0.0008**	0.05293525
OR	OrN2 copy 5	0.2495	0.94941	-0.6183	0.08611496
OR	OrN2 copy 6	0.24826	1.05198	-0.8924	0.1151226
OR	OrN2 copy 1 and 4 ancestor	0.26274	0.16626	-0.4861	0.5728869
OR	Orco	0.26137	0.17749	-0.2037	**0.02049901**
OR67b	*D. grimshawi*	0.24934	0.29184	-0.3355	0.7873258
OR67b	*S. flava* copy 3	0.1041	**1.803**	**0.0375**	**1.44394E-21**
OR67b	*S. flava* copy 2	0.10935	**0.7393**	-0.0208	**1.05769E-10**
OR67b	*S. flava* copy 1	0.11251	**0.62856**	-0.022	**1.5606E-06**
OR67b	*S. pallida*	0.1236	0.0506	-0.3076	**0.01604471**
OR67b	*D. mojavensis*	0.1137	**0.2053**	-0.8484	**0.009320281**
OR67b	*D. virilis*	0.119	0.1185	-1.4528	0.9933245
OR67b	*D. melanogaster*	0.11971	0.04714	-0.0423	0.4999046
OR67b	*D. sechellia*	0.1193	0.0796	-0.0368	0.6982622
OR67b	*D. simulans*	0.1202	0.0197	-0.0523	**0.04609565**
OR67b	*D. yakuba*	0.1198	0.0914	-0.1247	0.614736
OR67b	*D. erecta*	0.1193	0.1051	-0.1126	0.7873258
OR67b	*D. ananassae*	0.1329	0.0303	-1.7359	**5.2654E-10**
OR67b	*D. persimilis*	0.1188	0.1803	-0.0103	0.7873258
OR67b	*D. pseudoobscura*	0.119	0.1149	-0.0078	0.9933245
OR67b	*D. willistoni*	0.1288	0.0427	-1.4866	**1.97605E-05**
OR67b	Subgenus *Drosophila* Ancestor	0.1236	0.0564	-0.9778	**0.01967709**
OR67b	*Sophophora* minus D. *willistoni*	0.1212	0.0711	-0.5393	0.3034497
OR67b	virilis-repleta radiation	0.1196	0.0905	-0.1587	0.7873258
OR67b	Hawaiian radiation ancestor	0.1198	0.0979	-0.2595	0.7817405
OR67b	*Scaptomyza* genus ancestor	0.1221	0.0435	-0.2512	**0.04498955**
OR67b	*S. flava* copy 1, 2 and 3 ancestor	0.1094	**0.4081**	-0.1282	**4.06804E-07**
OR67b	*S. flava* copy 1 and 2 ancestor	0.1135	∞	**0.0276**	**4.5389E-09**
OR67b	*D. simulans* and *D. sechelia* ancestor	0.1205	0.0001	-0.0564	**0.01202915**
OR67b	*D. yakuba* and *D. erecta* ancestor	0.1221	0.0196	-0.1444	**0.00128517**
OR67b	melanogaster subgroup minus Dmel	0.1201	0.048	-0.0727	0.2412147

OR67b	melanogaster subgroup ancestor	0.125	0.0446	-0.5921	**0.003348616**
OR67b	melanogaster group ancestor	0.1207	0.0634	-0.2523	0.4817083
OR67b	pseudoobscura subgroup ancestor	0.1291	0.0342	-0.9693	**2.39766E-06**

Table S1: Parameters for background *versus* foreground branch tests in PAML. * OR = Branch tests for differences in dN/dS rates among *Scaptomyza flava* copies of odorant receptor genes. OR67b = Branch tests for differences in dN/dS rates among all individual branches in a phylogeny of *Or67b* homologs in *Scaptomyza* and *Drosophila* spp. † Terminal branches are indicated with gene names. Ancestral branches are indicated with the term "ancestor" and a list of the terminal branches or the clade subtended.

TableS2 PCR Primers

Primer Name	Sequence	Primer role
67b12Fb	cATGAAGGACTTATTGGATCTGGAGCTAG	CDS amplification
67b1Rb	agcgcaTTATTTGTCTCTCATATTGCTC	CDS amplification
67b2Rb	agcgcaTTATTCATCTCTCATGTTACGC	CDS amplification
67b3Fb	cATGAACTTATTGGATATGGAGCTAG	CDS amplification
67b3Rb	agcgcaTTATTTGTTAGATAAATTACGC	CDS amplification
Sp67bFb	cATGAAAAACCTATTGGACCGGAAG	CDS amplification
Sp67bRb	aaagtacaTCATTTGTTATTCATATTACGCAAAAG	CDS amplification
b1EcoRIF	GACgaattcATGAAGGACTTATTGG	RE cut site addition
b1KpnIR	GACggtaccTTATTTGTCTCTCATATTG	RE cut site addition
b2KpnIR	GACggtaccTTATTCATCTCTCATG	RE cut site addition
b3KpnIR	GACggtaccTTATTTGTTAGATAAATTACG	RE cut site addition
PalEcoRIF	GACgaattcATGAAAAACCTATTGGACC	RE cut site addition
PalKpnIR	GACggtaccTCATTTGTTATTCATATTACG	RE cut site addition

Tables S2: 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.

