

# Evolution of chemosensory tissues and cells across ecologically diverse *Drosophilids*

Received: 10 May 2023

Accepted: 19 December 2023

Published online: xx xx 2024

Check for updates

Gwénaëlle Bontonou<sup>1,2,5</sup>✉, Bastien Saint-Leandre<sup>1,2,5</sup>✉, Tane Kafle<sup>1,2</sup>, Tess Baticle<sup>1</sup>, Afrah Hassan<sup>1</sup>, Juan Antonio Sánchez-Alcañiz<sup>3</sup> & J. Roman Arguello<sup>1,2,4</sup>✉

Chemosensory tissues exhibit significant between-species variability, yet the evolution of gene expression and cell types underlying this diversity remain poorly understood. To address these questions, we conducted transcriptomic analyses of five chemosensory tissues from six *Drosophila* species and integrated the findings with single-cell datasets. While stabilizing selection predominantly shapes chemosensory transcriptomes, thousands of genes in each tissue have evolved expression differences. Genes that have changed expression in one tissue have often changed in multiple other tissues but at different past epochs and are more likely to be cell type-specific than unchanged genes. Notably, chemosensory-related genes have undergone widespread expression changes, with numerous species-specific gains/losses including novel chemoreceptors expression patterns. Sex differences are also pervasive, including a *D. melanogaster*-specific excess of male-biased expression in sensory and muscle cells within its forelegs. Together, our analyses provide new insights for understanding evolutionary changes in chemosensory tissues at both global and individual gene levels.

**Q1** Animal's abilities to perceive their chemical environments are remarkably variable. Chemosensory receptor protein families and the cell types in which they are expressed have multiple evolutionary origins<sup>1–6</sup>, and the tissues that contain them can differ dramatically across species in morphology and anatomy<sup>7–12</sup>. For example, while taste perception in mammals is largely restricted to gustatory cells located in the mouth, and primarily the tongue, aquatic vertebrates have taste cells distributed externally on their skin<sup>13–17</sup>. Insects have evolutionarily distinct taste receptors and cells that are also broadly distributed across their bodies, including their mouth parts, legs, wings, and ovipositors<sup>18</sup>. Appendages involved in smell are generally more restricted to animals' heads but also differ dramatically among taxa, as exemplified by the bulbous nose of the Proboscis Monkey or the feathery antennae of moths. In addition to differences among species, striking evolutionary changes have also arisen between sexes

within species. Sexual dimorphisms in chemosensory perception and organ morphology often evolve rapidly and have been attributed to differences in sex-specific physiological states, sexual selection, and sex-specific nutritional needs, among other factors<sup>19–22</sup>.

**Q2–Q6** Understanding the molecular basis of chemosensory evolution is important for fundamental and applied biology. Insights into the genes and expression changes that underlie species' chemosensory differences help us understand how nervous systems adapt in response to varying ecologies and provide the basis for managing disease vectors and agricultural pests. For instance, research on insect chemosensation has advanced our understanding of how mosquitoes track human odors, with important implications for human health<sup>23–25</sup>, and has aided in the development of novel farming methods that reduce crop infestation<sup>26</sup>. While these applications draw on knowledge of

<sup>1</sup>Department of Ecology & Evolution, Faculty of Biology and Medicine, University of Lausanne, Lausanne, Switzerland. <sup>2</sup>Swiss Institute of Bioinformatics, Lausanne, Switzerland. <sup>3</sup>Instituto de Neurociencias, UMH & CSIC, San Juan de Alicante, Spain. <sup>4</sup>Present address: School of Biological and Behavioural Sciences, Queen Mary University of London, London, UK. <sup>5</sup>These authors contributed equally: Gwénaëlle Bontonou, Bastien Saint-Leandre.

✉ e-mail: [gwenaelle.bontonou@unil.ch](mailto:gwenaelle.bontonou@unil.ch); [bastien.saint-leandre@unil.ch](mailto:bastien.saint-leandre@unil.ch); [r.arguello@qmul.ac.uk](mailto:r.arguello@qmul.ac.uk)

chemosensation from a broad range of biological models, much of what we know derives from research on *Drosophila melanogaster*.

Research on *D. melanogaster* has led to extensive knowledge about the development of its nervous system and chemosensory appendages and has generated a nearly complete mapping of its full set of olfactory and gustatory receptor proteins to specific neuron populations. This work has provided the basis for many pioneering functional and behavioral studies<sup>27–32</sup>. In addition, advances in connectomics and single-cell transcriptomics applied to *D. melanogaster*'s nervous system are helping to identify new developmental factors, describe cellular diversity in chemosensory tissues, and characterize synapse-level connections from the peripheral chemosensory neurons to the central brain<sup>33–41</sup>. Beyond its role as a preeminent model for chemosensory biology, *D. melanogaster* and its closely-related species have also long been a model system for evolutionary genetics and speciation<sup>42–47</sup>. The phylogenetic relationships among the *D. melanogaster* species group are well-resolved and include lineages of diverse ages and ecologies. This system, therefore, provides a valuable opportunity to ask how evolutionary forces and environments shape chemosensory systems<sup>22,48,49</sup>. However, beyond the meticulous molecular and cellular characterization of *D. melanogaster*'s chemosensory tissues, little is known about how they evolve between species.

To address this question, we have carried out a comparative transcriptomic experiment in which we generated bulk RNA-sequencing (RNA-seq) datasets for five chemosensory tissues: larva head (mixed sex), ovipositor (female), forelegs (female and male), antennae (female and male), and proboscis with maxillary palps (female and male). These samples were collected from six ecologically diverse species in the *D. melanogaster* species group that share common ancestors between ~0.25–15 million years ago<sup>44,50–53</sup>: *D. melanogaster*, *D. sechellia*, *D. simulans*, *D. santomea*, *D. erecta*, and *D. sukukii* (Fig. 1A). *D. sechellia*, is endemic to the Seychelles island and an extreme specialist on the fruit of *Morinda citrifolia*, which is toxic to the other species<sup>54</sup>. *D. santomea* is endemic to the island of São Tomé and adapted to high elevation mist forests<sup>55,56</sup>. *D. erecta* is restricted to west-central Africa and is thought to be an opportunistic specialist on the fruits of *Pandanus*<sup>57</sup>. *D. sukukii* originated in Eastern Asia but has expanded rapidly worldwide in the last decade<sup>58,59</sup>. Unlike the other species, *D. sukukii* females oviposit in ripe soft-bodied fruits and, as a result, have become a global agricultural pest<sup>60–65</sup>. Both *D. simulans* and *D. melanogaster* are generalists that feed on a broad range of decaying fruits and have nearly worldwide distributions<sup>66</sup>. We found that stabilizing selection has been the predominant force shaping the evolution of chemosensory transcriptomes. Still, several thousand genes have evolved expression changes in each tissue. Intriguingly, genes that have changed expression in one tissue have usually changed expression in multiple other tissues but at different times in the past, suggesting widespread tissue-specific regulatory changes. The fast evolution of chemosensory gene families and sex differences are also prominent and highlight distinct ecological differences. These data can be explored with our dashboard available at: <https://ctct.unil.ch/>.

## Results

### Relationships between sensory tissue transcriptomes

To study the evolution of gene expression in the main chemosensory tissues of closely-related *Drosophila*, we generated bulk RNA-seq datasets for six ecologically diverse species and five sensory tissues (Fig. 1A; Methods). On average, we obtained 43 million mapped reads per sample with high correlations across triplicates (average Pearson correlation coefficient = 0.98). To overcome annotation biases, we used these datasets to produce equivalent de novo gene annotations and used the resulting gene sets for orthology/paralogy assignments. This approach resulted in similar genome annotations with BUSCO scores ranging from 91.9–97.3%, indicating a well-balanced dataset for cross-species comparisons.

We began investigating the relationship between chemosensory tissue transcriptomes by conducting a principal component (PC) analysis on expression levels of 12,096 genes with a one-to-one relationship across the six species (1:1 orthologs; Fig. 1B). The first principal component (PC1) separates the three appendage samples from the larval head and ovipositor samples. The genes that contribute the most to the negative loading of PC1 are enriched for gene ontology (GO) terms related to cilia, cell projections/axons, and synapses, among other neural categories. These terms contrast with the enrichment of cell cycle, organelle, and nucleus-related terms that most contribute to the top positive loadings of PC1 (Supplementary Fig. 1). Soft-clustering analysis of correlated expression changes across multiple genes also identified appendage-specific expression modules that are enriched for cilium, dendrite, and chemosensory terms that load negatively on PC1 and larval/ovipositor-specific modules that are enriched for cell cycle ontology terms that load positively on PC1 (Supplementary Fig. 1). The second principal component (PC2) separates the antenna from the other samples and is enriched for GO terms related to olfactory, dendrite, and sensory function for the top positive loadings. We again identified an antenna-specific module that is enriched for olfactory receptor, odorant binding, and dendrite terms that load positively on PC2 (Supplementary Fig. 1). The gene set defining this antenna module negatively correlates with a muscle-related gene module that is enriched in the forelegs and proboscis datasets (Supplementary Fig. 1), highlighting both neural and structural genes underlying the chemosensory tissue transcriptome differences.

We observed further separation between the antenna, ovipositor, and larva clusters with additional PC pairings, but the foreleg and proboscis+palps transcriptomes always overlap (Supplementary Fig. 2). The latter overlap is likely driven primarily by the foreleg and proboscis tissues, given relatively small tissue contribution of the maxillary palps. Among the five tissues, the ovipositor samples varied the most in PC space, reflecting the lower correlation across some replicates (Methods). Despite this variation, the clustering separated the *D. sukukii* samples, for which the replicates were highly correlated (average Pearson correlation coefficient = 0.97). This *D. sukukii* difference is notable because the females of this species differ from the others in their preference for ovipositing in ripening fruits (instead of overripe/rotting fruits) and have evolved an elongated serrated ovipositor that punctures fruit skins<sup>67</sup>.

### Sensory transcriptomes exhibit low rates of divergence, with a few exceptions

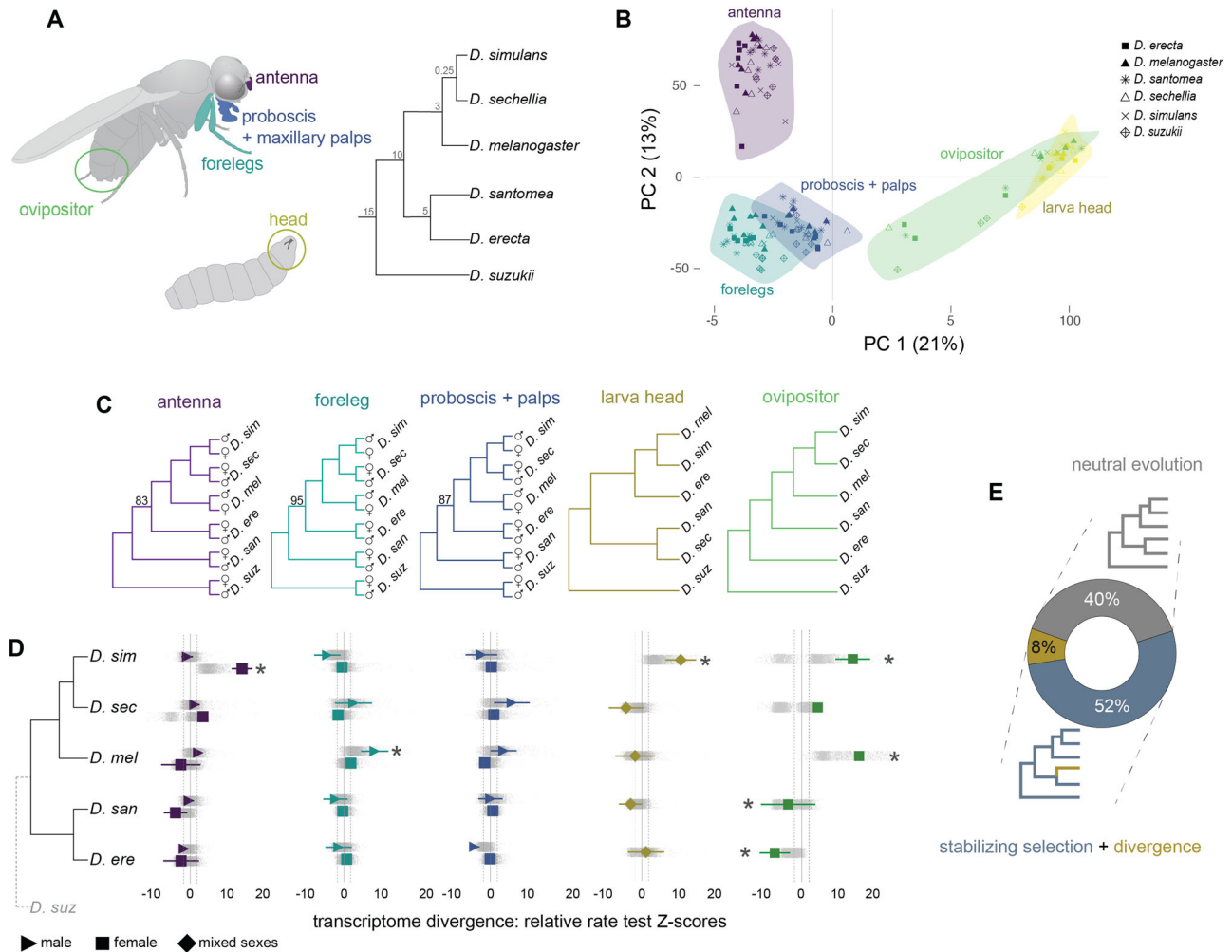
To investigate the clustering of the transcriptomic datasets on a species level, we estimated expression distances by applying an evolutionary model of transcriptome divergence<sup>68</sup>. The clustering largely recapitulates the known phylogeny (Fig. 1C). For all tissues except the larva head, the consistent difference between the species' genetic relationships and the transcriptomic clustering is the lack of an internal node shared by *D. erecta* and *D. santomea*. The transcriptomic clustering of the larval head dataset results in additional discrepancies, with *D. erecta* grouping with *D. melanogaster* and *D. simulans*, *D. sechellia* and *D. santomea* grouping together (Fig. 1C). This pattern points to a more complex evolutionary history of gene expression evolution for the larva head compared to the other tissues.

