Olfactory receptor pseudo-pseudogenes

Lucia L. Prieto–Godino1, Raphael Rytz†, Benoîte Bargéton1, Liliane Abuin1, J. Roman Arguello1, Matteo Dal Peraro2,3 & Richard Benton4

Pseudogenes are generally considered to be non-functional DNA sequences that arise through nonsense or frame-shift mutations of protein-coding genes1. Although certain pseudogene-derived RNAs have regulatory roles2, and some pseudogene fragments are translated3, no clear functions for pseudogene-derived proteins are known. Olfactory receptor families contain many pseudogenes, which reflect low selection pressures on loci no longer relevant to the fitness of a species4. Here we report the characterization of a pseudo-pseudogene in the chemosensory variant ionotropic glutamate receptor repertoire5,6 of Drosophila sechellia, an insect endemic to the Seychelles that feeds almost exclusively on the ripe fruit of Morinda citrifolia7. This locus, D. sechellia Ir75a, bears a premature termination codon (PTC) that appears to be fixed in the population.

However, D. sechellia Ir75a encodes a functional receptor, owing to efficient translational read-through of the PTC. Read-through is detected only in neurons and is independent of the type of termination codon, but depends on the sequence downstream of the PTC. Furthermore, although the intact Drosophila melanogaster Ir75a orthologue detects acetic acid—a chemical cue important for locating fermenting food8,9, found only at trace levels in Morinda fruit10—D. sechellia Ir75a has evolved distinct odour-tuning properties through amino-acid changes in its ligand-binding domain. We identify functional PTC-containing loci within different olfactory receptor repertoires and species, suggesting that such ‘pseudo-pseudogenes’ could represent a widespread phenomenon.

Figure 1 | Ir75a encodes an acetic acid receptor in D. melanogaster and is a transcribed pseudogene in D. sechellia. a, Top, schematic of the third antennal segment covered with porous olfactory sensilla of various morphological classes. Bottom, schematic of the ac2 sensillum class, which houses three OSNs that each express different ionotropic receptor genes. b, Electrophysiological responses in ac2 sensilla to the indicated odours (mean ± s.e.m.; mixed genders) in D. melanogaster30 (n = 9), D. simulans31 (n = 9) and D. sechellia31 (n = 8). The colours of the columns on the histogram distinguishes two broad chemical classes of odours: acids (magenta) and amines (black). c, Immunostaining with anti-Ir75a (magenta) and anti-Ir8a (green) antibodies on antennae of wild-type (left) or Ir75a-mutant (Ir75aMB00251, right) animals. Insets show the co-localization of Ir75a and Ir8a in the OSN soma and dendritic compartment (arrowheads). Scale bars, 10 μm. d, Representative traces of extracellular recordings of neuronal responses to the indicated stimuli in ac2 sensilla in control (Ir75a-GAL4/Ir75aMB00251/+), Ir75a hemizygous mutant (Ir75a−/−; Ir75aMB00253/Df(3L)BSC415) and Ir75a rescue (UAS-D. melanogaster (Dm)/Ir75aIr75a-GAL4,Ir75aMB00251/Df(3L)BSC415) animals. Bars above the traces mark 1 s stimulus time. e, Quantification of solvent-corrected responses in d. Data are mean ± s.e.m.; mixed genders. Control 1: Df(3L)BSC415/+ (n = 12); control 2: Ir75a-GAL4,Ir75aMB00251/+ (n = 11); Ir75a−/− (n = 12); Ir75a rescue (n = 13). Statistical differences between genotypes were tested using pairwise Wilcoxon rank-sum tests among the responses to each odorant, and p values adjusted for multiple comparisons using the Benjamini–Hochberg method. Significant comparisons to Ir75a−/− are shown in the figure (***p = 0.0001, **p = 0.001; for full information about p values see Source Data & Methods). f, Top, gene model of Ir75a indicating the position of the C640T nucleotide substitution in the D. sechellia orthologue. Bottom, genomic sequence spanning this nucleotide position in D. melanogaster and several geographically distributed D. simulans and D. sechellia strains (Methods). The bottom italicized sequence is of the D. sechellia cDNA. The D. sechellia C640T substitution (red) creates a PTC (underlined).
Comparative electrophysiological analysis of olfactory sensory neuron (OSN) responses in closely-related drosophilid species revealed a loss of sensitivity to acetic acid in D. sechellia neurons housed in the antennal coelomic 2 (ac2) sensillum class of sensory hairs (Fig. 1a, b). In D. melanogaster, acetic acid is detected by ac2 OSNs expressing Ir75a\(^{11}\); these sensilla house two other neurons that are sensitive to amines and express Ir41a and Ir75d\(^{11,12}\) (Fig. 1a). Two lines of evidence support Ir75a as the acetic acid receptor. First, the protein is expressed exclusively in these acetic-acid-sensing ac2 neurons, where it co-localizes with the ionotropic receptor co-receptor Ir8a\(^{13}\) in somata and sensory dendrites (Fig. 1c). Second, protein-null Ir75a-mutant animals (Fig. 1c) lack responses to acetic acid (and other organic acids) in these sensilla, while amine ligand-evoked action potentials are unaffected (Fig. 1d, e). Acid sensitivity is restored by expression of an Ir75a cDNA in these neurons (Fig. 1d, e).