Table S3: Net number of spikes/second; average ± SE

	"Empty neuron"		WT Or22a	
Odors	**spikes**	**n**	**spikes**	**n**
ethyl hexanoate	0	7	118±7.7	7
mineral oil	0	7	6.33±1.7	6
acetophenone	0	6	NA	NA
butyl ITC	0	6	NA	NA
mandelonitrile	0	6	NA	NA
cis-3-hexenyl butyrate	0	6	NA	NA
trans-2-pentenal	0	6	NA	NA
geranyl acetate	0	6	NA	NA
mineral oil	0	7	NA	NA

	Dmel67b		Spal67b		Sfla67b1		Sfla67b2		Sfla67b3	
Odors	**spikes**	**n**	**spikes**	**n**	**spikes**	**n**	**spikes**	**n**	**spikes**	**n**
apple cider vinegar	96.3±8.0	6	62±10.4	7	-4.33±2.0	6	6.67±4.6	6	15±2.1	6
arugula	19.7±3.0	6	14±3.4	7	64±7.8	6	24±8.7	6	22.3±2.3	6
wasabi	2.67±2.0	6	2.67±1.11	6	15.3±1.9	8	16.2±2.5	8	13.7±4.4	6
horseradish	3.67±1.8	6	-0.33±1.0	6	8.75±1.3	8	12.9±2.0	8	31.7±4.4	6
turnip	1±1.8	6	0.67±0.7	6	11.8±1.6	8	22±2.7	8	34.3±1.4	6
daikon	2±3.2	6	-0.67±0.8	6	14±2.4	8	17.1±3.4	8	19±3.4	6
beet	8±3.8	6	-0.33±1.1	6	2±1.30	6	-0.33±1.7	6	4.33±2.2	6
Dmel67b activators										
acetophenone	176±18	7	78±2.4	20	3.27±1.2	11	4.5±2.0	8	1.57±1.3	14
cis-3-hexen-1-ol	56.9±20	7	22±4.1	7	8.67±2.0	6	6.75±1.3	8	2.67±2.3	6
4-methylcyclohexenol	80.7±8.7	6	33±4.4	13	8.89±4.3	9	1.56±1.4	9	2.67±1.2	6
phenethyl alcohol	140±19	8	26.3±4.5	6	2.67±1.2	6	1±1.2	6	3.67±1.5	6
1-hexanol	118±13	7	19.142857	7	13.7±3.8	6	10.4±4.6	9	8±1.9	6
ethyl (S)-(+)-3-hydroxybutanoate	15.7±3.3	7	12±2.8	7	4.33±4.4	6	10.4±2.9	9	8±2	6
2-hexanol	12.3±5.2	8	11.7±2.8	7	7.33±2.1	6	9.5±1.8	8	7±3.1	6
benzyl alcohol	48.6±2.9	6	14.3±3.7	7	-1.33±0.8	6	6.75±1.8	8	5.67±1.4	6

2-hexanone	40±16	7	7.14±2.0	7	9.33±4.8	6	7.5±2.0	8	7.33±3.5	6
hexyl-formate	39.3±8.7	6	16.3±5.6	7	7±2.0	6	8.89±4.5	9	3±0.9	6
phenethyl aldehyde	51.4±10	7	29.4±3.8	7	2±2.3	6	2.86±3.0	7	1±0.7	6
2-heptanone	28±4.1	7	13.1±2.3	7	6.33±2.4	6	15.3±2.5	9	6±1.9	7
ITCs										
butyl ITC	4.57±1.6	7	15±4.2	6	196±15	7	22.7±2.8	6	171±11	7
ethyl ITC	0.33±1.5	6	2.33±0.95	6	24±2.7	6	3.33±3.9	6	13±1.7	6
isobutyl ITC	1.67±2.4	6	5.33±2.6	6	74±12	7	4.33±1.8	6	139±17	7
sec-butyl ITC	4.29±1.0	7	5±1.8	6	9.14±2.7	7	5.33±2.3	6	121±17	8
allyl ITC	3.43±0.8	7	9.75±4.2	8	28.9±5.3	7	6±2.4	7	4.75±3.7	8
3-methylthio propyl ITC	2±1.8	6	6.33±2.2	6	179±9.2	6	33±6.0	6	137±17	8
benzyl ITC	2.29±1.4	7	3.75±1.9	8	81.7±14.4	7	3.14±1.3	7	134±14	8
phenethyl ITC	6.57±2.2	7	5.25±2.8	8	4.8±1.3	10	1.14±0.86	7	158±15	8
3-methoxybenzyl ITC	-0.67±1.1	6	7±3.3	6	4.57±2.2	7	1.67±0.80	6	3.14±2.1	7
Nitriles										
mandelonitrile	45.3±11	6	1±1	6	1.33±1.8	6	-0.29±0.7	7	4±3.0	6
adiponitrile	2.67±2.5	6	0.33±0.33	6	0.33±0.3	6	0.67±2.2	6	4.33±1.8	6
propionitrile	5.33±1.6	6	1.33±0.4	6	2.3±2.0	6	1.33±1.1	6	4.33±2.5	6
GLVs										
cis-3-hexenyl butyrate	91±21	7	49.3±8.0	6	5±2.6	8	2±1.8	6	11.8±2.5	8
trans-2-hexen-1-ol	33±6.0	6	51±11	6	23.5±5.6	8	3±2.5	6	6.29±2.2	7
trans-2-hexenyl acetate	55.7±19	7	34±6.8	6	2±2.1	6	17.3±2.5	6	10.3±1.9	8
trans-2-hexenyl butyrate	29.7±6.7	6	35.6±8.2	6	1.71±1.0	7	4±1.5	6	12±4.3	7
cis-3-hexenyl acetate	43.3±9.3	6	13.7±2.6	6	1.67±1.7	6	16±5.3	8	11±5.8	6
TrpA1 activators/ electrophiles										
trans-2-pentenal	27.1±20	7	6.33±0.8	6	4.33±2.9	6	4±1.4	6	4±1.3	6
citronellal	1.43±1.6	7	7±2.3	6	4.57±1.6	7	9.14±3.5	7	12±2.8	11
cinnamaldehyde	12.3±2.8	7	9±3.8	6	9.25±2.7	8	3±1	6	6±1.7	12