The distinct ecologies and evolutionary histories among these six species led us to hypothesize that their chemosensory transcriptomes have evolved at different rates. We tested for these differences by applying relative rate tests, which use a pair of ingroup species with an outgroup species to determine whether one of the two ingroup lineages has a significantly elevated rate of transcriptomic change. We applied this test to all 12,096 1:1 orthologs for all species-pairs (setting *D. sukukii* as the outgroup) and found that the distribution of test scores (Z-scores) for the majority of the species-tissue-sex comparisons are largely consistent with equal rates of transcriptomic change

across species, indicating that sensory transcriptomes exhibit low rates of divergence (Fig. 1D). We obtained consistent results when examining the distribution of Z-scores based on subsampled sets of the 1:1 orthologs (Fig. 1D) and when using either *D. erecta* or *D. santomea* as outgroup species to *D. simulans*, *D. sechellia*, and *D. melanogaster* (Supplementary Fig. 3).

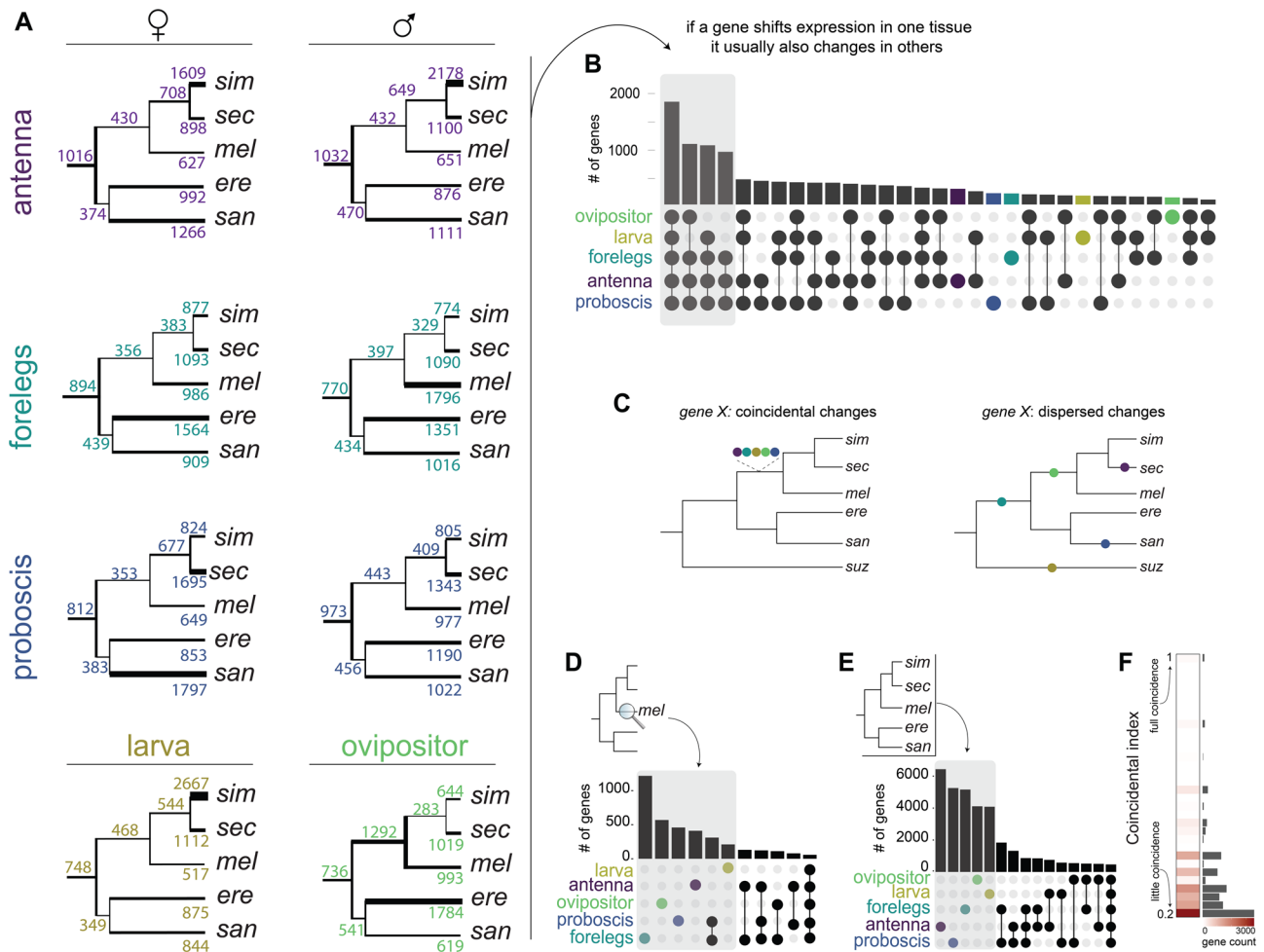
Although most tissue's transcriptomic divergence was low, we identified several tissues that stand out with elevated species-specific and sex-specific differences. *D. simulans* has a significantly elevated rate of evolution for its female antenna transcriptome (Wilcoxon signed-rank test  $V = 4.6e + 06$ ,  $p < 0.001$ ), larva head transcriptome (Wilcoxon signed-rank test  $V = 7.1e + 06$ ,  $p < 0.001$ ), and ovipositor transcriptome (Wilcoxon signed-rank test  $V = 2.9e + 06$ ,  $p < 0.001$ ). In

addition, *D. melanogaster's* male forelegs and ovipositor transcriptomes have significantly elevated rates of transcriptome change (Wilcoxon signed-rank test  $V = 6.6e + 06$ ,  $p < 0.001$  and  $V = 3.1e + 06$ ,  $p < 0.001$ , respectively). In contrast, the ovipositor transcriptomes from *D. santomea* and *D. erecta* were both found to have significantly lower expression divergence compared to the other species (Wilcoxon signed-rank test  $V = 3.1e + 06$ ,  $p < 0.001$  and  $V = 2.3e + 06$ ,  $p < 0.001$ , respectively). Overall, these global analyses of transcriptomic differences highlight a limited set of sensory tissues as rapidly evolving among the species, possibly reflecting key ecological and/or functional differences. They also provide evidence for significant sex differences that exist within species (see below).



**Fig. 1 | Chemosensory tissue transcriptome evolution.** **A** Overview of the chemosensory tissues and species used in this study. The numbers at the nodes of the species tree are the estimated divergence dates in millions of years. Antennas, proboscis+maxillary palps, forelegs, ovipositors and larvae transcriptomic datasets are labeled in purple, blue, cyan, green and dark yellow, respectively. **B** PCA of the transcriptomic datasets using 12,096 1:1 orthologs. The percentages on the axes are the amount of variation explained by the PCs. Each dot corresponds to one dataset. **C** Clustering of the transcriptomic datasets (12,096 1:1 orthologs) according to species and sex. Numbers above branches are bootstrap values for nodes with support <100. Species names are abbreviated to the first three letters. **D** Relative rate test results arranged by the species phylogeny. Tissues are arranged to align vertically with **C**. Colored shapes and lines display the mean and standard deviation of Z-scores from the full set of 1:1 orthologs. *D. suzukii*, noted with the dashed line and gray font, was used as the outgroup species. Gray data points are Z-scores that resulted from repeating the tests with subsampled datasets (Methods). Asterisks

denote the significantly elevated (positive values) or reduced (negative values) rates of gene expression change (Wilcoxon test comparing Z-score distribution to the minimum and maximum values of non-significant Z-scores; dotted lines). Only significant one sample right-tailed Wilcoxon tests are displayed (*D. simulans* female antennae,  $n = 4000$ ,  $p < 2.2^{-16}$ ; *D. melanogaster* male legs,  $n = 4000$ ,  $p = 3.73^{-288}$ ; *D. simulans* larvae,  $n = 4000$ ,  $p < 2.2^{-16}$ ; *D. simulans* ovipositor,  $n = 4000$ ,  $p = 9.65^{-7}$ ; *D. melanogaster* ovipositor,  $n = 4000$ ,  $p < 2.2^{-16}$ ) as well as significant one sample left-tailed Wilcoxon tests (*D. santomea* ovipositor,  $n = 4000$ ,  $p = 2.38^{-33}$ ; *D. erecta* ovipositor,  $n = 4000$ ,  $p = 6^{-125}$ ). Species names are abbreviated to the first three letters. **E** Fraction of expression tree branches (for all 1:1 orthologs and all tissues) found to best fit one of three evolutionary models: neutrally evolving, constrained, or divergent. Note that the divergent branches were found within otherwise constrained gene trees. Location of source data for this figure can be found in "Source\_data.xlsx".



**Fig. 2 | Expression changes over branches and tissues.** **A** Expression changes inferred across species for the five tissues and sexes. The number above each branch is the total number of expression changes (up and down), and the thickness of the branch is proportional to that number. **B** Quantification of the genes that change in expression across multiple tissues. The height of each bar indicates the number of genes that have changed expression across the set of tissues indicated by the darkened/colored circles. **C** Schematic illustrating a hypothetical *gene X* having expression changes involving multiple tissues that were coincidental (occurring on a single branch) or dispersed (occurring across multiple branches). In this illustration, the color of the dot indicates the tissues in which *gene X*'s expression changed. **D** Quantification of the coincidental expression changes along the branch leading to *D. melanogaster*. The height of each bar indicates the number

of genes that have changed expression across the set of tissues indicated by the darkened/colored circles. The plot was truncated at the bin containing the overlap of the five tissues. **E** Quantification of the coincidental expression changes over all branches in the phylogeny. The height of each bar indicates the number of genes that have changed expression across the set of tissues indicated by the darkened/colored circles. The plot was truncated at the bin containing the overlap of the five tissues. **F** Summary of the distribution of the “coincidental index” for all expression changes. A coincidental index of 1 indicates that a gene has changed in expression in all tissues within a single branch (full coincidence), decreasing values indicating less coincidence. Location of source data for this figure can be found in “Source\_data.xlsx”.

The observation that the chemosensory transcriptomes generally display low rates of divergence is suggestive of stabilizing selection but could also result from similar rates of neutral evolution over the relatively short timespans. We thus quantified the relative contributions of genetic drift and stabilizing selection to the evolution of gene expression. We did this by testing if the data for each gene expression tree is better explained by a phylogenetically-based model of neutral expression evolution or a model that assumes evolutionary constraint (with one or more branches experiencing expression divergence (Methods)). Across all gene expression trees, 52% of branches were inferred to have been under selective constraint with 8% of branches inferred to have diverged within an otherwise constrained expression tree. Approximately 40% of the branches display evidence of neutrally evolving expression (Fig. 1E).

### Genes change expression in multiple tissues but at different evolutionary times

As illustrated by the outliers in the relative rate tests and the detection of divergent branches in constrained expression trees, broad selective constraint has not precluded individual genes from evolving species-specific differences. Therefore, we investigated the genes that have changed in expression between species and when in the past the changes took place. Using phylogenetically-informed tests applied to our set of 1:1 orthologs, we detected several thousand differentially expressed genes for each tissue. Most of these expression changes occurred in only one species. The total number of expression changes ranges from 8499 (involving 6697 of the 1:1 orthologs) in male antennae to 7501 (involving 5927 of the 1:1 orthologs) in female legs (Fig. 2A, Supplementary Data 1). Analysis of the functional categories enriched by these differentially expressed genes highlighted combinations of developmental/morphological, neural/sensory, and gene regulation terms, among others, in varying proportions along extant



and past lineages (Supplementary Fig. 4, Supplementary Data 2). The elevated number of expression changes identified in *D. simulans* female antenna (1609) and larva (2667) and in *D. melanogaster* male forelegs (1796), confirms a history of elevated rates of expression evolution for these tissues (Fig. 2A; see also Fig. 1D).

Having identified gene expression changes across species for each tissue on its own, we next questioned how often a given gene changed expression in multiple tissues. Quantifying these overlaps revealed that genes that have changed expression in only one tissue are rare (~7%). Instead, most genes have changed in expression across multiple tissues, with the set of genes displaying changes across all five tissues being the largest set by almost twofold (Fig. 2B). Importantly, we find similar results for tissue overlaps and functional category enrichment when identifying differentially expressed genes using a standard alternative (non-phylogenetic) approach, confirming the robustness of our findings (Supplementary Fig. 5; Methods).

When genes have changed their expression across multiple tissues, this could have occurred simultaneously (e.g., as a result of pleiotropic mutations) or it could have resulted from the accumulation of tissue-specific expression changes at dispersed times in the past (e.g., as a result of the evolution of *cis*-acting regulators or of changes in cell abundances) (Fig. 2C). To gauge the importance of these two contrasting possibilities, we estimated the number of times a gene changed in expression across multiple tissues on individual branches of the phylogeny. Our analysis revealed very few coincidental changes. For example, on the branch leading to *D. melanogaster*, a vast majority of the expression changes occurred in only one of the five tissues (Fig. 2D). The same trend holds when summarizing expression changes over all branches of the phylogeny (Fig. 2E), as well as when quantifying the rate of coincidental changes (Fig. 2F). Collectively, these analyses imply most differentially expressed genes have evolved expression changes across different tissues at independent times in the past, consistent with independent evolutionary changes in gene regulation and/or cellular abundances.

Further inspection of the rare coincidental expression changes indicated that the probability of their occurrence is independent of branch lengths (Supplementary Fig. 6A). This finding confirms the intuition that many of these expression changes have arisen by pleiotropic mutations (and are not primarily a result of low resolution for detecting independent expression changes along longer branches). Interestingly, the most frequent coincidental change among all tissue combinations involved the forelegs and proboscis+palps samples (Supplementary Fig. 6B; see also Fig. 2D, E). This observation is coherent with the transcriptomes of these two tissues being the most similar among the five (Fig. 1B) and points to the likelihood that they share gene regulatory networks.

### Evolution of gene expression is often cell-type specific

Our finding that most differentially expressed genes across species are expressed in all or many of the tissues (Fig. 2B) led us to question if their expression specificity differs from genes that have not changed across species. We first compared the specificity of expression between differentially and non-differentially expressed genes at the tissue level and found that genes that have changed in expression have similar modes of tissue specificity but tend to be somewhat more tissue-restricted than genes that have not changed (Fig. 3A, B; Wilcoxon signed-rank test  $V = 4.3e + 10$ ,  $p < 0.001$ ). We then asked a similar question but at the level of cell types instead of tissues. Using the recently generated *D. melanogaster* single-cell atlases for antenna, legs, and proboscis<sup>33</sup> (Fig. 3A), we measured expression specificity at the level of cell types. We found that differentially expressed genes are significantly more likely to be expressed in a limited number of cell types than genes that have not changed in expression across species (Fig. 3B).