Ir75a orthologues are present across drosophilids,\(^6\) but D. sechellia Ir75a is a predicted pseudogene. A C640T nucleotide substitution in the open reading frame (ORF) creates a PTC (CAA→TAA) in exon 4 (Fig. 1f) that is predicted to truncate the protein within the ligand-binding domain (LBD). This PTC is present in all D. sechellia strains that we sequenced (from at least two islands of the Seychelles archipelago), but not in any D. melanogaster or D. simulans strain (Fig. 1f), suggesting that it is a derived change that is fixed in the D. sechellia population. We could, however, amplify Ir75a cDNA from D. sechellia antennal RNA. Sequencing of this cDNA, in addition to data from D. sechellia antennal RNA sequencing (Methods), verified that the PTC is not edited or spliced out of the transcript to maintain an intact ORF (Fig. 1f).

The pseudogenization of D. sechellia Ir75a provided a logical explanation for the loss of responses to acetic acid in this species (Fig. 1b). Nevertheless, D. sechellia ac2 sensilla house a neuron that responds to other acidic odours (Fig. 1b), suggesting that another receptor is expressed in these OSNs. We were, however, unable to detect other acid-sensing ionotropic receptors in these cells. We therefore wondered whether the Ir75a pseudogene might encode a functional receptor. Indeed, the anti-Ir75a antibody stains OSNs in D. sechellia with a similar distribution to that of ac2 sensilla (Fig. 2a). Because the epitope of this antibody is encoded upstream of the PTC, we generated a second antibody that recognizes an epitope of the protein that is encoded downstream of the PTC (anti-Ir75a\(^{\text{D}}\)), and found that it labelled the same cells (Fig. 2a). We also generated a transgene comprising D. sechellia Ir75a cDNA in which the terminal stop codon was removed and the coding sequence for GFP inserted in-frame with the last coding codon (DsIr75a:GFP). As D. sechellia is not yet amenable to transgenesis, we expressed this construct in D. melanogaster Ir75a neurons. GFP fluorescence was detected from D. sechellia Ir75a:GFP (Fig. 2b and Extended Data Fig. 1), indicating that the PTC is read through, permitting translation of the downstream GFP sequence. No GFP signal was observed with a control construct that retained the terminal stop codon (DsIr75aSTOP:GFP) (Fig. 2b and Extended Data Fig. 1).

Next we asked whether D. sechellia Ir75a encodes a functional receptor by misexpressing it in heterologous ‘ionotropic receptor decoder’ neurons (that is, ac4 Ir84a-mutant neurons that lack the endogenous ligand-specific Ir84a, but which still express the co-receptor Ir8a). In these cells, D. melanogaster Ir75a endowed sensitivity to acetic and propionic acids (Fig. 2c, d), consistent with the expected endogenous responses of ac2 Ir75a OSNs (Fig. 1b). By contrast, D. sechellia Ir75a conferred responses to propionic, butyric and 2-oxopentanoic acids (Fig. 2c, d). Cluster analysis revealed that the responses of D. sechellia Ir75a in ionotropic receptor decoder neurons and the endogenous responses of ac2 sensilla neurons group together (Fig. 2e, f). These results provide evidence that the D. sechellia Ir75a pseudogene encodes a functional olfactory receptor that accounts for the ac2 acid-sensing properties of this species.