acrorelin	2.67±2.6	6	6.33±2.2	6	6.67±1.9	6	7±4.3	6	3.33±1.3	6
iodoacetamide	1.43±1.4	7	6±1.9	6	7.33±1.6	6	6±2.4	6	4±1.7	6
formaldehyde	3.14±2.0	7	5.33±1.8	6	6±2.9	6	3±1.4	6	2±1.0	6
N-methylmaleimide	3.71±1.0	7	8.67±2.6	6	7.33±1.8	6	10±1.9	6	3±1.3	6
N-hydroxysuccinimide	1.71±1.3	7	4.67±2.5	6	4.67±2.6	6	8.67±2.3	6	2±0.7	6
methyl salicylate	-3±1.8	7	0±0.5	6	3±3	6	-0.57±1.8	7	5.33±4.2	6
S-allyl 2-propene-1-sulfinothioate	5.66±1.0	7	-0.33±0.8	6	3.67±1.5	6	0.57±1.1	7	3±1.5	6
diallyl disulfide	3.33±3.3	7	0.67±0.8	6	1.33±0.7	6	1.14±0.6	7	3.67±2.1	6
Others										
geranyl acetate	5.67±1.0	6	26±6.5	6	8.5±2.4	8	4.67±2.0	6	5.25±1.4	8
limonene	2.29±2.4	7	5±4.1	6	4.33±1.6	6	1±0.7	6	2.71±1.5	14
indol-3-carbinol	4±1.5	6	3.33±2.6	6	5.67±1.6	6	8.67±1.1	6	3.33±0.8	6
benzyl thyocyanate	2.57±1.4	7	7.67±1.4	6	8.67±2.4	6	5±1.3	6	8.67±2.8	6
Solvent control										
DMSO	2.86±1.6	7	10.7±3.2	6	3.33±1.0	6	7.67±1.4	6	4.33±2.9	6
mineral oil	3.73±1.2	15	0.91±0.87	11	1±1.4	6	1±1.8	6	0±0.38	11

	Sfla67b1		**Sfla67b3**	
odorants	**spikes**	**n**	**spikes**	**n**
3-methylthio propyl ITC				
1/100	114±18	8	127±13	6
1/1,000	64±11	8	27.7±7.1	6
1/10,000	27.9±4.6	8	11.7±5.3	6
1/100,000	12±3.3	9	10±4.0	6
butyl ITC				
1/100	147±19	10	161±16	6
1/1,000	46.8±6.6	9	50.7±8.1	6
1/10,000	25.6±7.3	8	19.7±6.2	6
1/100,000	14.25±5	9	7.33±3.4	6
isobutyl ITC				

1/100	40.6±9.0	8	131±14	6
1/1,000	6.3±2.3	8	31.7±5.6	6
1/10,000	5.63±2.7	8	9±3.8	6
1/100,000	8.25±1.8	9	6±5.2	6
sec-butyl ITC				
1/100	5±1.7	9	124±13	6
1/1,000	4.5±1.6	9	26.3±4.2	6
1/10,000	5.63±1.8	8	13.7±4.6	6
1/100,000	10.5±3.5	8	9±4.5	6
benzyl ITC				
1/100	94.5±14.1	10	123±13	6
1/1,000	16.5±5.1	9	27.7±9.4	6
1/10,000	8.625±3.6	8	8±1.4	6
1/100,000	10.5±3.2	9	13±5.0	6
phenethyl ITC				
1/100	5.25±2.8	9	147±17	6
1/1,000	8.5±3.0	8	74±18	6
1/10,000	8.71±4.1	7	26.3±7.4	6
1/100,000	8±2.9	7	9.33±5.2	6
mineral oil	1±1.4	6	0±0.4	11

Tables S3: Responses (net number of spikes/second; average ± SE) of the ab3 sensilla of flies expressing either no Or22a ("Empty neuron"), wild-type *Or22a* (Canton-S), *Dmel Or67b, Spal Or67b, Sfla Or67b1, Sfla Or67b2 or Or67b3* to stimulation with the odors and number of recordings (n) are denoted.