The relationship between tissue specificity and cell specificity varies substantially. For example, we identified differentially expressed genes that are expressed narrowly at the cell and tissue levels, e.g., the olfactory receptor *Or56a*, a receptor used by *Drosophila* to detect the harmful mold odor geosmin<sup>69</sup> (Fig. 3B). In contrast, we also identified genes that are expressed intermediately at the tissue level but are highly cell-specific within tissues. Using previous cell annotations<sup>33</sup> and marker-based cell type identification across the three atlases, we verified that these latter cases can be attributed to the same cell types being shared across tissues, e.g., *sosie*, a membrane protein localizing to mechanosensory cells and *rho*, a serine protease that localizes to glial cells (Fig. 3B). These examples illustrate how measurements of expression specificity using bulk tissues can mask the cell specificity of a gene's expression<sup>70</sup>. They also demonstrate that species' expression differences that likely underlie phenotypic divergence can be ascribed to individual cell populations.

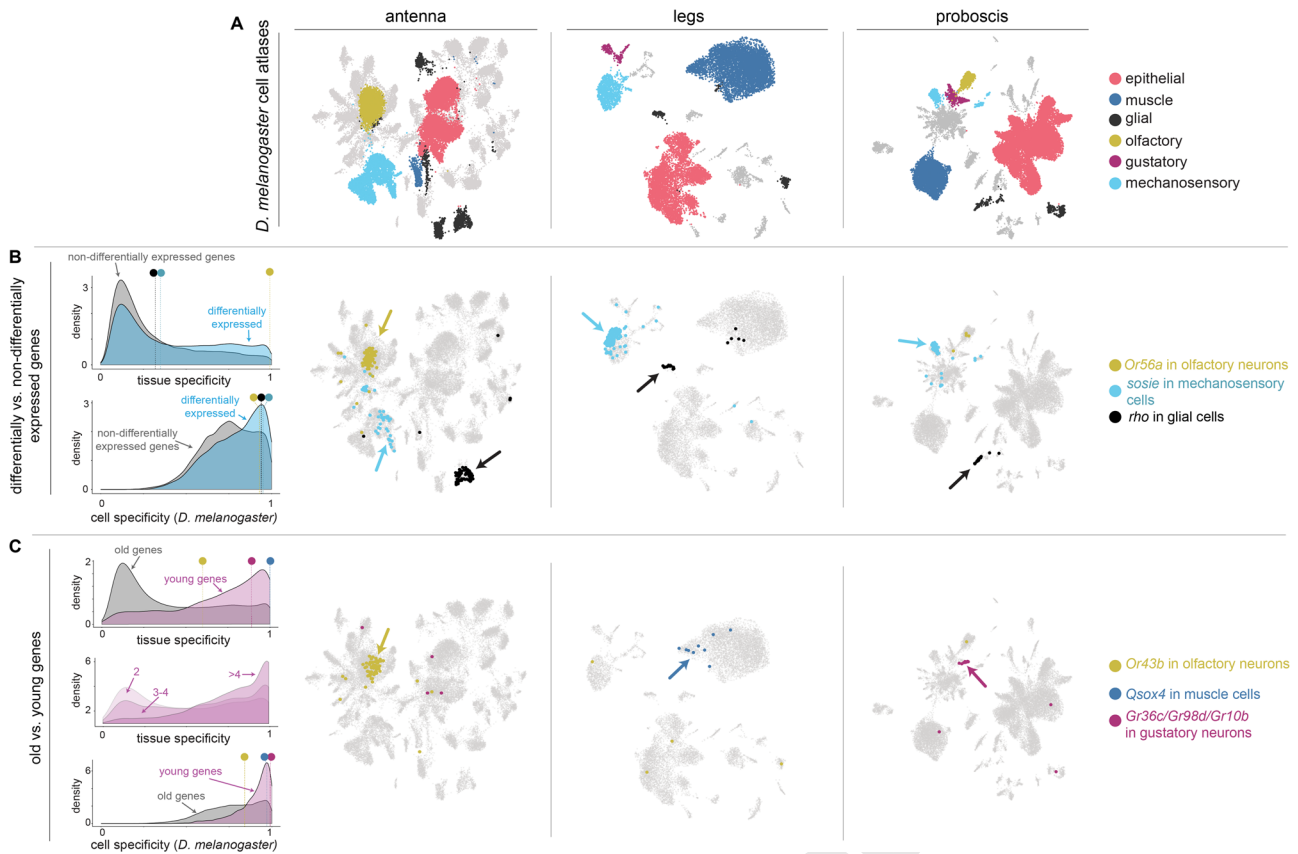
### New genes tend to cell-specific

New genes are a key source of evolutionary novelty<sup>71</sup>. Due to their potential contributions to species differences, we expanded our analyses to examine how gene age and duplication frequency relate to differences in the transcriptomes of sensory tissues. We compared the specificity of tissue expression between old genes (genes that predated the diversification of the *Drosophila* subgenus more than 50 million years ago) and new genes (genes that arose since). Consistent with previous work<sup>72–76</sup>, we found that new genes are significantly more likely to be expressed in fewer tissues than old genes ( $p < 0.001$ ; Fig. 3C). We also found that the more often a gene has been duplicated, the more tissue-restricted its paralogs are (Fig. 3C). We reasoned that the increased expression specificity of new genes would translate to their detection in a narrower number of cell types. We mapped the expression of new genes to the single-cell atlases for the antenna, legs, and proboscis and compared their expression specificity across cell types to that of old genes. Our analysis confirmed that new genes are indeed significantly more likely to be cell type specific than old genes (Fig. 3C; Wilcoxon signed-rank test  $V = 3.47e + 07$ ,  $p < 0.001$ ).

### Pervasive expression evolution of chemosensory genes

Insect genomes contain three large chemoreceptor gene families: odorant receptors (*Ors*), gustatory receptors (*Grs*), and ionotropic receptors (*Irs*)<sup>27</sup>. In addition, members of the chemosensory protein family (*CSPs*), and other diverse protein families, including the odor binding proteins (*Obps*), transient receptor potential channels (*Trps*), and pickpocket ion channels (*ppks*), are chemoreceptors or otherwise involved in the peripheral sensing of environmental chemicals<sup>27,77–79</sup>. The patterns of expression for most of these “chemosensory genes” have been mapped to specific tissues and cell populations in *D. melanogaster* and have provided the foundation for numerous functional and behavioral studies<sup>27</sup>. While multiple RNA-seq experiments have detected expression differences among developmental stages or species (or both) for chemosensory genes<sup>20,80–83</sup>, the heterogeneous combination of samples, experimental design, and sequencing approaches have limited evolutionary analyses. We, therefore, manually curated a set of 368 chemoreceptor genes for the above seven gene families and used our uniformly generated RNA-seq dataset to investigate how their expression patterns have evolved between species, tissues, and developmental stages (Supplementary Data 3).

Out of the 368 chemosensory genes, we detected expression for 299 in at least one of the tissues. In the antenna, proboscis, and forelegs samples, the expression patterns largely matched previous reports<sup>20,32,80,81,83–111</sup> (Supplementary Data 4). However, we detected few of the described *D. melanogaster* chemosensory genes in larva head samples, likely due to their very low expression levels. In each tissue, we identified a core set of genes that were expressed across all six species (antenna = 98, proboscis = 71, forelegs = 63, ovipositor = 28,



**Fig. 3 | Specificity of gene expression at the level of tissues and cells.** **A** Single cell atlases from *D. melanogaster* antenna legs and proboscis. Colors highlight the same cell types of interest across tissues. **B** Density plots for differentially and non-differentially expressed genes (leftmost panel) relative to expression specificity for tissues (top) and cell types (bottom). Colored circles with lines above the density plots indicate the expression specificity values of three genes (*Or56a*, *sosie* and *rho*) chosen to illustrate the varying relationships between expression specificity at the level of tissues and cells. Expression of the three genes has been mapped onto the *D. melanogaster* cell atlases (right three panels). **C** Density plots for old and young genes relative to expression specificity for tissues (top left) and cell types (bottom

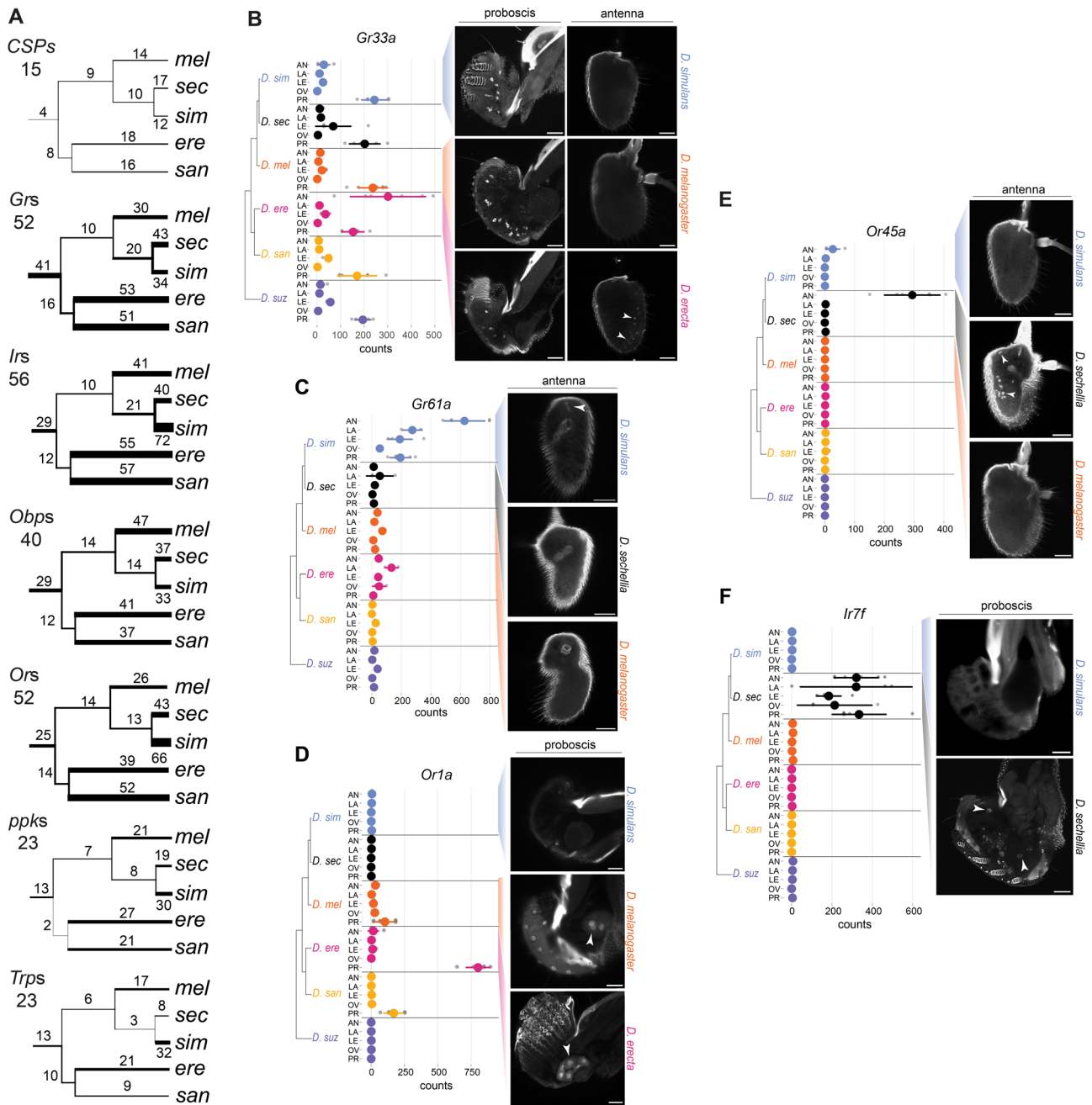
left). The middle left density plot shows the distribution of expression specificity values for genes grouped by duplication levels (2 = paralog group size of 2, 3–4 = paralog group size of 3–4, >4 = paralog group size greater than 4). Colored circles with lines above the density plots indicate the expression specificity values of three genes (*Or43b*, *Qsox4* and *Gr36c/Gr98d/Gr10b*) chosen to illustrate varying relationships between expression specificity at the level of tissues and cells (see text). Expression of the three genes has been mapped onto the *D. melanogaster* cell atlases (right three panels). Location of source data for this figure can be found in “Source\_data.xlsx”.

larva head = 37; Supplementary Data 5). Fourteen chemosensory genes were found to be expressed across all tissues, including two members of the *ppk* family, *ppk* and *ppk26*, which have previously been implicated in the detection of noxious mechanical stimuli in larvae. Their broad expression suggests additional sensory roles for these proteins in adults. The detection of multiple *Che* members in each of the tissues is also notable, given their suspected roles in detecting contact pheromones and pathogens have hitherto been limited to the legs<sup>98,112,113</sup>.

When we screened the set of 1:1 orthologous chemosensory genes for differential expression, we found that nearly all of them have evolved expression changes in at least one branch of the species tree (Fig. 4A, 93% of *CSPs*, 96% of *Grs*, 100% of *Irs*, 100% of *Ppks*, 100% of *Obps*, 98% of *Ors*, 100% of *Trps*; Supplementary Data 6) Furthermore, most genes have experienced recurrent expression changes, with the *CSP* family showing the greatest number. Consistent with genome-wide patterns (Fig. 2A), most expression changes were species-specific. Among the differentially expressed chemoreceptor genes, those that have gained or lost expression in a particular tissue were of particular interest because they may indicate novel gains (or losses) of sensory capabilities. We defined a gene with an average transcript per million (TPM) greater than 3 as expressed and genes with an average TPM less than 0.5 as unexpressed. Using these thresholds, we identified 95 chemosensory orthologs (32%) that have either gained or lost

expression in at least one tissue. Some of these expression gains/losses have occurred once, as illustrated by the gain of expression of *Gr98a* and *Gr98b* in *D. melanogaster*'s ovipositor or *Gr59e* in *D. erecta*'s larva, while others have involved recurrent changes, as in the foreleg-expressed *CheB74a* or the antenna-expressed *Ir31a* (Supplementary Figs. 7–13). Similar analyses of the set of 85 chemosensory genes (from 31 gene families) that have duplicated since the common ancestor of the six species revealed that nearly all recent paralogs have retained expression in the same tissues but often at lower levels Supplementary Figs. 14–15, Supplementary Data 7). Two exceptions are *Gr59a4* and *Ir52f2* in *D. sukukii*, which show gains of increased expression levels that may indicate the neofunctionalization of these genes (Supplementary Fig. 16).