Reversion of the PTC to the ancestral glutamine-encoding codon (TAA→CAA; *214Q) in transgenic constructs had no effect on expression or function of D. sechellia Ir75a (Fig. 3a, b and Extended Data Fig. 1), indicating that the PTC is read through efficiently and does not influence odour responses. Translational read-through of
terminal stop codons, resulting in C-terminal extensions, has been characterized for several eukaryotic genes15,16; in these cases, the 'leakiness' of translation arrest is predicted to depend  on the termination codon (TGA > TAA > TAG) and the immediate 3′ nucleotide (C > T > G > A)14,15. We investigated the cis-regulatory  elements that determine the high efficiency of read-through of the ir75a PTC— which has the second most leaky termination codon context (TAAC)14—by generating additional read-through GFP reporters bearing mutations in this sequence (Fig. 3c, e). Replacement of the TAA PTC with either TGA or TAG did not affect GFP expression (Fig. 3c, d and Extended Data Fig. 1). By contrast, replacing the immediate 3′ cytosine nucleotide with an adenosine almost completely blocked GFP expression (Fig. 3e and Extended Data Fig. 1), although this transgene still produces a truncated protein, as detected by the anti-Ir75a antibody (Fig. 3e). These results indicate that sequence context outside, but not within, the ir75a PTC is critical for determining read-through efficiency.

The expression of full-length protein by PTC-containing D. sechellia ir75a transgenes in two populations of D. melanogaster OSNs (that is, ir75a and ionotropic receptor decoder neurons) indicates that the mechanisms that permit read-through are not species- or OSN-class-specific. We investigated whether read-through occurs in other cell types by using an actin5C-GAL4 driver to broadly express D. sechellia ir75a:GFP and, as a control, D. sechellia ir75a214Q:GFP. D. sechellia ir75a214Q:GFP was detected in many, but not all, cells (Extended Data Fig. 2); this heterogeneity may arise from the variable expression of actin5C-GAL4 in different cell types.

Nevertheless, the GFP-positive cells encompassed neurons and non-neuronal support cells (Extended Data Fig. 2, arrowheads). By contrast, D. sechellia ir75a:GFP was detected exclusively in ir75a neurons (Extended Data Fig. 2). To confirm this finding, we compared the expression of D. sechellia ir75a214Q:GFP and ir75a:GFP transgenes induced either by a pan-neuronal (elav-GAL4) or pan-glial (repo-GAL4) driver. Both transgenes produced similar GFP signals in sensory neurons throughout the antenna (Fig. 3f). However, only ir75a214Q:GFP was detectable in glia (Fig. 3g). Similarly, we detected broad neuronal expression of both ir75a214Q:GFP and ir75a:GFP in the brains of these animals (Fig. 3h), but only ir75a214Q:GFP was expressed in glia (Fig. 3i). Thus, efficient read-through of the D. sechellia ir75a PTC occurs in diverse neuronal classes, but not in non-neuronal cells.

We next investigated the molecular basis of the different ligand-response profiles of D. sechellia and D. melanogaster Ir75a. As Ir75a-dependent neuron responses are conserved between D. melanogaster and D. simulans (Fig. 1b), we reasoned that ligand-specificity determinants would be conserved in Ir75a sequences from these species but differ in D. sechellia ir75a; of eight such positions (excluding the PTC), six lie within the predicted bi-lobed BLD (Extended Data Fig. 3). Three of these positions are located within the putative ligand-binding pocket in a protein homology model of the D. sechellia ir75a BLD (Fig. 4a). We tested their function by generating a version of D. melanogaster ir75a in which each of these positions was mutated to encode the amino acid present in D. sechellia ir75a, and expressed this mutant (D. melanogaster ir75a214Q:GFP).
in ionotropic receptor decoder neurons. This engineered receptor conferred responses indistinguishable from those of *D. sechellia* Ir75a (Fig. 4b), indicating the importance of one or more of these residues as odor-specificity determinants.