Figure S1

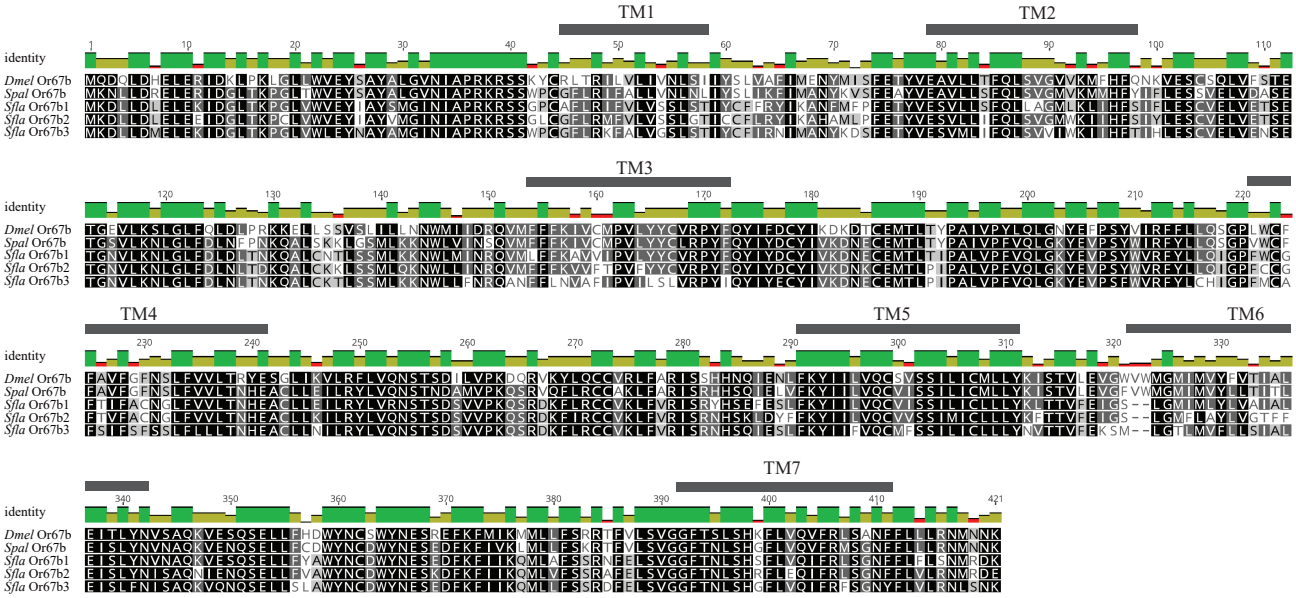


Figure S2

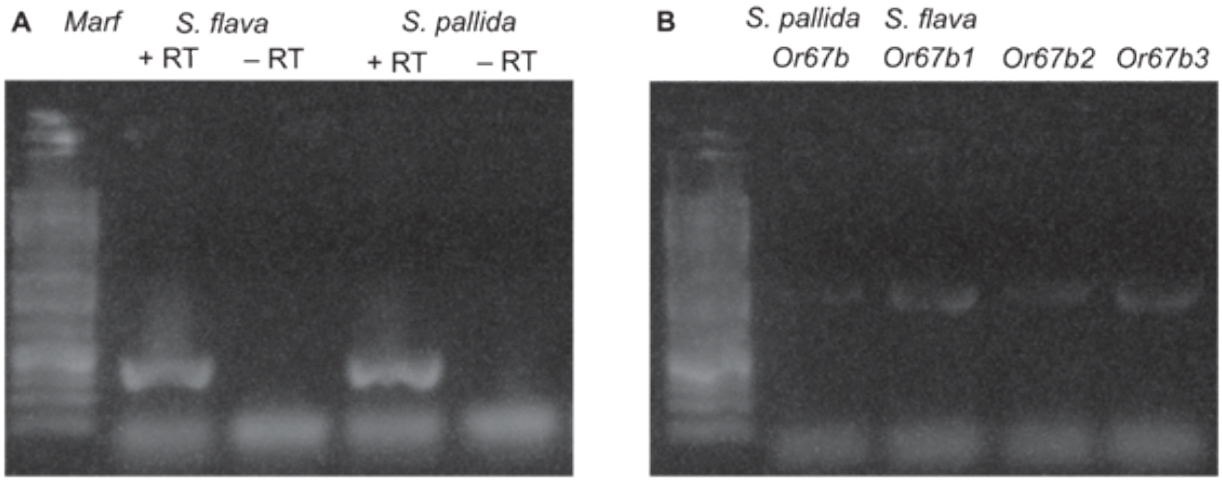
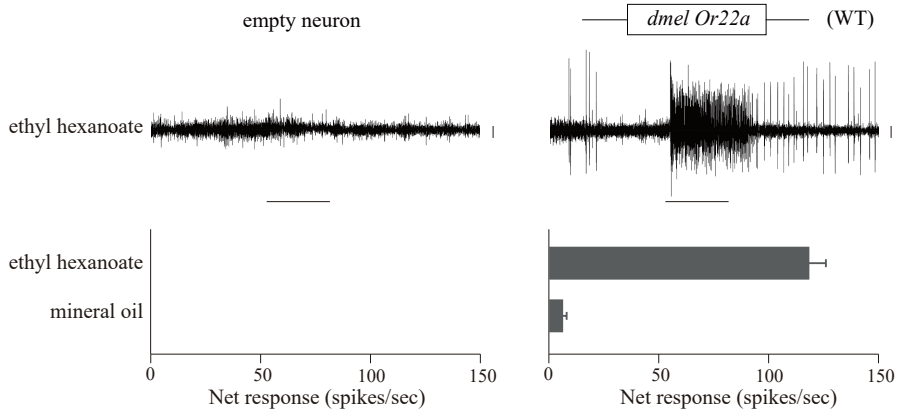
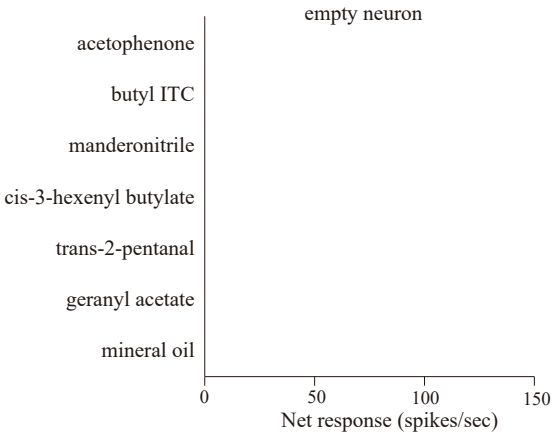