To gain spatial and cellular resolution for the expression of a subset of 95 chemosensory genes with novel species-specific expression patterns, we designed in situ hybridization chain reaction (HCR) experiments for six of them: *Gr32a*, *Gr33a*, *Gr61a*, *Ir7f*, *Or1a*, *Or45a* (Methods). We detected expression that was consistent with our RNA-seq results for all of these genes except *Gr32a* (Fig. 4B–F, Supplementary Fig. 17). For unknown reasons, we were unable to detect *Gr32a* in *D. sukukii* antenna despite detecting its expected expression in the labial palps (Supplementary Fig. 18). Additional co-labeling experiments using probes for other cell type markers resulted in the



**Fig. 4 | Evolution of chemosensory gene expression.** **A** Expression changes mapped onto the species tree for genes belonging to the main chemosensory gene families (*Grs* gustatory receptors, *Irs* ionotropic receptors, *Ors* odorant receptors, *CSPs* chemosensory proteins including *CheA* and *CheB* family members, *Obps* odorant binding proteins, *ppk* pickpocket, *Trps* transient receptor potential channel). The number above each branch is the total number of expression changes (up and down) across tissue samples, with the thickness of the branch proportional to this count. The number under the gene family name in parenthesis corresponds to the number of 1:1 orthologs used for the analysis. **B–F** In situ HCR results for chemosensory genes with a species-specific gain of expression. On the left is the

species tree (not to scale) with the mean normalized read counts obtained for each sample. The mean and standard error are represented by the colored dots and the vertical line, respectively, with the individual data points in gray. AN antenna, LA larva head, LE forelegs, OV ovipositor, PR proboscis. RNA in situ hybridizations on the right with the targeted tissue(s) above each column. White arrows indicate cells with species-specific receptor expression. Scale bars: 30  $\mu$ m. See also Supplementary Figs. 17–18. In **C** and **E**, autofluorescence highlights the sacculus in the three species. Location of source data for this figure can be found in “Source\_data.xlsx”.

discovery of unexpected patterns of cellular expression including for the two gustatory receptors, *Gr33a* and *Gr61a*. In *D. melanogaster*, *Gr33a* is characterized as a bitter receptor expressed in taste cells in the legs and proboscis and involved in aversion to male-male courtship<sup>102,114,115</sup>. We found that expression of *Gr33a* has expanded from bitter taste neurons into olfactory sensory neurons (*Orco*+) in the antenna of *D. erecta* (Fig. 4B, Supplementary Fig. 17A).

Interestingly, antennal expression of *Gr33a* was previously observed in *D. melanogaster* when programmed cell death was experimentally blocked in olfactory sensory neurons<sup>116</sup>, possibly indicating a *D. erecta*-specific developmental “escape” from cell death for this neuron population. Analogously, *Gr61a*, a glucose receptor in *D. melanogaster* that is expressed in neurons in the labellum, legs, and the labral sense organ<sup>117,118</sup>, was also found to have expanded into olfactory neurons of



*D. simulans*' antenna (Fig. 4C, Supplementary Fig. 17B). Analyses of two odorant receptors, *Or1a* and *Or45a*, also revealed species-specific expression patterns. We found that *Or1a*, which was previously described as being larva-specific in *D. melanogaster*<sup>104</sup>, is expressed in non-neuronal cells that are likely part of the labellar glands in *D. melanogaster* and *D. santomea* (Fig. 4D, Supplementary Fig. 17C). To our knowledge, no chemosensory function for this gland has yet been described. We found that *Or45a*, previously described as larva-specific in *D. melanogaster*<sup>109</sup>, is also expressed in the adult antenna in *D. sechellia* (Fig. 4E, Supplementary Fig. 17D). Finally, *Ir7f*, which has yet to be functionally characterized, was one of the most distinct differently expressed chemosensory genes because it has uniquely gained high expression in all chemosensory tissues in *D. sechellia* (an example of a "coincidental" gain of expression). We observed expression of *Ir7f* within cells that also express a pan-neuronal marker (*nsyb*) in the labial palps, indicating that this gene is likely encodes a taste receptor (Fig. 4F, Supplementary Fig. 17E). Together, these expression analyses underscore the remarkable evolutionary flexibility in transcript abundance, developmental timing, and spatial expression of chemosensory genes.

### Fast evolution of sex differences

Next, we identified sex differences in our dataset and placed them in a phylogenetic context. *Drosophila* chemosensory tissues are involved in sex-specific functions and often vary between the sexes in morphology and neuroanatomy<sup>19,67,119–121</sup>. While previous single gene and transcriptomic analyses identified sex differences in gene expression within some of these tissues<sup>20,81,82,122</sup>, their evolutionary histories between tissues and species remain unclear.

For each species, we computed the number of genes with significantly different expression levels between males and females ( $\geq 1.5$ -fold change with adjusted  $p < 0.01$ ) within our proboscis + palps, antenna, and foreleg datasets and examined their variation among the six species. Our analysis revealed extensive evolution in the number of sex-biased genes across species, the proportion of genes having male-versus female-biased expression, and in the identity of the sex-biased genes (Fig. 5A; Supplementary Data 8). Remarkably, the patterns of sex-biased gene expression do not reflect the genetic relationships among the species, in line with previous findings that expression differences between the sexes evolve quickly<sup>123–127</sup>. We observed an approximately ten-fold difference in the number of sex-biased genes between the species with the fewest and the species with the most (*D. santomea* and *D. sechellia* with 135 and 178, respectively, versus *D. erecta* and *D. melanogaster* with 1350 and 1132, respectively). Although the number of male-biased genes outnumbers female-biased genes (2098 vs. 1287), this ratio varied considerably across tissues. Genes expressed in the forelegs and the proboscis are mainly male-biased, while female-biased genes are predominant in antennae. These results suggest that different modes of sexual selection may have shaped the male/female expression balance in a tissue-specific manner.

We then asked if the identity of sex-biased genes is shared across species and tissues. These analyses once again highlighted pervasive variation in the sets of genes that differ between the sexes. In a majority of cases, the sets of female- and male-biased genes are private to each species (Fig. 5B). Intriguingly, among the few overlaps between species, we found enrichments of genes involved in or activated by cell-autonomous and non-autonomous control of sex differences (including *fruitless*, *doublesex*, *insulin-like peptide 7*, and members of the cytochrome P450 family; hypergeometric tests  $p < 0.001$ ), suggesting they may play roles in the maintenance of sexually dimorphic traits in adult sensory tissues, similar to what has been observed for *D. melanogaster*'s intestine<sup>21</sup>. We also observed enrichment in chemosensory proteins among conserved sex-biased genes (hypergeometric tests  $p < 0.001$ ) which have been shown to be sex-biased and involved

in pheromone-induced behaviors<sup>98,112</sup>. Finally, within species, if a gene is sex-biased in one tissue, it is rarely sex-biased in the other two (Fig. 5C).

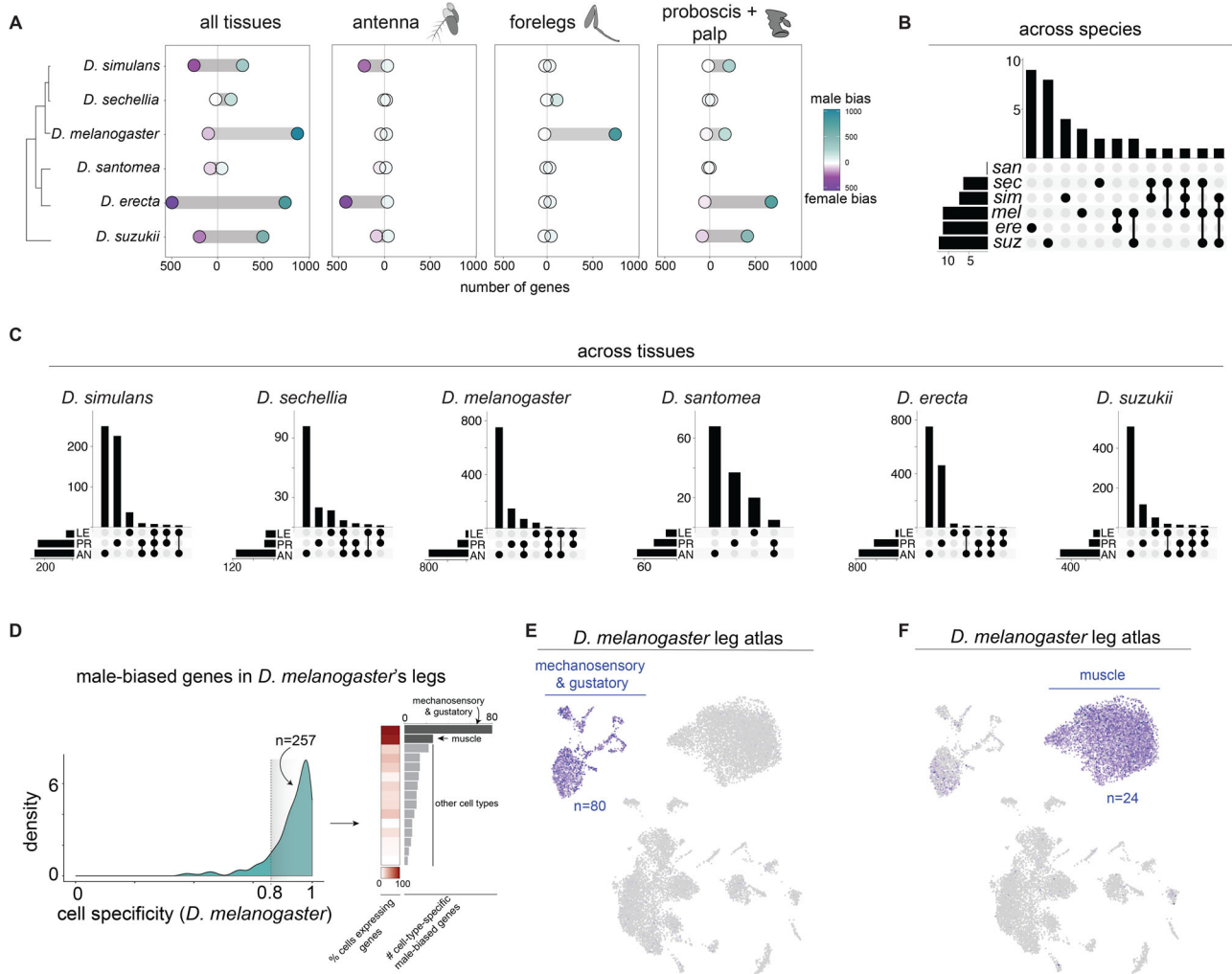
We sought further insight into the cell types underlying the derived *D. melanogaster* male-biased foreleg expression. Of the 806 male-biased genes identified in our RNA-seq experiment, 285 were detected in the leg samples of the Fly Cell Atlas<sup>33</sup>. Examination of the cell-type-specificity of the 285 genes revealed that most are expressed very narrowly (number of genes with cell specificity  $> 0.8 = 257$ ) and are enriched in cell populations related to mechanosensation, gustation, and muscle (Fig. 5D–F). These cell types are particularly compelling in light of the extensive literature identifying key roles for these sensory modalities in *D. melanogaster*'s courtship<sup>22,128–130</sup> and because musculature is dimorphic among *Drosophilids*<sup>131</sup>. Sexually dimorphic genes in muscle include those with functions in mitochondrial respiratory (*ND-B8*), actin assembly (*forked*), and vision (*Rh2* and *Culd*). Sexually dimorphic genes in the mechanosensation/gustation cell populations include a putative pheromone receptor (*Ir52c*), two *trp* channels involved in temperature sensing (*brv3* and *pkd2*), and several genes involved in neuron development and signaling (e.g., *Unc-104* and *Stathmin*).

Sex-bias expression that is detected in bulk tissue samples could result from differences in cell abundances between the sexes, transcript abundance differences between the sexes, or a combination of both. We examined these possibilities using the sex-specific Fly Cell Atlas leg data, which includes two pooled male and two pooled female samples. Though preliminary due to the limited number of replicates, we found that the population of muscle cells enriched for male-biased genes (identified in the bulk RNA-seq analysis) is more abundant in the male sample compared to the female sample. We also found higher mean expression levels for the male-biased genes (identified in the bulk RNA-seq analysis) in the male sample compared to the female sample (Wilcoxon test  $p < 0.01$ ). No differences in cell abundance of expression levels between the sexes were identified in the mechanosensation/gustation cell population (Supplementary Fig. 19). These results suggest that, at least in the muscle cells, regulation of the cell population size and transcript abundance have both contributed to the sexual dimorphism in *D. melanogaster*. Additional male and female samples, along with cross-species single-cell data sets will provide more rigorous tests of these observations.

## Discussion

By conducting comparative transcriptomic analyses of chemosensory tissues across species and linking them with single-cell datasets and additional experiments, we have expanded our understanding of how these sensory systems evolve on a global and individual gene level. Globally, we have found that stabilizing selection has been the dominant evolutionary force shaping chemosensory transcriptomes, with ~60% of the branches on the 1:1 ortholog expression trees displaying evidence of evolutionary constraint. Because protein abundances are more directly relevant to phenotypic changes than mRNA levels, we suggest there is additional stabilizing selection at the level of translation. Support for this comes from cross-species comparisons of co-profiled translomes and transcriptomes which have found fewer between-species changes at the level of the translome compared to the transcriptome, reflecting additional stabilizing selection for protein abundance<sup>132</sup>. Nevertheless, evolutionary constraint has not precluded a subset of tissues and genes from experiencing accelerated rates of expression change. At the transcriptomic level, *D. melanogaster* (forelegs and ovipositor) and *D. simulans* (antenna, larva head, and ovipositor) are distinct for having significantly increased expression divergence. This was initially curious as the two are ecological generalists while the other species have evolved ecological specializations. However, it is consistent with *D. melanogaster* and *D. simulans* having the largest effective population sizes (and likely substantially so)<sup>47,51</sup>





**Fig. 5 | Evolution of sex-biased expression in chemosensory tissues.** **A** Number of male- and female-biased genes across species and tissues. Sex-biased gene expression does not match the species' phylogenetic relationships demonstrating many species-specific changes. **B** Overlap across species (regardless of tissue) for the number of genes that share the same direction of sex-bias (male or female), illustrating that sex-biased genes are species-specific. Species names are abbreviated to their first three letters. **C** Number of sex-biased genes that overlap across tissues within species. Most sex-biased genes are tissue-specific (LE = forelegs, AN = antenna, PR = proboscis+palps). **D** (right) Density plot for *D. melanogaster*'s

male-biased genes relative to cell-type specificity. (left) Heat map showing the fraction of cells in a given cell population that express the male-biased genes and bar plot displaying the total number of male-biased genes found expressed within a given cell population. Most male-biased genes are cell type specific and predominantly found within cells associated with mechanosensation, gustation, and muscles. **E, F** Cell atlas for *D. melanogaster* legs with the total mean expression of 104 male-biased genes displayed, highlighting their restricted expression in mechanosensory/gustatory and muscle cells. Location of source data for this figure can be found in "Source\_data.xlsx".

resulting in positive selection playing a greater role within these species compared to the others. If true, this result would suggest an important role for positive selection in driving gene expression changes.