Finally, we sought to identify other functional olfactory receptor pseudogenes. Within wild-caught isolates of *D. melanogaster* (Methods), we identified several strains in which the Ir75b ORF contains a C517T substitution, creating a PTC (Fig. 5a and Methods). Immunostaining of antennae from flies of these lines with an Ir75b antibody (recognizing an epitope encoded downstream of the PTC) showed a pattern comparable to controls (Fig. 5a, b), indicating that this PTC is read through. Consistently, electrophysiological recordings of Ir75b neurons in one of these lines (Raleigh707, RAL707) revealed robust responses when presented with known agonists (Fig. 5c). In another strain (RAL441), the Ir31a ORF contains a T1805G substitution that is predicted to truncate the receptor before the third transmembrane domain (Fig. 5d). Nevertheless, electrophysiological recordings of RAL441 Ir31a neurons stimulated with 2-oxo-pentanoic acid, revealed clear responses (Fig. 5e). We also identified a segregating PTC-containing allele of the odorant receptor gene Or35a (in Tasmanian strains T09 and T29) (Fig. 5f). In strain T09, responses of Or35a neurons to octanol were readily detected (Fig. 5g). These findings indicate that the phenomenon of functional olfactory pseudogenes is restricted neither to a particular species nor to a specific receptor repertoire.

Our efforts to understand the molecular basis of the loss of olfactory sensitivity to acetic acid in *D. sechellia* led us to discover a notable and, to our knowledge, unprecedented evolutionary trajectory of a presumed pseudogene. Efficient read-through of a PTC in *D. sechellia* Ir75a permits production of a full-length receptor protein, in which reduction in acetic acid sensitivity and gain of responses to other acids is due to lineage-specific amino acid substitutions in the LBD pocket. The PTC does not noticeably influence the activity of *D. sechellia* Ir75a, suggesting that it is selectively neutral from an evolutionary standpoint. We propose that it became fixed through genetic drift, given the low read-through rates (26). Our characterization of four alleles with low read-through rates (26) or the decoder neurons of *D. sechellia* Ir75a, indicating the position of the PTC in the RAL707 strain (an identical sequence is found in all other strains containing a PTC at this position; see Methods). The region encoding the epitope recognized by anti-Ir75b antibodies is indicated with a red bar. b, Immunofluorescence with anti-Ir75b antibodies on antennae of reference *D. melanogaster* (Methods) and RAL707 flies. Scale bars, 10 μm.

Quantification of odour-evoked responses to 2-oxo-pentanoic acid in a3 sensilla of control (reference *D. melanogaster*) and RAL441 flies. Box plots indicate the median and first and third quartile of the data; each dot corresponds to one recording. d, Gene structure of *D. melanogaster* Ir31a, indicating the position of the PTC in the RAL441 strain. e, Quantification of odour-evoked responses to 2-oxo-pentanoic acid in a3 sensilla of control (reference *D. melanogaster*) and RAL441 flies. f, Gene structure of *D. melanogaster* Or35a, indicating the position of the PTC in the T09 (and T29) strain. g, Quantification of odour-evoked responses to octanol in a3 sensilla of control (reference *D. melanogaster*) and T09 flies.

selenocysteine (which is incorporated at UGA) (19). Moreover, no suppressor tRNAs are known in *D. melanogaster* (19) and ribosomal frame-shifting is also unlikely because there is no change in the reading frame after the PTC. We suggest that read-through is due to PTC recognition by a near-cognate tRNA that allows insertion of an amino acid instead of translation termination. Although the trans-acting factors regulating read-through are unclear, the neuronal specificity of this process is reminiscent of RNA editing and micro-exon splicing, in which key recognizable regulatory proteins are neuronally enriched (20, 21). We therefore speculate that tissue-specific expression differences in tRNA populations underlie neuron-specific read-through.