Figure S3

A



B



C

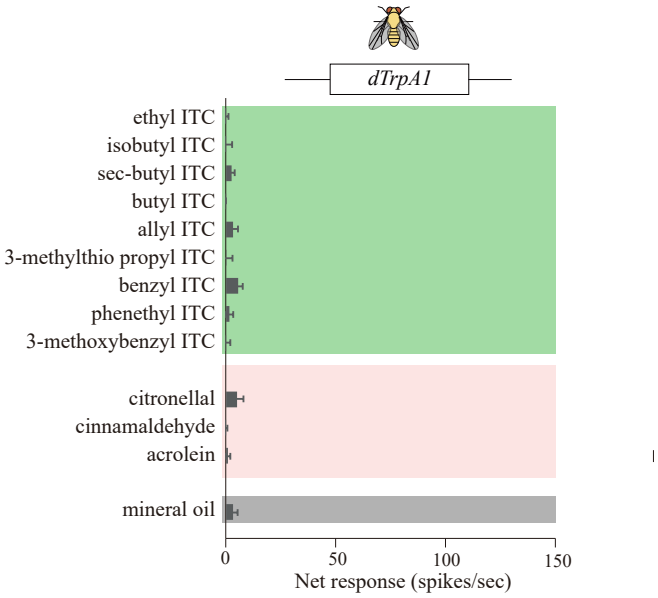


Figure S4

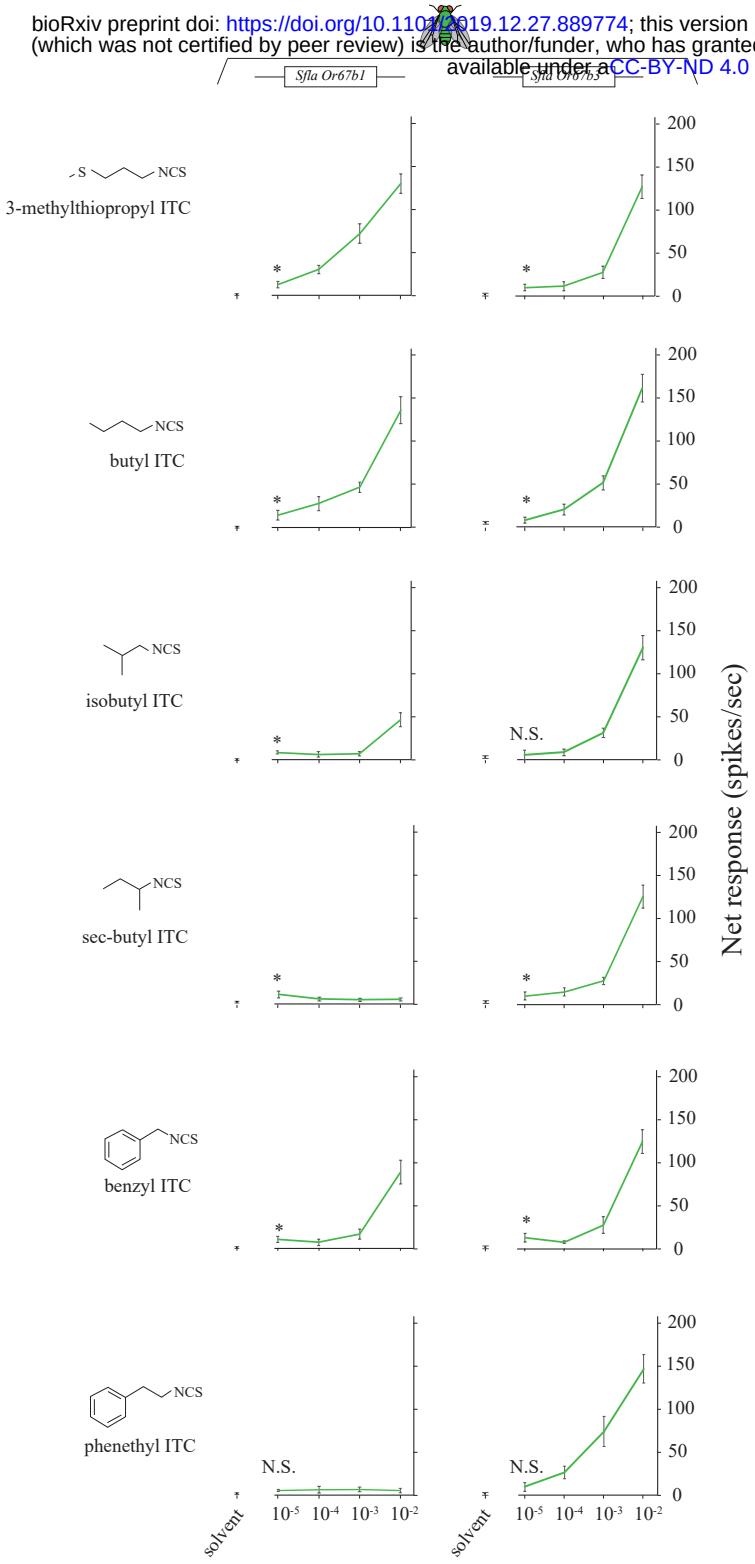
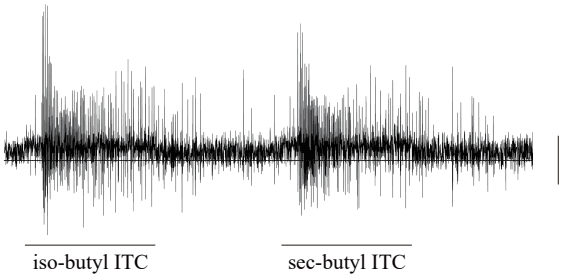