At the level of individual genes, we have identified numerous instances of significant expression differences across species for each of the five chemosensory tissues. We estimate that ~8% of all expression tree branches display changes consistent with adaptation. A previous estimate based on a "fitness seascape" model of gene expression evolution inferred that ~60% of expression divergence among (partially overlapping) *Drosophila* species was driven by positive selection, concluding that adaptation is the principle evolutionary force underlying the changes<sup>133</sup>. Though seemingly at odds with our observations, we propose that this previous estimate is comparable to ours if we summarize our results at the level of genes instead of trees. If we summarize the percentage of gene expression trees that have at least one divergent branch, we find that ~50% have experienced historical change(s) in expression optima that may be consistent with adaptation

(Supplementary Fig. 20). However, a gene-focused summary obscures our finding that most branches in the expression trees are under constraint. It is also noteworthy that most of the expression changes we have identified have occurred in only one species, indicating that most differences are recent. As it becomes more feasible to carry out population surveys for expression polymorphism, it will be important to quantify how many of these changes are fixed between species and how many are polymorphic<sup>126,134</sup>.

The expression changes that we identified could have resulted from differences in transcript abundance (e.g., *cis*-regulatory changes) or cellular composition (e.g., expanded or contracted cell populations). Though we cannot separate these possibilities with bulk tissue samples, the fact that most changes occurred in one tissue ("dispersed") supports an evolutionary model of modular change. We suggest that the same argument applies to the cell specificity of young genes (Fig. 3C). Both observations are important because a key factor in determining anatomical evolution is the pleiotropy of mutations. Due to the functions that individual genes have across multiple tissues,

it is expected that the diversification in any one tissue (or subset) will arise through mutations in the genes' modular cis-regulatory regions<sup>135,136</sup>. To the extent that transcript abundance drives the differences in our datasets, our results are consistent with previous findings that indicate that most between-species expression changes are driven by cis-regulatory modifications<sup>133,137–140</sup>. We expect that the close relationships between these species will foster the identification of candidate regulatory differences that can be studied to further understand the molecular basis of transcript abundance evolution. Much less is known about the evolution of cell population sizes. In the case of *D. melanogaster*'s species-specific male-biased foreleg expression, we have found preliminary evidence that both transcript abundance and cell abundance evolution may be involved. We will soon be able to address this question more thoroughly through cross-species comparisons of single-cell atlases.

Molecular evolutionary studies of chemosensory genes have consistently highlighted their rapid protein coding and copy number evolution<sup>141–144</sup>. Our analyses demonstrate that changes in transcript abundance and species-specific expression gains and losses also fuel their fast evolution. It has been suggested that the cell-specific expression patterns of most chemosensory genes, along with partially overlapping molecular functions (e.g., promiscuous ligand-binding), result in relatively fewer pleiotropic constraints and, as a result, increased evolutionary freedom to change<sup>141</sup>. It is likely that their narrow cellular expression also allows for increased flexibility to fine-tune their levels of expression. Though the phenotypic implications of chemoreceptor expression levels remain unclear, it is plausible that they shape neuronal sensitivity or other cellular kinetics that impact a neuron's encoding of chemical information. We also have evidence for several peripheral sensory neuron population that they can expand/contract quickly<sup>54,145,146</sup> and are likely contributing to species differences in chemoreceptor expression levels. More comprehensive studies are needed to assess how frequently such changes are occurring. Of potentially greater immediate phenotypic consequences are chemoreceptors' ability to gain (or lose) expression in novel tissues. We estimated that approximately a third of the chemosensory genes may have done so over the diversification of these six species. And while instances of unusual or "ectopic" receptor expression, as illustrated by *Or1a* (Fig. 4D), call for additional functional characterization, they are also a reminder of the first step that all receptors and receptor operated channels have taken as they have diversified across tissues throughout the animal kingdom.

As with other comparative functional genomic studies, identifying the specific changes that are translated into phenotypic differences remains an outstanding challenge. The phylogenetic framework provided here will help to devise future experiments for addressing this question, as illustrated by our investigation of five chemosensory genes with species-specific expression patterns. One line of evidence pointing towards a substantial fraction of the expression differences being functionally important is our observation that they tend to be cell-specific. Though it is conceivable that a similar trend could be produced by neutral evolution (e.g., expression drift being more common among sets of genes that are cell-type-specific), we argue that this observation nonetheless provides important genome-wide evidence consistent with them being functionally relevant. This is most convincing in the context of sex differences, where nearly all expression changes are species-specific and where, in *D. melanogaster*, we linked species-specific sex-biased genes to specific cell populations involved in sexually dimorphic functions<sup>22,128–130</sup>.

## Methods

### Fly strains, rearing, and dissections

*D. melanogaster* (NDSSC, GDL B54<sup>147</sup>), *D. simulans* (NDSSC,14021-0251.008), *D. sechellia* (NDSSC,14021-0271.07), *D. santomea* (NDSSC,14021-0271.00), *D. erecta* (NDSSC,14021-0224.01) and *D.*

*suzukii* (Kyorin, K-AWA036) flies were reared on a standard yeast/cornmeal/agar medium supplemented with Carolina 4-24 Formula and maintained in a 12:12 h light:dark cycle at 25 degrees. Adults between 2 to 10 days old were sex-sorted on CO<sub>2</sub> at least 24 h before the dissections. Third instar larvae were taken directly from the food medium the day they were dissected. For each replicate, 10 third instar larval heads, 25 proboscis, 50 legs, 5 ovipositors, and ~100 antennae were collected. Three replicates were made per sex and species for the proboscis, the legs, and the antennae; 3 replicates were made per species for ovipositors and larval heads.

### Tissue collection

All adult samples were collected from flies aged between 2-10 days. Antennas were collected by flash-freezing flies in liquid nitrogen and agitating them over a mini-sieve connected to a collection dish<sup>148</sup>. Antennas were selected from the collection dish using a pipette under a dissecting scope. Forelegs and proboscis with maxillary palps were collected from individual flies using forceps and a micro scalpel under a dissecting scope. Though the proboscis and maxillary palps are distinct appendages, we combined them to reduce library and sequencing costs. Third instar larvae were collected from vials by floating them in 75% sucrose water and washed. Larva heads were removed under a dissecting scope using a micro scalpel.

### mRNA library preparation and sequencing

Dissected tissues were homogenized in 200ul of Trizol (Invitrogen) using a Precellys24 (6800 rpm, 2x30s with 10 s breaks; Bertin Technology) followed by a standard Trizol RNA extraction. The final mRNA concentration was measured using a DeNovix Ds-11 FX spectrophotometer. mRNA libraries were prepared using KAPA Stranded mRNA-seq Kit (Roche) following the manufacturer's instructions (Version 5.17). Briefly, 500 ng of total RNA diluted in 50ul of RNAase-free water was first placed on supplied mRNA capture magnetic beads to allow the isolation of mature, polyadenylated mRNA, which was subsequently fragmented to a size of 100–200 bp. Double-strand cDNA was then synthesized, marked by A-tailing and barcoded with 2.5 ul of TruSeq RNA UD Indexes (Illumina). SPRI select beads (Beckman Coulter) were used for cleanup. Library concentrations were measured using Qubit dsDNA HS Assay Kits (Invitrogen). Fragment analysis and HiSeq 4000 single-end Illumina sequencing were performed by the Lausanne Genomic Technologies Facility. Information for all molecular reagents used in this project can be found in Supplementary Data 9.

### In situ hybridization chain reaction experiments

**Gene choice:** *Gr32a*, *Gr33a*, *Gr61a*, *Ir7f*, *Or1a*, and *Or45a* were chosen based on information on their previous functional characterization and/or expression in *D. melanogaster* (see Supplementary Fig. 17 legend) and because we found their species-specific expression differences to be the most intriguing in light of existing data.

**Probe sets:** HCR probes set, amplifiers, and buffers were purchased from Molecular Instruments. The list and the sequences of the probes used can be found in Supplementary Data 9. Coding sequences and 5' and 3'UTRs, were extracted from the species reference genomes and aligned. *D. melanogaster* sequences were used to design HCR probe sets for genes sharing >91% identity across our target species. If sequence identity was less than 91%, or if we failed to detect a signal using a *D. melanogaster* probe set in a different species where transcripts were detected in our RNA-seq dataset, we designed species-specific probe sets. Based on these criteria, species-specific probes were designed for *D. simulans Gr61a*, *D. suzukii Gr32a*, *D. sechellia Ir7f*, and *D. suzukii Gr66a*.

**In situs:** Flies between 2 to 9 days old were cold anesthetized and dissected on ice. Samples were collected on PBT (1XPBS, 0.1% Triton X-100) and fixed in 2 ml of a 4% paraformaldehyde, 1X PBS, 0.1% Triton

X-100 solution at 4 °C on a rotator set at low speed (<20 rpm) during 2 h for antennae, 4 h for *D. sechellia* proboscis and 24 h for other the species' proboscis. Following fixation, samples were washed twice in PBS + 3% Triton X-100 and three times in PBT. The protocol suggested by Molecular instruments for generic samples in solution was then followed with minor adjustments (<https://files.molecularinstruments.com/MI-Protocol-RNAFISH-GenericsSolution-Rev9.pdf>). Samples were pre-hybridized in 300 µl of probe hybridization buffer for 30 min at 37 °C. For antenna samples 3.5 µl of control probe (*Orco* or *nsyb*) and 5 µl of experimental probes were used. For proboscis, 5 µl of control (*nsyb*, *Gr66a*) and experimental probes were added to the amplification buffer. Samples were also pre-amplified in 300 µl of amplification buffer. For antenna samples 6 µl of hairpin solution designed to amplify the signal of control probes was used, 10 µl otherwise. For proboscis samples, 10 µl of hairpin solutions were used to amplify both the controls and the experimental probes. After washes, samples were mounted in Vectashield and stored at 4 °C. Information for all molecular reagents used in this project can be found in Supplementary Data 9.

**Image acquisition:** Antennae, proboscis and larvae images were acquired on inverted confocal microscopes (Zeiss LSM 710 or LSM 880) equipped with an oil immersion 40X objective (Plan Neofluar 40X oil immersion DIC objective; 1.3 NA). The images were processed in Fiji (v1.53)<sup>149</sup>.

### Gene annotations

Annotations in General Feature Format were generated for all species using BRAKER v2.1.6 and Augustus v3.4.0<sup>150,151</sup>. We ran BRAKER with the `-etpmode` flag as we provided evidence from both our aligned RNA-seq data and an orthologous protein dataset for arthropods (arthropoda\_odb10). The quality of annotations was checked with BUSCO v3.0.2<sup>152</sup>. First, we generated fasta files with coding sequence from the annotations using Cufflinks v2.2.1<sup>153</sup> `gffread` function (`-w exons.fa -W -F -D -E -o filtered.gff` flags). Completeness was checked against the diptera\_odb9 dataset. BUSCO scores were similar across species: *D. simulans* 97.3%, *D. melanogaster* 97.1%, *D. erecta* 97.0%, *D. santomea* 94.5%, *D. sukukii* 91.9%. The species' GTFs are in Supplementary Data 10.

### OrthoFinder-based orthology analysis

Our next goal was to group our annotated sequences into their respective orthologue groups using OrthoFinder v2.3.8<sup>154</sup>. The input peptide sequence was generated for each species by the following steps: (1) fasta files of coding sequence from annotations were converted to peptide sequence using the `transeq` function from EMBOSS v6.6.0<sup>155</sup>, (2) duplicate genes introduced from BRAKER's pipeline were removed using a custom script (`rmduplicategenes.sh`), (3) OrthoFinder's `primarytranscript.py` was run on each of the resulting peptide fasta files. These input peptide sequences were then placed in the same directory and we ran OrthoFinder to generate our orthologue groupings. We additionally added the `-M msa` flag to generate gene trees.

### Opposum-based orthology analysis and gene IDs

We used Possvm<sup>156</sup> (v1.1) to refine orthology relationships (1 to many and many to many) inferred by OrthoFinder (above). We first aligned non-1:1s orthologs using MAFFT<sup>157</sup> (v7.490; `mafft --auto protein.fa`) and outputted alignments in phylip format. We then generated phylogenetic trees containing bootstrap information at each node using IQ-TREE<sup>158</sup> (v2.2.0.5; `iqtree2 -s./MAFFT_ortho/{spe} -mset WAG, LG -b 200`), testing for the best substitution model (WAG or LG) and performing 200 bootstrap replicates. Finally, we used Possvm to identify new orthogroups. For this step, we first parsed phylogenies using the species overlap algorithm, and second, we clustered orthogroups using the MCL clustering method. We updated the former orthoFinder Orthogroup.tsv with the list of newly generated

orthogroups which included 2,066 new 1:1s genes. Orthogroups were renamed according to the *D. melanogaster* reference genes, which were identified through iterative BLAST (v2.10.1+)<sup>159</sup>. For this step, we used `tblastn` to query our list of protein orthogroups on a *D. melanogaster* gene database containing nucleotide fasta from all annotated CDS. BLAST results were sorted according to their best hit (bit score selection) and the matching gene names were appended to our inferred orthogroups IDs. This "lookup table" is available as Supplementary Data 11.

### RNA-seq read mapping

Each species' Illumina reads were mapped to its own soft-masked reference genome using STAR (v2.7.8<sup>160</sup>), inputting the GTF files generated above. On average we obtained 43 million mapped reads per replicate, with mapping rate modes ranging between 0.79–0.90. A single *D. simulans* proboscis sample resulted in fewer mapped reads due to the amplification of a viral sequence, but was otherwise highly correlated with the two other replicates and was therefore retained. Sample replication across all tissues was high, with an overall average Pearson correlation coefficient of 0.98; the range of Pearson correlation coefficients within each tissue's replicates was between 0.97–0.99. The one exception was the ovipositor dataset, likely reflecting less precise dissections (above). Pearson correlation coefficients for the ovipositor samples ranged from 0.93–0.99, with the replicates of *D. erecta*, *D. santomea*, *D. sechellia* being more variable (average Pearson correlation coefficients = 0.93, 0.95, 0.96, respectively) than the other three species (average Pearson correlation coefficients = 0.97, 0.97, 0.99). QC files and plots can be found within the project's repository (<https://gitlab.com/EvoNeuro/sensory-naseq>), in particular, see the `Full_Data_normPCA_trimmed.html` and `plot_mapping_stats_good_samples.html` files.

### Read count table generation

**Full-length gene:** Expression count tables were generated using HTseq (v0.11.2<sup>161</sup>), inputting the GTF files generated above (Supplementary Data 12; the corresponding TPM table for the 1:1 orthologs is File Supplementary Data 13).