Although it has long been known that viruses display PTC read-through (22), case studies in eukaryotes are largely limited to artificial scenarios in which nonsense mutations have been introduced through random or site-directed mutagenesis (23–25), or in human disease-causing alleles with low read-through rates (26). Our characterization of four protein-coding PTC-containing genes demonstrates that the read-through of naturally occurring PTCs can be sufficiently efficient to permit the functionality of pseudogenes and the maintenance of these variants in populations. This finding further highlights the plasticity of translational regulation, allowing for the phenotypic buffering of genetic changes (27). It should also prompt the experimental examination...
of the hundreds of PTC-containing presumed pseudogenes, both within and beyond chemosensory gene families in insects, humans and other organisms.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 5 February; accepted 6 September 2016.
Published online 24 October 2016.

1. Salmena, L. Pseudogene redux with new biological significance. Methods Mol. Biol. 1167, 29449–29461 (2016).
2. Poliseno, L. et al. Many lncRNAs, mRNAs regulates tumour biology. Nature 530, 48–54 (2016).
3. Ji, Z., Song, R., Regev, A. & Struhl, K. Many lncRNAs, 5′UTRs, and pseudogenes are translated and some are likely to express functional proteins. eLife 4, e08890 (2015).
4. Rei, M., Nimmura, Y. & Nozawa, M. The evolution of animal chemosensory receptor gene repertoires: roles of chance and necessity. Nat. Rev. Genet. 9, 951–963 (2008).
5. Benton, R., Vannice, K. S., Gomez-Diaz, C. & Vossalli, L. B. Variant ionotropic glutamate receptors as chemosensory receptors in Drosophila. Cell 136, 149–162 (2009).
6. Croset, V. et al. Ancient protostome origin of chemosensory ionotropic glutamate receptors and the evolution of insect taste and olfaction. PLoS Genet. 6, e1001064 (2010).
7. Stensmyr, M. C. Drosophila sechellia as a model in chemosensory neuroecology. Ann. NY Acad. Sci. 1170, 468–475 (2009).
8. Gorter, J. A. & Lowe, T. M. GtRNAdb: a database of transfer RNA genes detected in genomic sequence. Nucleic Acids Res. 37, D93–D97 (2009).
METHODS
No statistical methods were used to predetermine sample size. These experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Molecular biology and transgenesis. *D. melanogaster* and *D. sechellia* Ir75a ORFs (including the PTC in *D. sechellia* Ir75a) were cloned into puAST-attB43. Site-directed mutagenesis of *D. sechellia* Ir75a and *D. melanogaster* Ir75a, reverse-transcription PCR amplification and sequencing of genomic amplicons to verify the presence of the *D. sechellia* Ir75a PTC were performed using standard procedures. For the transgene in which the nucleotide 3’ of the PTC was mutated (Fig. 3e), we changed codon 215 from CGT → AGT to maintain the identity of the encoded amino acid (arginine). Oligonucleotide and plasmid sequences are available upon request. New transgenes were integrated in attP40 using the phiC31 site-specific integration system32 by Genetic Services Inc. or BestGene Inc. All transgenes were sequence-verified both before and after integration into *D. melanogaster*.

*Drosophila* strains. Flies were maintained at 25 °C in 12 h light:12 h dark conditions. *D. melanogaster* wild-type refers to w1118, unless noted otherwise. *D. sechellia* wild-type was 14021-0248.25 (*Drosophila* Species Stock Center, UCSD). The sequenced region of Ir75a shown in Fig. 5c was amplified from the following strains (from the *Drosophila* Species Stock Center, unless noted otherwise): *D. sechellia*: 14021-0248.08, 14021-0248.11, 14021-0248.13, 14021-0248.15, 14021-0248.19, 14021-0248.25, 14021-0248.27, 14021-0248.30. *D. simulans*: 14021-0251.195, 14021-0251.196, 14021-0251.197, as well as a Seychelles-isolated *D. simulans*33. Other published mutant and transgenic lines used were: Ir84a-GAL4 (ref. 34), Ir75a-GAL4 (ref. 11), UAS-CD8-GFP35, Mi(E/T)Ir75aB(M000253)36, Df(3L) BSC415 (Ir75a deficiency)37, actin5C-GAL4 (ref. 38), elav-GAL4 (Bloomington 458), repo-GAL4 (Bloomington 7415), RAL441, RAL707 (ref. 39) and Tasmania T09 (ref. 40).