Figure S5



833 **Supplemental Figure Legends:**

834

835 **Supplemental Figure 1: Protein alignment of *D. melanogaster*, *S. pallida* and *S. flava***

836 **Or67bs.**

837 Note the high number of identical amino acids across orthologs/paralogs (black squares). Also

838 shown are each of the seven predicted transmembrane domains (TM1-7). Darker colors illustrate

839 higher degrees of sequence similarity, and lighter colors denote residues with high variability in

840 sequence across paralogs and orthologs.

841

842 **Supplemental Figure 2: *Or67bs* are expressed in *Scaptomyza* spp.**

843 (A) Amplification of *Marf* from cDNA generated from whole body extracts of *S. flava* and *S.*

844 *pallida* adults (+RT). As expected, *Marf* does not amplify in templates treated with DNaseI

845 without reverse transcriptase (– RT). (B) Amplification of *Or67b* genes from *S. pallida* and *S.*

846 *flava* whole adult cDNA, indicative of *in vivo* transcription of *Or67b* genes in adult *Scaptomyza*.

847 Ladder is Gene Ruler™ 1 kb plus (ThermoFisher, USA).

848

849 **Supplemental Figure 3:**

850 (A) Lack of *Or22a* expression in the ab3A neuron in the M2-MD fly line ("empty neuron"

851 mutant) was verified by lack of electrophysiological responses to ethyl hexanoate (left). In

852 contrast, in wild-type flies (Canton-S), ab3A neurons respond strongly to simulation with this

853 odor compound (right). Data indicates the net number of spikes/sec (average ± SE, n= 6-7 for

each odor and control stimulation, obtained from 3 females). (B) Single odorants used for statistical analysis in Fig. 3 were tested on the mutant ab3A neuron lacking its endogenous olfactory receptor. Stimulation with any of the odorants failed to evoke spiking activity. (C) *dTrpA1* was expressed in the ab3A neuron in the M2-MD fly line. None of the volatile ITCs or electrophiles elicited spiking activity. For a list of chemical categories see Fig. 3.

Supplemental Figure 4: Odorant dose–responses of *Sfla* Or67b1 and *Sfla* Or67b3

Responses (net spikes/sec, average +/- SE, n=6, obtained from 2 animals) of *Sfla* Or67b1 and *Sfla* Or67b3 to stimulation with increasing concentrations (vol/vol) of six different ITCs. The responses of *Sfla* Or67b3 steadily increased with increasing the concentration of all ITCs tested. *Sfla* Or67b1 and *Sfla* Or67b3 showed similar dose-response relationships in response to stimulation with 3-methyl-thio-propyl-ITC, butyl ITC, and benzyl ITC. Stimulation with sec-butyl ITC and phenethyl ITC produced only a very slight increase in spike activity (<10 spikes/second) from *Sfla* Or67b1 at any of the concentrations tested. The asterisks indicate significant differences between the responses to the 1:10,000 vol/vol concentration and the responses to the mineral oil control (Mann-Whitney U tests corrected for multiple testing using FDR; adjusted p-values: * $p<0.05$).

Supplemental Figure 5: *Sfla* Or67b3 has fast recovery.

Representative electrophysiological recording obtained from the ab3 sensilla of flies expressing *Sfla Or67b3* in response to stimulation with iso-butyl ITC following sec-butyl ITC with one second interval between chemical administration. The horizontal bar denotes one second and the

vertical bar denotes 10 mV. *Sfla* Or67b3 sensitivity is recovered within seconds after stimulation with ITCs.

Supplemental Material and Methods:

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 *Arabidopsis 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 about 7-10 days old at the time of experiments; *D. melanogaster* (wild-type or transgenic) were about 3-10 days old at the time of the experiments.

M2-MD refers to a CRISPR/Cas9 deletion of *Or22a/b* and a knock-in of *Gal4* and *DsRed* by homologous repair. *Gal4* is not functional but *DsRed* is expressed in the eye. The functional absence of *Or22a* and *Or22b* genes in M2-MD flies was confirmed by electrophysiological analysis on *Or22a/b* expressing neurons in wild-type flies (Fig. S3). The M2-MD line was used to generate flies expressing *Dmel Or67b*, *Spal Or67b* and *Sfla Or67b1-3* under the control of *Gal4* in the ab3A "empty neuron". The *UAS-SflaOr67b1, UAS-Sfla Or67b2, UAS-Sfla Or67b3,* and *UAS-Spal Or67b* strains were generated during this study.