**Trimmed genes:** Despite the six species being closely related, differences in orthologous gene lengths exist. If unaccounted for, these differences may lead to misleading cross-species differential expression results when using methods that assume identical gene lengths. To account for length differences in our PCA or clustering analyses and for analysis using DESeq2 (v1.34.0<sup>162</sup>), we generated count tables based on orthologous gene regions that were conserved across all six species. Conserved regions were identified based on DNA alignments (MAFFT v7.475<sup>157</sup>) of the 1:1 orthologs. We excluded gene regions if any of the six species contained a gap greater than 150 bp (using the script `get_aligned_blocks.py`). Using the coordinates of the conserved gene regions, we then generated a set of "trimmed" GTF files (using the script `make_trimmed_gtf.py`; the species' trimmed GTFs are found in Supplementary Data 14) that were passed to HTseq (v0.11.2<sup>161</sup>) for computing the "trimmed" count tables (Supplementary Data 15). The "trimmed" GTF file includes the full set of genes that were annotated in each species' genome but contains the modified genic coordinates based on the conserved alignments for the set of 1:1 orthologs.

The normalized count data for 1:1 orthologs can be explored and plotted with our CT<sup>2</sup> dashboard available at: <https://ctct.unil.ch/>.

### Transcriptomic clustering and Relative rate tests

Transcriptomes were clustered by species using the set of 1:1 orthologs and a phylogenetically-informed distance measure implemented in TreeExp (v0.99.3<sup>68</sup>). TreeExp implements a statistical framework assuming that gene expression changes are constrained by stabilizing selection (based on the Ornstein-Uhlenbeck (OU) model). For phylogenetic reconstruction, we generated "taxa.objects" from our TPM



normalized expression matrix specifying taxa (species) and sub-taxa (tissue) levels. Distance matrices were computed for each tissue by modeling gene expression changes under a stationary OU model (method="sou"). Finally, distance matrices were converted into phylogenetic trees using the neighbor-joining method, setting *D. suzukii* as an outgroup and performing 100 bootstrap replicates.

Relative rate tests were carried out in TreeExp (v0.99.3<sup>68</sup>) for all pairwise comparisons using its RelRate.test function. For these analyses, only genes with a TPM > 1 were included. To confirm that divergence score estimations were not driven by a subset of genes as well as to give stronger statistical power to the analysis, we computed divergence Z-scores by randomly sampling 1000 genes and 1000 times for each species pair and each tissue sample. We compared the per species per tissue Z-score distribution from randomly sampled genes to both the minimum and maximum value of the non-significant Z-score distribution using a Wilcox test statistic in R (v4.1.2<sup>163</sup>).

### Differential expression for 1:1 orthologs

Evolutionary changes in gene expression were detected using the lIou R package (v1.43<sup>164</sup>). lIou uses a phylogenetic lasso method to detect past changes in the expected mean trait value, assuming traits evolve under an Ornstein–Uhlenbeck (OU) process. We used a reference species tree that was previously inferred<sup>53</sup> and the species' mean TPM for each gene, for each tissue, as the evaluated traits. We set the maximum number of possible expression changes to 3 (half the number of taxa in the tree) and selected the best model for the number of expression changes using the phylogenetic-informed BIC approach (pBIC). We further tested the most likely evolutionary scenario for individual gene expression trees using EvoGeneX (v0.9.9.0<sup>165</sup>). Using both interspecific and intraspecific (within species triplicates) information, EvoGeneX carries out tests between three models of expression evolution: (1) neutral evolution (Brownian motion), (2) constrained evolution with one expression optimum (OU model), and (3) constrained evolution with one or more gene expression optima (OU model with multiple optima values). We found that gene expression changes detected by the lIou method strongly agreed with the most likely evolutionary scenario inferred by EvoGeneX (Supplementary Fig. 21). This analysis enabled us to quantify the number of neutrally evolving, constrained, and divergent branches for each expression tree (Fig1 E).

**Coincidental index:** For each gene, we calculated the frequency that it changed in expression in multiple tissues simultaneously by computing a simple "coincidental index" defined as:

$$\sum n_{obs(t)} / \sum n_{max.pos(t)} \quad (1)$$

Where  $n_{obs(t)}$  is the number of tissues an expression change occurred at time  $t$  and  $n_{max.pos(t)}$  the maximum number of possible changes at time  $t$ . This index takes a value between 0 to 1 where 0 reflects no expression changes, 0.2 reflects a change that occurred in only one tissue (dispersed) and 1 an expression change that occurred simultaneously in all tissues (coincidental).

**DESeq2 analyses:** Pair-wise based identification of differentially expressed genes was carried out with DESeq2 (v1.34.0<sup>162</sup>) specifying the following design: ~ 1 + species + tissue + species:tissue. For these tests, the set of 1:1 orthologs (above) and the "trimmed" count tables (above) were inputted.

**Gene module analyses:** We identified co-expressed gene modules between tissues using soft-clustering algorithms implemented in CEMITools (v1.18.1<sup>166</sup>).

### Analyses of gene duplicates and expression

We performed gene age analyses on gene lists derived from ref. 167. Genes predating the speciation of the *Drosophila* subgenus (~50 My ago) were classified as "old", while new genes that have emerged since

the *Drosophila* subgenus speciation event were classified as "young". Duplicated genes and their level of duplication are derived from our ortholog annotation on the set of non-1:1 orthologs (Supplementary Data 16).

### Manual curation of chemosensory gene set

Chemosensory genes were first extracted from the look-up table generated for the global dataset (Supplementary Data 16). Genes for which an ortholog was missing in one or more species, genes with multiple paralogs, or genes previously annotated in *D. melanogaster* or *D. suzukii* but missing in our datasets were investigated and manually corrected if an annotation error was identified. *D. melanogaster* coding sequences were obtained from flybase<sup>168</sup> and *D. suzukii* coding sequences from the literature<sup>143</sup>. Each species' reference genome was uploaded into Geneious (v2022.0.2) and annotated with the GTF files generated above. The coding sequences of the genes selected for manual correction were then combined with these annotated genomes using Minimap2 (v2.17<sup>169</sup>). A new GTF file for the chemosensory genes was generated for each species with annotation errors corrected and previously omitted missing genes added. The GTF files for these manually curated annotations are available in Supplementary Data 17).

For each tissue, the mean TPMs for each gene across replicates were calculated and the number of genes from each chemosensory family that were detected as expressed was evaluated with TPM thresholds of 0.25, 0.5, 1, 2, and 3. For the antenna and proboscis, the number of genes detected only slightly decreased with TPM thresholds between 0.5 and 2 TPM. However, for ovipositor, forelegs, and larva, the number of genes detected dropped significantly with the increase of the TPM threshold. This is likely because some genes are expressed in a few cells, leading to low TPM values. Therefore, to ensure that these genes were not excluded, the threshold for gene detection was settled at 0.5 TPM for all tissues. The TPM file for the chemosensory set of genes is available in Supplementary Data 18.

### Sex-biased gene expression

Genes that have significant differences between sexes were identified using the full set of species' genes and the "full gene" count tables. Read count data for the tissues of each species was read into DESeq2 (v1.34.0<sup>162</sup>) specifying the following design: ~ tissue + sex + tissue:sex. Only genes that had a normalized read count of five in three or more samples were kept for analysis. A Wald test was used to test for sex differences for each gene, requiring a log fold change of 1.5 and an adjusted  $p < 0.01$  for significance.

### Fly Cell atlas data manipulation

**Data importation:** We imported 10x stringent loom and H5DA atlases of legs proboscis and antennae from flycellatlas.org<sup>33</sup>. The H5DA files (that contain the clustering information and feature count matrix for a subset of Highly Variable Genes) were converted to Seurat objects (Seurat v4.3.0, SeuratObject v4.1.3) using the Convert function from the SeuratDisk (v0.0.0.9020; <https://mojaveazure.github.io/seurat-disk/>) and were exported as RDS files using the "saveRDS" function. We used the "Connect" function from SeuratDisk to convert loom files (containing count matrix for all *D. melanogaster* genes but no clustering information) to Seurat objects and exported them as RDS files.

**Mean gene expression per cell cluster:** We split Seurat objects by cluster (`subset(atlas.data, idents = "cluster_ID")`) and extracted their respective feature count matrices (`GetAssayData(object=atlas.data, slot="count")`). We then calculated the mean expression of individual features per cluster (`rowMeans()`) and log-transformed their expression for downstream analysis.

**Visualization of a gene of interest:** The AddModuleScore function from Seurat was used to select gene subsets and visualize their expression using the "FeaturePlot" function. To visualize subsets of cells expressing specific features, we used the "DimPlot" function



specifying cells of interest with the “cells.highlight” option. Gene expression cutoffs were determined after visual examination to highlight highly expressing cells only.

**Cell type homology between tissue:** We used the Seurat “FindAllMarkers” (`atlas.data, only.pos = TRUE, min.pct = 0.25, logfc.threshold = 0.25`) function from Seurat to identify significant markers ( $p < 0.001$ ) among the top 100 list of markers per cell cluster. The list of unique shared markers was retrieved across all tissues, and we generated pairwise correlation matrices based on cluster-mean expression values for each cell cluster across each tissue. In addition, we generated pairwise matrices of the percentage of cell markers shared across tissue cell clusters. The product of these two matrices gives a score between 0 and 1, where 0 corresponds to completely unrelated cell types, and 1 corresponds to identical cell types. This homology score enabled us to cross-validate the Fly-CellAtlas annotation and to identify cell type homology across tissues at a finer scale.

### Measurements of tissue specificity and cell type specificity

We measure gene expression specificity as  $\tau^{170}$  defined as:

$$\sum_{i=1}^n (1 - \hat{x}_i) / (n - 1); \hat{x}_i = x_i / \max(x_i) \quad (2)$$

Where  $x_i$  is the expression of the gene in tissue  $i$ ,  $n$  the number of tissues.

We apply the same formula to define  $\tau$  at the level of cell clusters where  $x_i$  is the mean expression of the gene in cell cluster  $i$  and  $n$  the number of clusters in a given atlas. We also investigated measuring cell specificity  $\tau$  index by considering  $x_i$  as the percentage of cells expressing the gene in cluster  $i$ , which gave very similar distributions. All count values were log-transformed before applying the  $\tau$  formula for stringency purposes.

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

### Data availability

The transcriptomic data generated in this study have been deposited in the ArrayExpress database under accession code [E-MTAB-12656](https://www.ebi.ac.uk/ena/arrayexpress/experiments/E-MTAB-12656) and on our lab’s “sensory RNAseq” GitLab repository [<https://gitlab.com/EvoNeuro/sensory-rnaseq>]. The publicly available single-nucleus datasets used in this study were part of the Fly Cell Atlas and are available at <https://flycellatlas.org>. The location of the source data for Figs. 1–5 and Figures Supplementary 1–3, 7–13 are provided in the table “Source\_data.xlsx”. Source data are provided with this paper.

### Code availability

Code used for this project is available on our lab’s “sensory RNAseq” GitLab repository: <https://gitlab.com/EvoNeuro/sensory-rnaseq>.

### References

- Benton, R., Dessimoz, C. & Moi, D. A putative origin of the insect chemosensory receptor superfamily in the last common eukaryotic ancestor. *eLife* **9**, e62507 (2020).
- 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).
- Brand, P. et al. The origin of the odorant receptor gene family in insects. *eLife* **7**, e38340 (2018).
- Toda, Y. et al. Evolution of the primate glutamate taste sensor from a nucleotide sensor. *Curr. Biol.* **31**, 4641–4649.e5 (2021).
- Poncellet, G. & Shimeld, S. M. The evolutionary origins of the vertebrate olfactory system. *Open Biol.* **10**, 200330 (2020).
- Frank, H. E. R. et al. The evolution of sour taste. *Proc. R. Soc. B Biol. Sci.* **289**, 20211918 (2022).
- Kopp, A. & Barmina, O. Interspecific variation in sex-specific gustatory organs in *Drosophila*. *J. Comp. Neurol.* **530**, 2439–2450 (2022).
- Baeckens, S. et al. Evolutionary morphology of the lizard chemosensory system. *Sci. Rep.* **7**, 10141 (2017).
- Iwasaki, S. Evolution of the structure and function of the vertebrate tongue. *J. Anat.* **201**, 1–13 (2002).
- Elgar, M. A. et al. Insect antennal morphology: the evolution of diverse solutions to odorant perception. *Yale J. Biol. Med.* **91**, 457–469 (2018).
- Leiser-Miller, L. B. & Santana, S. E. Morphological diversity in the sensory system of phyllostomid bats: Implications for acoustic and dietary ecology. *Funct. Ecol.* **34**, 1416–1427 (2020).
- Berta, A., Ekdale, E. G. & Cranford, T. W. Review of the cetacean nose: form, function, and evolution. *Anat. Rec.* **297**, 2205–2215 (2014).
- Kirino, M., Parnes, J., Hansen, A., Kiyohara, S. & Finger, T. E. Evolutionary origins of taste buds: phylogenetic analysis of purinergic neurotransmission in epithelial chemosensors. *Open Biol.* **3**, 130015 (2013).
- Kotrschal, K. Solitary chemosensory cells: why do primary aquatic vertebrates need another taste system? *Trends Ecol. Evol.* **11**, 110–114 (1996).
- Chandrashekar, J., Hoon, M. A., Ryba, N. J. P. & Zuker, C. S. The receptors and cells for mammalian taste. *Nature* **444**, 288–294 (2006).
- Liman, E. R. & Kinnamon, S. C. Sour taste: receptors, cells and circuits. *Curr. Opin. Physiol.* **20**, 8–15 (2021).
- Carleton, A., Accolla, R. & Simon, S. A. Coding in the mammalian gustatory system. *Trends Neurosci.* **33**, 326–334 (2010).
- Scott, K. Gustatory processing in *Drosophila melanogaster*. *Annu. Rev. Entomol.* **63**, 15–30 (2018).
- Tanaka, K. et al. Evolution and development of male-specific leg brushes in *Drosophilidae*. *Dev. Genes Evol.* <https://doi.org/10.1007/s00427-022-00694-3> (2022).
- Crava, C. M., Sassù, F., Tait, G., Becher, P. G. & Anfora, G. Functional transcriptome analyses of *Drosophila suzukii* antennae reveal mating-dependent olfaction plasticity in females. *Insect Biochem Mol. Biol.* **105**, 51–59 (2019).
- Millington, J. W. & Rideout, E. J. Sex differences in *Drosophila* development and physiology. *Curr. Opin. Physiol.* **6**, 46–56 (2018).
- Markow, T. A. & O’Grady, P. M. Evolutionary genetics of reproductive behavior in *Drosophila*: connecting the dots. *Annu. Rev. Genet.* **39**, 263–291 (2005).
- McBride, C. S. et al. Evolution of mosquito preference for humans linked to an odorant receptor. *Nature* **515**, 222–227 (2014).
- DeGennaro, M. et al. orco mutant mosquitoes lose strong preference for humans and are not repelled by volatile DEET. *Nature* **498**, 487–491 (2013).
- De Obaldia, M. E. et al. Differential mosquito attraction to humans is associated with skin-derived carboxylic acid levels. *Cell* **185**, 4099–4116.e13 (2022).
- Pickett, J. A., Woodcock, C. M., Midega, C. A. & Khan, Z. R. Push–pull farming systems. *Curr. Opin. Biotechnol.* **26**, 125–132 (2014).
- Montell, C. *Drosophila* sensory receptors—a set of molecular Swiss Army Knives. *Genetics* **217**, 1–34 (2021).
- Hallam, E. A., Dahanukar, A. & Carlson, J. R. Insect odor and taste receptors. *Annu. Rev. Entomol.* **51**, 113–135 (2006).
- Kwon, J. Y., Dahanukar, A., Weiss, L. A. & Carlson, J. R. A map of taste neuron projections in the *Drosophila* CNS. *J. Biosci.* **39**, 565–574 (2014).