Sequence analysis. We downloaded *D. sechellia* antennal RNA-sequencing datasets42 from the NCBI Gene Expression Omnibus repository (GE related accessions GSE67861 and GSE67587; SSR files SRR1952772, SRR1952777, SRR1937487, SRR1937490). The sra files were converted to fastq files and remapped to the *D. sechellia* genome (r1.3) using TopHat (v2.0.13) → b2-sensitive). The genomic index and splicing index were also generated with TopHat using the *D. sechellia* gtf (r1.3). The resulting bam files were visualized within IGV (v2.3.63) and we manually inspected reads that covered the Ir75a PTC. Within these four datasets, ~100% of the reads supported the presence of the PTC-causing ‘T’ allele (only 6/1777 reads within all four datasets supported an alternative nucleotide, within the noise of sequencing errors).

PTCs in other olfactory receptor genes were identified in the *D. melanogaster melanogaster* Genetic Reference Panel (DGRP; http://dgrp2.gnets.ncsu.edu/)39 and/or the Global Diversity Lines (GDL)40,41. For Ir75b, the following lines contain the same PTC: DGRP (RAL181, RAL189, RAL227, RAL320, RAL348, RAL358, RAL374, RAL379, RAL382, RAL385, RAL395, RAL399, RAL439, RAL461, RAL531, RAL596, RAL707, RAL712, RAL716, RAL730, RAL804, RAL805, RAL821, RAL855, RAL846), GDL (B10, B11, B12, I26, N01, N02, N07, N25, T29). We re-sequenced the PTC-containing and confirmed Ir75b protein expression in all strains from the GDL (data not shown). For Ir31a, only RAL441 contains the PTC. For Or35a, GDL Tasmanian strains T09 and T29 contain the same PTC.

Histology and morphological analyses. Immunofluorescence on whole-mount antennae or antennal crossections (used only in Fig. 1c, inset, and Fig. 2a), were performed as described1,13,43. Affinity-purified antibodies were generated by Proteintech Group, Inc., against the following peptides: KRSKYGNREQLTDVVLRV (anti-Ir75a, in rabbits; used at 1:100 for whole-mount antennae, and 1:500 for cryosections), and RPLTLSDDELIRFLSQEND (anti-Ir75b, in guinea pigs; used at 1:500); these peptides are predicted antigenic, with a maximum of three sensilla per animal. Exact sample sizes for each experimental or group condition are provided in the Source Data. Genotypes (not blinded to experimenter) were interlaced to minimize effects of time-of-day and animal age. All odour-evoked responses were corrected for solvent-evoked spikes. Chemicals were purchased from Sigma-Aldrich and were of the highest purity available. Odorants were used at 1% (v/v) in all experiments unless otherwise noted in the figure legends. Odor stimulus cartridges (10 μl odour dilution on ~5 × 5 mm Sugi strip placed in a 2-ml plastic syringe) were prepared freshly before each recording session; cartridges were interleaved, with a maximum of five stimuli. Stimuli, with CAS number and solvents used in brackets, are as follows: 1,4-diaminobutane (110-60-1; H2O), 2-oxopentanoic acid (1821-02-9; paraffin oil), acetic acid (64-19-7; H2O), butyric acid (107-92-6; H2O), hexanoic acid (142-62-1; H2O), octanol (111-87-5; paraffin oil), phenylenylamine (64-04-0; paraffin oil), propionic acid (79-09-4; H2O), pyridine (110-86-1; paraffin oil). Ionotropic receptor decoder neurons are ac3 h84a-mutant neurons that lack the endogenous ligand-specific Ir84a, but that still express the co-receptor Ir84f.44 For measurement of odour-evoked responses of Ir75b neurons in ac3 sensilla (Fig. 5c), we note approximately half of the analysed sensilla belong to ‘ac3H1’ class (expressing Ir75c) that are electrophysiologically indistinguishable from ac3 sensilla (expressing Ir75b; L.L-P., G., R.R. and R.B., unpublished data). Thus, if the RAL707 PTC-bearing allele is non-functional, we would expect half of the sensilla to show no responses, which is not the case.