896 **_Scaptomyza Or67b_ gene cloning, UAS line generation, and verification of _S. flava Or67b_**
897 **transcription**

898 The _UAS-Or67b_ transgene lines were constructed as follows: RNA was extracted from
899 20–25 male and female, laboratory-reared, adult flies of _S. pallida_ (collected from the White
900 Mountains, New Mexico) and _S. flava_ (collected from near Portsmouth, New Hampshire). RNA
901 was extracted using Trizol (Thermo-Fisher, Waltham, MA) and precipitated with isopropanol.
902 Extracted RNA was treated with DNaseI; cDNA was then generated using qScript cDNA
903 Supermix (Quantabio, Beverly, MA). Absence of gDNA in cDNA preparations was verified by
904 attempting to amplify fragments of the _Marf_ gene from reactions lacking reverse transcriptase
905 (Fig. S2; PCR conditions and primers detailed in (59)). CDS plus 7–9 bp of untranslated
906 sequence were amplified using High Fidelity Phusion Taq (New England BioLabs, NEB), 3%
907 DMSO v/v, and the PCR primers (Table S2) with the following program: initial denaturing at
908 98°C during 30 sec; 35 cycles at 98°C for 10 sec, 58°C for 30 sec, 72°C for 45 sec, and
909 extension at 72°C during 7 min. PCR fragments of the expected size were purified using
910 Qiaquick Gel purification kit protocol (Qiagen). An overhang was added to purified _Or67b_
911 amplicons with Taq polymerase (Fermentas) and cloned using the pGEM-T Easy cloning kit
912 protocol (Promega). Plasmids were extracted and purified using the GenElute plasmid miniprep
913 kit (Sigma-Aldrich, St. Louis, MO). EcoRI and KpnI cut sites were introduced using restriction
914 enzyme cut-site primers (Table S2) with 10 ng/µL diluted plasmids (as template) with 3%
915 DMSO vol/vol and the following program: initial denaturing at 98°C for 30 sec; 35 repetitions of
916 98°C for 10 sec, 55°C for 50 sec; 72°C for 45 sec; and final extension at 72°C for 7 min. The
917 pUAST attB plasmid (60) and the four _S. pallida_ and _S. flava Or67b_ PCR amplicons with RE
918 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 sent for injection into the y^1 w^{67c23}; P{CaryP}attP2 fly line (BestGene Inc., Houston, Texas, USA). Transformants were selected from individually injected flies with eye color rescue phenotypes. The four *UAS-Or67b* lines created in this study, the *UAS-Dmel Or67b* line, and the *UAS-dTrpA1* (stock no. 26264) line was each crossed into the M2-MD line. The progeny were then used for SSR recordings and behavioral experiments.

Behavioral tests of olfactory attraction

The olfactory responses of mated, non-starved *D. melanogaster* (Canton-S or transgenic), *S. pallida*, and *S. flava* were tested using a dual-choice "Y-shaped" olfactometer based on one previously published (20). The "Y piece" of the olfactometer was a propylene connector (43 mm long x 60 mm); the open ends of the arms of the "Y" were each connected to a 1-ml syringe containing a piece of filter paper (600 x 50 mm) loaded with the odor or control stimuli. Charcoal-filtered air was delivered to each of the two syringes using silicon tubing at 250 ml/min; thus, at the base of the maze air flow was approximately 500 ml/min. Two hours (in the case of *D. melanogaster* and *S. pallida*) or 20 hours before tests (in the case of *S. flava*) insects were gently anesthetized under CO_2 and placed in groups of four in open-top and mesh-bottom cylindrical release containers (20 mm long x 10 mm diameter) constructed using silicon tube. 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

942 honey solution (in the case of *S. flava*). Before tests, each release tube was placed in ice during
943 45-60 seconds to slow down insect activity, the cotton cap was removed, and the open-top of the
944 tube was carefully slid into the open end of the Y maze. Thus, upon being released, insects could
945 walk upwind towards the "decision point" (intersection of the short and long arms of the "Y")
946 and turn towards either the odor-laden or the odorless arm of the maze. Although four insects
947 were released at once (to increase experimental efficacy), only the first choice (and the time of
948 the choice) was recorded; a choice was considered as such only if the insect walked past at least
949 10 mm into one of the arms, orienting upwind. The test was discarded if two or more insects
950 choose the two arms of the maze within a 3-second window of each other. Each test lasted a
951 maximum of five minutes, and each group of insects was used only once. As much as possible,
952 insects from the same cohort were tested in the same day with different odors/odorants; insects
953 from the three different species were also, as much as possible, tested in the same day with a
954 given odor/odorant/different fly genotypes. Tests with each species/stimulus were conducted in
955 at least five different days to compensate for possible day-to-day variations. Results from an
956 individual test session with a given odor/odorant were discarded if insects did not make a choice
957 is more than 50% of tests (this happened in less than 5% of experimental sessions). The position
958 of the odor and odorless arms was switched every 1-2 tests to control for positional asymmetries;
959 the mazes and odor sources were changed and replaced for clean/new ones every four tests or 10
960 minutes, whichever comes first.

961 The odor/odorant was loaded in a piece of filter paper and inserted into the 1-ml syringe
962 right before tests; control syringes had a piece of filter paper (3 x 40 mm) loaded with the
963 mineral oil solvent (for odorant solutions) or water (for odor mixtures). Experimental and control
964 filter papers were replaced by fresh ones every 4 tests or 10 minutes, whichever came first. The

odorants (20 µl of 1:100 vol/vol mineral oil solution) used in experiments were butyl ITC (Sigma-Aldrich, CAS # 592-82-5, USA) and sec-butyl ITC (Sigma-Aldrich, CAS # 15585-98-5, USA). For tests with transgenic flies, responses were tested at 1:100 vol/vol and 1:1,000 vol/vol of butyl ITC. We also used apple cider vinegar (40 µl, O Organics, USA; 40 µl of distilled water was as control stimulus in these tests). For tests of host-orientation, two-four leaves from young arugula 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 tissue paper. In all cases the Y-mazes, tubing and syringes were cleaned 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*). A total of 1,061 (669 with wild-type flies of *D. melanogaster*, *S. pallida* and *S. flava*; 392 with transgenic *D. melanogaster*) tests were and insects made a choice in 84% (888) of tests.