30. Chai, P. C., Cruchet, S., Wigger, L. & Benton, R. Sensory neuron lineage mapping and manipulation in the *Drosophila* olfactory system. *Nat. Commun.* **10**, 643 (2019).
31. Jefferis, G. S. X. E. & Hummel, T. Wiring specificity in the olfactory system. *Semin. Cell Dev. Biol.* **17**, 50–65 (2006).
32. Sánchez-Alcañiz, J. A. et al. An expression atlas of variant ionotropic glutamate receptors identifies a molecular basis of carbonation sensing. *Nat. Commun.* **9**, 4252 (2018).
33. Li, H. et al. Fly Cell Atlas: a single-nucleus transcriptomic atlas of the adult fruit fly. *Science* **375**, eabk2432 (2022).
34. Scheffer, L. K. et al. A connectome and analysis of the adult *Drosophila* central brain. *eLife* **9**, e57443 (2020).
35. Schlegel, P. et al. Information flow, cell types and stereotypy in a full olfactory connectome. *eLife* **10**, e66018 (2021).
36. Corrales, M. et al. A single-cell transcriptomic atlas of complete insect nervous systems across multiple life stages. *Neural Dev.* **17**, 8 (2022).
37. McLaughlin, C. N. et al. Single-cell transcriptomes of developing and adult olfactory receptor neurons in *Drosophila*. *eLife* **10**, e63856 (2021).
38. Arguello, J. R. et al. Targeted molecular profiling of rare olfactory sensory neurons identifies fate, wiring, and functional determinants. *eLife* **10**, e63036 (2021).
39. Hopkins, B. R., Barmina, O. & Kopp, A. Charting the development of *Drosophila* leg sensory organs at single-cell resolution. 2022.10.07.511357 Preprint at <https://doi.org/10.1101/2022.10.07.511357> (2022).
40. Li, H. et al. Classifying *Drosophila* olfactory projection neuron subtypes by single-cell RNA sequencing. *Cell* **171**, 1206–1220.e22 (2017).
41. Winding, M. et al. The connectome of an insect brain. *Science* **379**, eadd9330 (2023).
42. Dobzhansky, T. Genetics of natural populations. Xix. Origin of heterosis through natural selection in populations of *Drosophila pseudoobscura*. *Genetics* **35**, 288–302 (1950).
43. McDonald, J. H. & Kreitman, M. Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature* **351**, 652–654 (1991).
44. Schrider, D. R., Ayroles, J., Matute, D. R. & Kern, A. D. Supervised machine learning reveals introgressed loci in the genomes of *Drosophila simulans* and *D. sechellia*. *PLoS Genet.* **14**, e1007341 (2018).
45. Clark, A. G. et al. Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* **450**, 203–218 (2007).
46. Coyne, J. A. & Orr, H. A. Patterns of speciation in *Drosophila*. *Evolution* **43**, 362–381 (1989).
47. Kliman, R. M. et al. The population genetics of the origin and divergence of the *Drosophila simulans* complex species. *Genetics* **156**, 1913–1931 (2000).
48. Markow, T. A. & O'Grady, P. M. *Drosophila* biology in the genomic age. *Genetics* **177**, 1269–1276 (2007).
49. Khallaf, M. A. & Knaden, M. Evolutionary neuroecology of olfactory-mediated sexual communication and host specialization in *Drosophila* – a review. *Entomol. Exp. Appl.* **170**, 289–302 (2022).
50. Obbard, D. J. et al. Estimating divergence dates and substitution rates in the *Drosophila* phylogeny. *Mol. Biol. Evol.* **29**, 3459–3473 (2012).
51. Ometto, L. et al. Linking genomics and ecology to investigate the complex evolution of an invasive *Drosophila* pest. *Genome Biol. Evol.* **5**, 745–757 (2013).
52. Garrigan, D. et al. Genome sequencing reveals complex speciation in the *Drosophila simulans* clade. *Genome Res.* **22**, 1499–1511 (2012).
53. Suvorov, A. et al. Widespread introgression across a phylogeny of 155 *Drosophila* genomes. *Curr. Biol.* **32**, 111–123.e5 (2022).
54. Auer, T. O., Shahandeh, M. P. & Benton, R. *Drosophila sechellia*: a genetic model for behavioral evolution and neuroecology. *Annu. Rev. Genet.* **55**, 527–554 (2021).
55. Lachaise, D. et al. Evolutionary novelties in islands: *Drosophila santomea*, a new melanogaster sister species from São Tomé. *Proc. Biol. Sci.* **267**, 1487–1495 (2000).
56. Matute, D. R., Novak, C. J. & Coyne, J. A. Temperature-based extrinsic reproductive isolation in two species of *Drosophila*. *Evolution* **63**, 595–612 (2009).
57. Linz, J. et al. Host plant-driven sensory specialization in *Drosophila erecta*. *Proc. R. Soc. B Biol. Sci.* **280**, 20130626 (2013).
58. Fraimout, A. et al. Deciphering the Routes of invasion of *Drosophila suzukii* by Means of ABC Random Forest. *Mol. Biol. Evol.* **34**, 980–996 (2017).
59. Santos, L., Ados et al. Global potential distribution of *Drosophila suzukii* (Diptera, Drosophilidae). *PLoS One* **12**, e0174318 (2017).
60. Keesey, I. W., Knaden, M. & Hansson, B. S. Olfactory specialization in *Drosophila suzukii* supports an ecological shift in host preference from rotten to fresh fruit. *J. Chem. Ecol.* **41**, 121–128 (2015).
61. Cavey, M. et al. Evolutionary divergence in sugar valuation shifts *Drosophila suzukii* oviposition choice towards ripe fruit. 2022.06.22.497171 Preprint at <https://doi.org/10.1101/2022.06.22.497171> (2022).
62. Green, J. E. et al. Evolution of ovipositor length in *Drosophila suzukii* is driven by enhanced cell size expansion and anisotropic tissue reorganization. *Curr. Biol.* **29**, 2075–2082.e6 (2019).
63. Knapp, L., Mazzi, D. & Finger, R. The economic impact of *Drosophila suzukii*: perceived costs and revenue losses of Swiss cherry, plum and grape growers. *Pest Manag. Sci.* **77**, 978–1000 (2021).
64. Goodhue, R. E., Bolda, M., Farnsworth, D., Williams, J. C. & Zalom, F. G. Spotted wing *Drosophila* infestation of California strawberries and raspberries: economic analysis of potential revenue losses and control costs. *Pest Manag. Sci.* **67**, 1396–1402 (2011).
65. Bolda, M. P., Goodhue, R. E. & Zalom, F. G. Spotted wing *Drosophila*: potential economic impact of a newly established pest. *ARE Update* **13**, 5–8 (2010).
66. Capy, P. & Gibert, P. *Drosophila melanogaster*, *Drosophila simulans*: so similar yet so different. *Genetica* **120**, 5–15 (2004).
67. Karageorgi, M. et al. Evolution of multiple sensory systems drives novel egg-laying behavior in the fruit pest *Drosophila suzukii*. *Curr. Biol.* **27**, 847–853 (2017).
68. Yang, J., Ruan, H., Xu, W. & Gu, X. TreeExp2: an integrated framework for phylogenetic transcriptome analysis. *Genome Biol. Evol.* **11**, 3276–3282 (2019).
69. Stensmyr, M. C. et al. A conserved dedicated olfactory circuit for detecting harmful microbes in *Drosophila*. *Cell* **151**, 1345–1357 (2012).
70. Price, P. D. et al. Detecting signatures of selection on gene expression. *Nat. Ecol. Evol.* **6**, 1035–1045 (2022).
71. Long, M., Betran, E., Thornton, K. & Wang, W., & others. The origin of new genes: glimpses from the young and old. *Nat. Rev. Genet.* **4**, 865–875 (2003).
72. Cardoso-Moreira, M. et al. Gene expression across mammalian organ development. *Nature* **571**, 505–509 (2019).
73. Assis, R. & Bachtrog, D. Neofunctionalization of young duplicate genes in *Drosophila*. *Proc. Natl Acad. Sci.* **110**, 17409–17414 (2013).
74. Makova, K. D. & Li, W.-H. Divergence in the spatial pattern of gene expression between human duplicate genes. *Genome Res.* **13**, 1638–1645 (2003).
75. Kaessmann, H. Origins, evolution, and phenotypic impact of new genes. *Genome Res.* **20**, 1313–1326 (2010).
76. Guschanski, K., Warnefors, M. & Kaessmann, H. The evolution of duplicate gene expression in mammalian organs. *Genome Res.* **27**, 1461–1474 (2017).

77. Xu, A. et al. Novel genes expressed in subsets of chemosensory sensilla on the front legs of male *Drosophila melanogaster*. *Cell Tissue Res.* **307**, 381–392 (2002).
78. Ben-Shahar, Y. et al. The *Drosophila* gene *CheB42a* is a novel modifier of *Deg/ENaC* channel function. *PLOS ONE* **5**, e9395 (2010).
79. Rihani, K., Ferveur, J.-F. & Briand, L. The 40-year mystery of insect odorant-binding proteins. *Biomolecules* **11**, 509 (2021).
80. Pan, J. W. et al. Patterns of transcriptional parallelism and variation in the developing olfactory system of *Drosophila* species. *Sci. Rep.* **7**, 8804 (2017).
81. 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).
82. Shiao, M.-S. et al. Transcriptional profiling of adult *Drosophila* antennae by high-throughput sequencing. *Zool. Stud.* **52**, 42 (2013).
83. Walker, W. B. et al. Comparative transcriptomic assessment of the chemosensory receptor repertoire of *Drosophila suzukii* adult and larval olfactory organs. *Comp. Biochem. Physiol. Part D. Genom. Proteom.* **45**, 101049 (2023).
84. Crava, C. M. et al. Structural and transcriptional evidence of mechanotransduction in the *Drosophila suzukii* ovipositor. *J. Insect Physiol.* **125**, 104088 (2020).
85. Menuz, K., Larter, N. K., Park, J. & Carlson, J. R. An RNA-seq screen of the *Drosophila* antenna identifies a transporter necessary for ammonia detection. *PLoS Genet.* **10**, e1004810 (2014).
86. Larter, N. K., Sun, J. S. & Carlson, J. R. Organization and function of *Drosophila* odorant binding proteins. *eLife* **5**, e20242 (2016).
87. Jaeger, A. H. et al. A complex peripheral code for salt taste in *Drosophila*. *eLife* **7**, e37167 (2018).
88. Galindo, K. & Smith, D. P. A large family of divergent *Drosophila* odorant-binding proteins expressed in gustatory and olfactory sensilla. *Genetics* **159**, 1059–1072 (2001).
89. Sánchez-Alcañiz, J. A., Zappia, G., Marion-Poll, F. & Benton, R. A mechanosensory receptor required for food texture detection in *Drosophila*. *Nat. Commun.* **8**, 14192 (2017).
90. Jeong, Y. T. et al. An odorant-binding protein required for suppression of sweet taste by bitter chemicals. *Neuron* **79**, 725–737 (2013).
91. Depetris-Chauvin, A., Galagovsky, D. & Grosjean, Y. Chemicals and chemoreceptors: ecologically relevant signals driving behavior in *Drosophila*. *Front. Ecol. Evol.* **3**, 41 (2015).
92. Fujii, S. et al. *Drosophila* sugar receptors in sweet taste perception, olfaction, and internal nutrient sensing. *Curr. Biol.* **25**, 621–627 (2015).
93. Hekmat-Scafe, D. S., Scafe, C. R., McKinney, A. J. & Tanouye, M. A. Genome-wide analysis of the odorant-binding protein gene family in *Drosophila melanogaster*. *Genome Res.* **12**, 1357–1369 (2002).
94. Jeong, Y. T. et al. Mechanosensory neurons control sweet sensing in *Drosophila*. *Nat. Commun.* **7**, 12872 (2016).
95. Zhang, Y. V., Aikin, T. J., Li, Z. & Montell, C. The basis of food texture sensation in *Drosophila*. *Neuron* **91**, 863–877 (2016).
96. Koh, T.-W. et al. The *Drosophila* *IR20a* clade of ionotropic receptors are candidate taste and pheromone receptors. *Neuron* **83**, 850–865 (2014).
97. Weiss, L. A., Dahanukar, A., Kwon, J. Y., Banerjee, D. & Carlson, J. R. The molecular and cellular basis of bitter taste in *Drosophila*. *Neuron* **69**, 258–272 (2011).
98. Starostina, E., Xu, A., Lin, H. & Pikielny, C. W. A *Drosophila* protein family implicated in pheromone perception is related to *Tay-Sachs* GM2-activator protein. *J. Biol. Chem.* **284**, 585–594 (2009).
99. Montell, C. A taste of the *Drosophila* gustatory receptors. *Curr. Opin. Neurobiol.* **19**, 345–353 (2009).
100. Akitake, B. et al. Coordination and fine motor control depend on *Drosophila* TRP. *Nat. Commun.* **6**, 7288 (2015).
101. Watanabe, K., Toba, G., Koganezawa, M. & Yamamoto, D. Gr39a, a highly diversified gustatory receptor in *Drosophila*, has a role in sexual behavior. *Behav. Genet.* **41**, 746–753 (2011).
102. Ling, F., Dahanukar, A., Weiss, L. A., Kwon, J. Y. & Carlson, J. R. The molecular and cellular basis of taste coding in the legs of *Drosophila*. *J. Neurosci.* **34**, 7148–7164 (2014).
103. Liu, T. et al. The receptor channel formed by *ppk25*, *ppk29* and *ppk23* can sense the *Drosophila* female pheromone 7,11-heptacosadiene. *Genes Brain Behav.* **19**, e12529 (2020).
104. Fishilevich, E. et al. Chemotaxis behavior mediated by single larval olfactory neurons in *Drosophila*. *Curr. Biol.* **15**, 2086–2096 (2005).
105. Liu, L. et al. Contribution of *Drosophila* *DEG/ENaC* genes to salt taste. *Neuron* **39**, 133–146 (2003).
106. Mast, J. D., De Moraes, C. M., Alborn, H. T., Lavis, L. D. & Stern, D. L. Evolved differences in larval social behavior mediated by novel pheromones. *eLife* **3**, e04205 (2014).
107. Kwon, J. Y., Dahanukar, A., Weiss, L. A. & Carlson, J. R. Molecular and cellular organization of the taste system in the *Drosophila* larva. *J. Neurosci.* **31**, 15300–15309 (2011).
108. Zhou, S., Stone, E. A., Mackay, T. F. C. & Anholt, R. R. H. Plasticity of the chemoreceptor repertoire in *Drosophila melanogaster*. *PLOS Genet.* **5**, e1000681 (2009).
109. Kreher, S. A., Kwon, J. Y. & Carlson, J. R. The molecular basis of odor coding in the *Drosophila* larva. *Neuron* **46**, 445–456 (2005).
110. Turner, H. N. et al. The TRP channels *Pkd2*, *NompC*, and *Trpm* act in cold-sensing neurons to mediate unique aversive behaviors to noxious cold in *Drosophila*. *Curr. Biol. CB* **26**, 3116–3128 (2016).
111. Dweck, H. K. et al. Olfactory channels associated with the *Drosophila* maxillary palp mediate short- and long-range attraction. *eLife* **5**, e14925 (2016).
112. Park, S. K. et al. A *Drosophila* protein specific to pheromone-sensing Gustatory hairs delays males' copulation attempts. *Curr. Biol.* **16**, 1154–1159 (2006).
113. Shang, J. et al. Sensing of a spore surface protein by a *Drosophila* chemosensory protein induces behavioral defense against fungal parasitic infections. *Curr. Biol.* **33**, 276–286.e5 (2023).
114. Hu, Y. et al. Gr33a modulates *Drosophila* male courtship preference. *Sci. Rep.* **5**, 7777 (2015).
115. Shim, J. et al. The full repertoire of *Drosophila* gustatory receptors for detecting an aversive compound. *Nat. Commun.* **6**, 8867 (2015).
116. Prieto-Godino, L. L. et al. Functional integration of “undead” neurons in the olfactory system. *Sci. Adv.* **6**, eaaz7238 (2020).
117. Kohatsu, S., Tanabe, N., Yamamoto, D. & Isono, K. Which sugar to take and how much to take? Two distinct decisions mediated by separate sensory channels. *Front. Mol. Neurosci.* **15**, 895395 (2022).
118. Miyamoto, T., Chen, Y., Slone, J. & Amrein, H. Identification of a *Drosophila* glucose receptor using Ca<sup>2+</sup> imaging of single chemosensory neurons. *PLoS One* **8**, e56304 (2013).
119. Datta, S. R. et al. The *Drosophila* pheromone *cVA* activates a sexually dimorphic neural circuit. *Nature* **452**, 473–477 (2008).
120. Luecke, D. M. & Kopp, A. Sex-specific evolution of relative leg size in *Drosophila* *prolongata* results from changes in the intersegmental coordination of tissue growth. *Evolution* **73**, 2281–2294 (2019).
121. Bray, S. & Amrein, H. A putative *Drosophila* pheromone receptor expressed in male-specific taste neurons is required for efficient courtship. *Neuron* **39**, 1019–1029 (2003).
122. Ahn, S.-J. et al. Sex-biased gene expression in antennae of *Drosophila suzukii*. *Arch. Insect Biochem Physiol.* **104**, e21660 (2020).
123. Ellegren, H. & Parsch, J. The evolution of sex-biased genes and sex-biased gene expression. *Nat. Rev. Genet.* **8**, 689–698 (2007).