Statistical analysis. Sample sizes were fixed before data analysis, based on preliminary studies. Data were analysed and plotted using ‘R project’ (http://R-project.org). Data were analysed statistically using the Shapiro–Wilk test to assess for normality followed by a two-tailed Student’s t-test or a Wilcoxon rank-sum test as appropriate. When a P value correction for multiple comparisons was needed, the Benjamini–Hochberg method was used. Full statistical test results are provided in the Source Data files for each figure.

Cluster analysis of odour responses. Spike-count data (from responses of endogenous or ionotropic receptor decoder neurons) were imported to Matlab (Mathworks), and standardized using z-scores across each recording. An unbiased k-means cluster analysis was performed using the response properties to the indicated number of odours. The optimal number of clusters was determined by the silhouette method45. The silhouette value is a measure of both how tightly a data point is associated with its assigned cluster and how dissimilar it is from other clusters, and it peaks at the ‘correct’ number of clusters. If the distribution is unimodal (that is, all data fall within one cluster) the silhouette value does not peak and is similar for each value of k. In brief, we first ran iteratively the Matlab k-means algorithm 100 times for k values between 2 and 6. We then calculated the silhouette value for each k-means solution, and subsequently computed and plotted the mean silhouette value and its standard deviation of all of the solutions within each k value. The results of the clusters from the k-mean analysis were plotted in Matlab on the principal component space after performing a principal component analysis using the same odour dataset. The scripts used to analyse data are available upon request.