For each odor/odorant, species, sex and genotype, the number of tests with the first insect orienting towards the odorous arm and the odorless arm of the Y-maze were tested against a 50% expected random distribution using two-tailed Binomial tests (48). P-values were adjusted for multiple comparisons using the false discovery rate (FDR) method of Benjamini-Hochberg (49); results were considered significant if the Benjamini-Hochberg adjusted p-value was <0.05. For all tests, we verified that the power was >0.8.

Molecular phylogeny of drosophilid Odorant 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, (50)). *S. flava* sequences were previously published (19). Three hundred and nine sequences were aligned in MAFFT v7.017 with the E-INS-I algorithm and then manually adjusted (51). Models were fitted to the alignment using IQ-Tree and tested using the AIC criterion (52). 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 (53). Orco sequences were designated as the outgroup.

Molecular phylogeny of drosophilid *Or67b* genes:

Or67b CDS from *D. grimshawi*, *D. mojavensis, D. virilis D. sechellia, D. simulans*, *D. erecta, D. yakuba, D. pseudoobscura, D. persimilis, D. ananassae, D. willistoni* 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, dwil r1.3 and dmel r6.28, respectively) were downloaded from Flybase (www.flybase.org, (50)). The *S. pallida* DNA sequence was obtained through PCR as described above; *S. flava* DNA sequences were previously published (19). Sequences were aligned, models fitted and chosen according to AIC (GTR+I+G) in IQ-Tree (52). Trees were inferred using RAxML (v8.2.10) with the GTRCATI 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 (54).

Analysis of molecular evolution:

CDS of homologs of every *Or* gene in *S. flava* found in the 12 *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 above; *S. flava* sequences were previously published (19). Sequences were aligned in MAFFT (v7.017) (51) 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 sequences examined of *Or46a* contain two alternatively spliced exons, so this gene was analyzed with all gene exon sequences in a single alignment as single taxa. *Or69a*, however, contains alternatively spliced exons only among *Sophophora* species. These alternative splice forms were analyzed as separate taxa. Phylogenies were generated for every alignment using PhyML (55) with the GTR+G substitution models. If these trees showed >70% bootstrap support 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. Branch models of sequence evolution were fit using PAML 4.9h (26). 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 for every unique phylogeny (75 tests 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 with identical likely ratio tests performed as above (29 tests total Fig. 2*B*; Table S1). P-values were adjusted for multiple-testing using the FDR method (49).

Distance tree obtained from responses of ab3A neurons expressing *Or67b* transgenes

A matrix of average response of five *Or67b* transgene receptors to 44 odorant compounds (1:100 vol/vol concentration) was produced (Table S3). Net responses were calculated by subtracting the response to mineral oil or dimethyl sulfoxide solvent from the odor-evoked responses. A Euclidean distance matrix was generated using the dist function of the R stats package(56). Receptor responses were clustered by using the neighbor joining (NJ) algorithm on this distance matrix. Support for clusters was assessed using 1000 bootstraps of the original response matrix by generating distance matrices and NJ trees on the pseudo-datasets in ape (v5.3) (57).

Single sensillum recordings (SSR)

Female fed flies were prepared for SSR as previously described (58). We identified sensilla of targeted olfactory receptor cells 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 the ab3 sensillum. Signals were amplified 10,000 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: 3k Hz), and analyzed off-line with WinEDR (v3.9.1; University of Strathclyde, Glasgow). A tube delivering a constant flow of charcoal-filtered air 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 Stimulus Controller CS 55 (Syntech, Germany). We used three standard odors for identification of ab3 sensilla (all from Sigma-Aldrich, US, purity > 1%): ethyl hexanoate (CAS # 123-66-0), ethyl acetate (CAS # 141-78-6) and 2-heptanone (CAS

110-43-0). The following odor sources (all purchased in Berkeley, California, USA unless otherwise mentioned; 20 µl of material loaded on filter paper) were used: apple cider vinegar (40 µl, O Organics, USA), grated roots of *Wasabia japonica* (wasabi), organic roots of *Armoracia rusticana* (horseradish), *Brassica rapa* (turnip), *Raphanus sativus* (daikon), and *Beta vulgaris* (beet). *Eruca vesicaria* (arugula) was grown from seeds at 23°C and 60% relative humidity in a 12-hours light: 12-hours dark cycle, and leaves from 3-8 weeks old plants were used for odor stimulation. Roots and leaves were grated and homogenized using a vegetable grater to a volume equivalent to ~500 µl. The following odorants (all from Sigma-Aldrich, US) were diluted 1:100 vol/vol in dimethyl sulfoxide: iodoacetamide (CAS # 144-48-9), N-methylmaleimide (CAS # 930-88-1), N-hydroxysuccinimide (CAS # 6066-82-6), and 3-indol-carbinole (CAS # 700-06-1) and benzyl thiocyanate (CAS # 3012-37-1). All the other chemicals were diluted in mineral oil at 1:100, 1:1,000, and 10,000 vol/vol. The "net number of spikes" were obtained by counting the number of spikes during a 1-second window 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). Data was analyzed using Mann-Whitney Rank sum tests (for comparisons involving two means) and p-values were adjusted for multiple comparisons using the FDR method, or by Kruskal-Wallis ANOVAs followed by Dunn's tests (for comparisons involving more than two means).

Principal Component Analysis (PCA). For visualization, Principal component analysis (PCA) was done using Python.

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), Python, and Geneious (10.0.9).