124. Catalan, A., Hutter, S. & Parsch, J. Population and sex differences in *Drosophila melanogaster* brain gene expression. *BMC Genom.* **13**, 654 (2012).
125. Parker, D. J. et al. Sex-biased gene expression is repeatedly masculinized in asexual females. *Nat. Commun.* **10**, 4638 (2019).
126. Khodursky, S., Svetec, N., Durkin, S. M. & Zhao, L. The evolution of sex-biased gene expression in the *Drosophila* brain. *Genome Res.* **30**, 874–884 (2020).
127. Rodríguez-Montes, L. et al. Sex-biased gene expression across mammalian organ development and evolution. *Science* **382**, eadf1046 (2023).
128. Yu, J. Y., Kanai, M. I., Demir, E., Jefferis, G. S. X. E. & Dickson, B. J. Cellular organization of the neural circuit that drives *Drosophila* courtship behavior. *Curr. Biol.* **20**, 1602–1614 (2010).
129. Yamamoto, D. & Koganezawa, M. Genes and circuits of courtship behaviour in *Drosophila* males. *Nat. Rev. Neurosci.* **14**, 681–692 (2013).
130. McKelvey, E. G. Z. et al. *Drosophila* females receive male substrate-borne signals through specific leg neurons during courtship. *Curr. Biol.* **31**, 3894–3904.e5 (2021).
131. Tracy, C. B., Nguyen, J., Abraham, R. & Shirangi, T. R. Evolution of sexual size dimorphism in the wing musculature of *Drosophila*. *PeerJ* **8**, e8360 (2020).
132. Wang, Z.-Y. et al. Transcriptome and translome co-evolution in mammals. *Nature* **588**, 642–647 (2020).
133. Nourmohammad, A. et al. Adaptive evolution of gene expression in *Drosophila*. *Cell Rep.* **20**, 1385–1395 (2017).
134. Cridland, J. M., Majane, A. C., Sheehy, H. K. & Begun, D. J. Polymorphism and divergence of novel gene expression patterns in *Drosophila melanogaster*. *Genetics* **216**, 79–93 (2020).
135. Carroll, S. B. Evolution at two levels: on genes and form. *PLOS Biol.* **3**, e245 (2005).
136. Stern, D. L. Perspective: evolutionary developmental biology and the problem of variation. *Evolution* **54**, 1079–1091 (2000).
137. Wittkopp, P. J. & Kalay, G. Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. *Nat. Rev. Genet.* **13**, 59–69 (2012).
138. Coolon, J. D., McManus, C. J., Stevenson, K. R., Graveley, B. R. & Wittkopp, P. J. Tempo and mode of regulatory evolution in *Drosophila*. *Genome Res.* **24**, 797–808 (2014).
139. Metzger, B. P. H., Wittkopp, P. J. & Coolon, J. Evolutionary dynamics of regulatory changes underlying gene expression divergence among *Saccharomyces* species. *Genome Biol. Evol.* **9**, 843–854 (2017).
140. Hill, M. S., Vande Zande, P. & Wittkopp, P. J. Molecular and evolutionary processes generating variation in gene expression. *Nat. Rev. Genet.* **22**, 203–215 (2021).
141. Arguello, J. R. et al. Extensive local adaptation within the chemosensory system following *Drosophila melanogaster*'s global expansion. *Nat. Commun.* **7**, ncomms11855 (2016).
142. McBride, C. S. & Arguello, J. R. Five *Drosophila* genomes reveal nonneutral evolution and the signature of host specialization in the chemoreceptor superfamily. *Genetics* **177**, 1395–1416 (2007).
143. Ramasamy, S. et al. The evolution of olfactory gene families in *Drosophila* and the genomic basis of chemical-ecological adaptation in *Drosophila suzukii*. *Genome Biol. Evol.* **8**, 2297–2311 (2016).
144. Robertson, H. M. Molecular evolution of the major arthropod chemoreceptor gene families. *Annu Rev. Entomol.* **64**, 227–242 (2019).
145. Prieto-Godino, L. L. et al. Evolution of acid-sensing olfactory circuits in *Drosophilids*. *Neuron* **93**, 661–676.e6 (2017).
146. Takagi, S. et al. *Sensory neuron population expansion enhances odour tracking through relaxed projection neuron adaptation.* <https://doi.org/10.1101/2023.09.15.556782> (2023) 10.1101/2023.09.15.556782.
147. Grenier, J. K. et al. Global diversity lines—a five-continent reference panel of sequenced *Drosophila melanogaster* strains. *G3 Genes Genomes Genet.* **5**, 593–603 (2015).
148. Gomez-Diaz, C., Reina, J. H., Cambillau, C. & Benton, R. Ligands for pheromone-sensing neurons are not conformationally activated odorant binding proteins. *PLOS Biol.* **11**, e1001546 (2013).
149. Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).
150. Hoff, K. J., Lomsadze, A., Borodovsky, M. & Stanke, M. Whole-Genome Annotation with BRAKER. in *Gene Prediction: Methods and Protocols* (ed Kollmar, M.) 65–95 (Springer, 2019). [https://doi.org/10.1007/978-1-4939-9173-0\\_5](https://doi.org/10.1007/978-1-4939-9173-0_5).
151. Brůna, T., Hoff, K. J., Lomsadze, A., Stanke, M. & Borodovsky, M. BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR Genom. Bioinform.* **3**, lqaa108 (2021).
152. Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015).
153. Trapnell, C. et al. Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat. Biotechnol.* **31**, 46–53 (2013).
154. Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**, 238 (2019).
155. Rice, P., Longden, I. & Bleasby, A. EMBOSS: the European molecular biology open software suite. *Trends Genet. TIG* **16**, 276–277 (2000).
156. Grau-Bové, X. & Sebé-Pedrós, A. Orthology clusters from gene trees with possvm. *Mol. Biol. Evol.* **38**, 5204–5208 (2021).
157. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
158. Nguyen, L.-T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).
159. Camacho, C. et al. BLAST+: architecture and applications. *BMC Bioinforma.* **10**, 421 (2009).
160. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2012).
161. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
162. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
163. R Core Team. *R: A Language and Environment for Statistical Computing.* (R Foundation for Statistical Computing, 2012).
164. Khabbazian, M., Kriebel, R., Rohe, K. & Ané, C. Fast and accurate detection of evolutionary shifts in Ornstein–Uhlenbeck models. *Methods Ecol. Evol.* **7**, 811–824 (2016).
165. Pal, S., Oliver, B. & Przytycka, T. M. Stochastic modeling of gene expression evolution uncovers tissue- and sex-specific properties of expression evolution in the *Drosophila* genus. *J. Comput. Biol.* **30**, 21–40 (2023).
166. Russo, P. S. T. et al. CEMiTool: a Bioconductor package for performing comprehensive modular co-expression analyses. *BMC Bioinforma.* **19**, 56 (2018).
167. Shao, Y. et al. GenTree, an integrated resource for analyzing the evolution and function of primate-specific coding genes. *Genome Res.* **29**, 682–696 (2019).
168. Gramates, L. S. et al. FlyBase: a guided tour of highlighted features. *Genetics* **220**, iyac035 (2022).



169. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).
170. Kryuchkova-Mostacci, N. & Robinson-Rechavi, M. A benchmark of gene expression tissue-specificity metrics. *Brief. Bioinform.* **18**, 205–214 (2017).

## Acknowledgements

We thank Margarida Cardoso-Moreira for discussions and advice throughout the project. Margarida Cardoso-Moreira, John Pannell, Thomas Auer, Giulia Zancolli, and Lucia L. Prieto Godino provided important comments on an earlier version of the paper. T. M. Balls provided support and inspiration. Wolf Huetteroth provided anatomical assistance with *Or1a* expression in the proboscis. Sequencing was performed at the Lausanne Genomic Technologies Facility, and the University of Lausanne's HPC services provided resources and support. Microscopy was performed at the University of Lausanne's Cellular Imaging Facility. Research in JRA's laboratory was supported by the University of Lausanne and the Swiss National Science Foundation (grants PPOOP3\_176956 and 310030\_201188).

## Author contributions

Conceptualization: G.B., B.S.L., and J.R.A.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-023-44558-4>.

**Correspondence** and requests for materials should be addressed to Gwénaëlle Bontonou, Bastien Saint-Leandre or J. Roman Arguello.

**Peer review information** *Nature Communications* thanks Philipp Brand, Trevor Sorrelis and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024

# QUERY FORM

NATURECOMMUNICATIONS	
Manuscript ID	[Art. Id: 44558]
Author	
Editor	
Publisher	

## Journal: NATURECOMMUNICATIONS

**Author** :- The following queries have arisen during the editing of your manuscript. Please answer by making the requisite corrections directly in the e-proofing tool rather than marking them up on the PDF. This will ensure that your corrections are incorporated accurately and that your paper is published as quickly as possible.

Query No.	Description	Author's Response
AQ1	Please confirm or correct the city/country name inserted in affiliation 4.	
AQ2	Please check your article carefully, coordinate with any co-authors and enter all final edits clearly in the eproof, remembering to save frequently. If edits are needed to figures, please attach a corrected figure file containing the relevant changes. Once corrections are submitted, we cannot routinely make further changes to the article.	
AQ3	Note that the eproof should be amended in only one browser window at any one time; otherwise changes will be overwritten.	
AQ4	Author surnames have been highlighted. Please check these carefully and adjust if the first name or surname is marked up incorrectly, as this will affect indexing of your article in public repositories such as PubMed. Also, carefully check the spelling and numbering of all author names and affiliations, and the corresponding author(s) email address(es). Please note that email addresses should only be included for designated corresponding authors, and you cannot change corresponding authors at this stage except to correct errors made during typesetting.	
AQ5	You cannot alter accepted Supplementary Information files except for critical changes to scientific content. If you do resupply any files, please also provide a brief (but complete) list of changes. If these are not considered scientific changes, any altered Supplementary files will not be used, only the originally accepted version will be published.	
AQ6	If applicable, please ensure that any accession codes and datasets whose DOIs or other identifiers are mentioned in the paper are scheduled for public release as soon as possible, we recommend within a few days of submitting your proof, and update the database record with publication details from this article once available.	

# QUERY FORM

NATURECOMMUNICATIONS	
<b>Manuscript ID</b>	[Art. Id: 44558]
<b>Author</b>	
<b>Editor</b>	
<b>Publisher</b>	

## Journal: NATURECOMMUNICATIONS

**Author** :- The following queries have arisen during the editing of your manuscript. Please answer by making the requisite corrections directly in the e.proofing tool rather than marking them up on the PDF. This will ensure that your corrections are incorporated accurately and that your paper is published as quickly as possible.

Query No.	Description	Author's Response
AQ7	If ref. 39 (preprint) has now been published in final peer-reviewed form, please update the reference details if appropriate.	
AQ8	If ref. 61 (preprint) has now been published in final peer-reviewed form, please update the reference details if appropriate.	