Protein homology modelling. A multiple-sequence alignment of the LB of *D. melanogaster* antennal ionotropic receptors (Ir86a, Ir21a, Ir25a, Ir31a, Ir40a, Ir41a, Ir46a, Ir75a, Ir75b, Ir75c, Ir76a, Ir76b, Ir84a, Ir92a, Ir93a)46, Rattus norvegicus GluK2 (UniProt ID P42260), Rattus norvegicus GluA2 (P19491), Adneta vaga GluR1 (EP95SP) and Synancinetes PCC 6803 GluR0 (P73797) was generated by PROMAL53D47. A GT dipetptide sequence was introduced within the S1 and S2 domains of the *Drosophila* proteins to facilitate alignment with the linker sequence included in the crystallized GluA2, GluK2, ArGlur1 and GluR0 LBDs. Alignment of Ir75a proteins was curated according to PIPERED secondary structure predictions46. Models of the *D. sechellia* Ir75a LBD (V205 → T318 → GT-K434 → C579) were built using MODELLER (mod9.12)48 using as templates the apo and ligand-bound cryostructure of GluA2 (PDB ID: 1FTM apo state; 1FTM ligand bound)49. The results of standard MODELLER energy functions, molpdf and DOPE, were highly similar between generated models. We illustrate in Fig. 4a the model with the lowest DOPE energy function score.
32. Bischof, J., Maeda, R. K., Hediger, M., Karch, F. & Basler, K. An optimized transgenesis system for Drosophila using germ-line-specific phiC31 integrases. Proc. Natl Acad. Sci. USA 104, 3312–3317 (2007).
33. Carracedo, M. C., Asenjo, A. & Casares, P. Genetics of Drosophila simulans male mating discrimination in crosses with D. melanogaster. Heredity 91, 202–207 (2003).
34. Grosjean, Y. et al. An olfactory receptor for food-derived odours promotes male courtship in Drosophila. Nature 478, 236–240 (2011).
35. Lee, T. & Luo, L. Mosaic analysis with a repressible cell marker for studies of gene function in neuronal morphogenesis. Neuron 167, 761–781 (2004).
36. Cook, R. K. et al. The generation of chromosomal deletions to provide extensive coverage and subdivision of the Drosophila melanogaster genome. Genome Biol. 13, R21 (2012).
37. Ito, K., Awano, W., Suzuki, K., Hiromi, Y. & Yamamoto, D. The Drosophila mushroom body is a quadruple structure of clonal units each of which contains a virtually identical set of neurones and glial cells. Development 124, 761–771 (1997).
38. Mackay, T. F. et al. The Drosophila melanogaster Genetic Reference Panel. Nature 482, 173–178 (2012).
39. Grenier, J. K. et al. Global diversity lines—a five-continent reference panel of sequenced Drosophila melanogaster strains. G3 (Bethesda) 5, 593–603 (2015).
40. Shiao, M. S. et al. Expression divergence of chemosensory genes between Drosophila sechellia and its sibling species and its implications for host shift. Genome Biol. Evol. 7, 2843–2858 (2015).
41. Arguello, J. R. et al. Extensive local adaptation within the chemosensory system following Drosophila melanogaster’s global expansion. Nat. Commun. 7, ncomms11855 (2016).
42. Benton, R. & Dahanukar, A. Electrophysiological recording from Drosophila olfactory sensilla. Cold Spring Harb. Protoc. 2011, 824–838 (2011).
43. Kaufman, L. & Rousseeuw, P. J. Finding Groups in Data: an Introduction to Cluster Analysis. (Wiley-Interscience, 2005).
44. Pei, J., Kim, B. H. & Grishin, N. V. PROMALS3D: a tool for multiple protein sequence and structure alignments. Nucleic Acids Res. 36, 2295–2300 (2008).
45. Jones, D. T. Protein secondary structure prediction based on position-specific scoring matrices. J. Mol. Biol. 292, 195–202 (1999).
46. Salt, A. & Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints. J. Mol. Biol. 234, 779–815 (1993).
47. Armstrong, N. & Gouaux, E. Mechanisms for activation and antagonism of an AMPA-sensitive glutamate receptor: crystal structures of the GluR2 ligand binding core. Neuron 28, 165–181 (2000).
Extended Data Figure 1 | Quantification of efficiency and tissue-specificity of translational read-through of the D. sechellia Ir75a PTC. Quantification of GFP staining in the cell bodies of neurons expressing different read-through reporter constructs in different populations of OSNs (see Figs 2, 3 for genotypes). GFP fluorescence levels were normalized by anti-Ir75a fluorescence levels in the Cy3 channel within each analysed cell. Box plots indicate the median and first and third quartile of the data. *P < 0.05, **P < 0.0005, not significant (n.s.) P > 0.05 (all data analysed using pairwise Wilcoxon rank-sum test, Benjamini–Hochberg correction).
Extended Data Figure 2 | Tissue specificity of translational read-through of the *D. sechellia* Ir75a PTC. Immunofluorescence with anti-GFP (green) and the neuron nuclear marker anti-Elav (magenta) on whole-mount *D. melanogaster* antennae in which *actin5C-GAL4* drives broad expression of *D. sechellia* Ir75a*214Q*-GFP (UAS-DsIr75a*214Q*:GFP/act5C-GAL4) or Ir75a::GFP (UAS-DsIr75a::GFP/act5C-GAL4). Arrowheads indicate examples of GFP-expressing, Elav-negative, non-neuronal cells that were observed in 6 out of 6 antennae expressing the control transgene lacking the PTC, and in 0 out of 6 antennae expressing the PTC-containing transgene. Note that the neuronal GFP signal of both transgenes is heterogeneous across the antenna, possibly because of the variable strength of driver expression and/or instability of the GFP-tagged receptors in heterologous neurons. Scale bars, 10 μm.
Extended Data Figure 3 | Alignment of drosophilid Ir75a orthologues.

Protein-sequence alignment of *D. melanogaster*, *D. simulans* and *D. sechellia* Ir75a. The position of the PTC (X) is highlighted in yellow. Dark grey columns in the alignment highlight amino acids conserved only in two of the three species. Pink and red shading represents *D. sechellia*-specific amino acid changes within the LBD; red denotes the subset located in the internal cavity of the binding pocket (Fig. 4a). The locations of the peptide epitopes for the Ir75a antibodies are highlighted with green dashed boxes.