RESEARCH ARTICLE

Class-wide sequence, structural, and docking-profile diversity of insect odorant receptors and structural specialization of Orco

Tianmin Zhang[1], Xuanxiao Yang[1], Yufang Fu[2], Wei Xue[1], Yifeng Zhang[1], Suyang Duan[1], Yangming Yin[1], Yi Guo[1], Chenxi Gao[1], Yang Liu[2], Gang Li[2], Chang Xu[2]*, Huimeng Lu[1]*

[1]School of Life Science and Technology, Northwestern Polytechnical University, Xi'an, China; [2]College of Life Sciences, Shaanxi Normal University, Xi'an, China

eLife Assessment

This **important** large-scale comparative study of odorant receptor (OR) genes across more than 100 insect species, combining sequence- and structure-based approaches, represents a significant contribution in the exploration of the evolution of this large gene family involved in the detection of odorant signals by olfactory neurons. Most notably, the authors uncover a structural feature unique to the odorant receptor co-receptor Orco that reduces ligand binding affinity. From this **solid** data emerges a general integrative framework to apprehend the diversity and evolution of odorant receptor genes in insects.

Abstract Insect olfaction is facilitated by a heterotetrameric odorant receptor–odorant receptor co-receptor (OR-Orco) complex. However, extensive sequence divergence and lineage-specific expansion have complicated large-scale comparisons of OR repertoires across insects. Here, we integrated sequence similarity, structural similarity, and molecular docking to analyze ORs from 115 insect species. OR repertoire size was associated with adult diet, larval diet, and habitat, but not with circadian rhythm. We developed a 'trunk–branch' network framework to classify ORs according to sequence and structural similarity and used large-scale molecular docking to derive comparative docking profiles. Despite marked differences in sequence- and structural-community composition, almost all sampled orders contained ORs assigned to all six docking-profile communities. Holometabolous insects showed higher OR sequence-community diversity than non-holometabolous insects, and docking-profile composition differed between extant lineages whose order-level origins predate or postdate the end-Permian mass extinction. Most notably, we identified a highly conserved extracellular β-sheet in Orco that restricts ligand access to the binding pocket and may contribute to its reduced capacity to bind natural odorants. The emergence of Orco represents a key evolutionary transition from homomeric receptors to heteromeric OR–Orco complexes, and our results suggest that this transition may have been accompanied by pronounced specialization of the extracellular domain and binding pocket. Together, these analyses provide a class-wide framework for comparing insect OR diversity and generate testable hypotheses for future experimental and evolutionary studies.

***For correspondence:** xuchang@snnu.edu.cn (CX); luhuimeng@nwpu.edu.cn (HL)

Competing interest: The authors declare that no competing interests exist.

Funding: See page 21

Preprint posted 13 February 2026
Sent for Review 12 February 2026
Reviewed preprint v1 12 May 2026
Reviewed preprint v2 25 September 2026
Reviewed preprint v3 14 October 2026
Version of Record 03 November 2026

Reviewing Editor: Virginie Courtier-Orgogozo, CNRS - Universite Paris Cite, France

© Copyright Zhang *et al*. This article is distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use and redistribution provided that the original author and source are credited.

Introduction

Olfactory systems play a critical role in animal survival strategies, governing critical behaviors from nutrient acquisition to mate selection (**Bargmann, 2006**; **Nei et al., 2008**; **Kaupp, 2010**). Within the kingdom Animalia, insects—the most speciose metazoan group—have evolved exceptionally sophisticated chemosensory systems to interpret complex volatile landscapes (**Nei et al., 2008**; **Kaupp, 2010**; **Mora et al., 2011**; **Robertson, 2019**). The extraordinary sensitivity of insect olfaction to hydrophobic volatile organic compounds (VOCs) is facilitated by a heterotetrameric channel complex consisting of odorant receptors (ORs) and odorant receptor co-receptors (Orco; **Clyne et al., 1999**; **Vosshall et al., 1999**; **Sato et al., 2008**). Insect ORs represent a distinct protein superfamily that originated independently of vertebrate ORs, with no shared homology between the two (**Sato et al., 2008**; **Hansson and Stensmyr, 2011**). Studies have shown that this ancient structural fold has a deeper phylogenetic origin (**Benton and Himmel, 2023**; **Himmel et al., 2023**). While vertebrate ORs have been extensively studied (**Niimura and Nei, 2005**; **Hughes et al., 2018**; **Pacalon et al., 2023**; **Policarpo et al., 2024**), the extremely low sequence similarity among insect ORs has made resolving their internal sequence, structural, and functional differences challenging (**Robertson and Wanner, 2006**). Extensive studies have reported advances in the olfactory recognition mechanism and OR function (**Guo et al., 2021**; **Chang et al., 2023**; **Depetris-Chauvin et al., 2023**; **Jiang et al., 2024**), which have greatly facilitated a deeper understanding of insect olfactory perception. However, a class-wide integrative framework linking OR sequence divergence, structural variation, and docking-derived binding-profile diversity is still lacking. Such a framework is needed to distinguish broadly conserved patterns from lineage-specific features and to place species-level OR diversification in a broader macroevolutionary context.

Within the class Insecta, Orcos exhibit exceptional conservation (>400 Mya evolutionary stability), whereas lineage-specific ORs display unprecedented diversity, far exceeding that of vertebrate ORs (**Larsson et al., 2004**; **Nei et al., 2008**; **Brand et al., 2018**; **Robertson, 2019**; **Wicher and Miazzi, 2021**). Although several studies have linked OR repertoire evolution to ecological adaptation in specific insect lineages, a unified class-wide comparison remains lacking (**Jongepier et al., 2022**; **Yin et al., 2022**; **Singh et al., 2025**). Traditional phylogeny-driven subfamily categorization, constrained by extreme cross-order sequence divergence, fails to capture the continuum of OR evolutionary innovation (**Zhou et al., 2012**; **Zhou et al., 2015**; **Mitchell et al., 2020**; **Legan et al., 2021**; **Tian et al., 2022**; **Yin et al., 2022**). Unlike vertebrate G-protein-coupled ORs, insect ORs function as ligand-gated ion channels by assembling with the conserved co-receptor Orco, with recent cryo-EM structures supporting a 1:3 OR–Orco heterotetrameric model (**Jones et al., 2005**; **Nei et al., 2008**; **Kaupp, 2010**; **Jones et al., 2011**; **Wang et al., 2024**; **Zhao et al., 2024**). However, the exact in vivo stoichiometry remains to be fully resolved, and alternative arrangements such as 2:2 may also be possible (**Wang et al., 2024**). The mechanistic drivers of this functional divergence raise key evolutionary questions: This functional specialization raises two key questions: which structural features distinguish Orco from ligand-selective ORs, and how might these features contribute to its reduced capacity to bind natural odorants?

Identifying the binding profiles of ORs for VOCs is fundamental to understanding how insect olfactory systems perceive and adapt to their environments (**McBride et al., 2014**; **Auer et al., 2020**). To date, numerous studies have characterized the ligand-binding profiles of various insect ORs using heterologous expression systems (**Hallem and Carlson, 2006**; **Carey et al., 2010**; **de Fouchier et al., 2017**; **Guo et al., 2021**; **Chang et al., 2023**; **Jiang et al., 2024**). However, when attempting to investigate the functional relevance of tens of thousands of ORs across the class Insecta at a macroevolutionary scale, experimental methods are no longer feasible due to their long timeframes and low throughput. For comparative analyses at this scale, computational approaches provide a practical means of generating hypotheses that can subsequently be prioritized for experimental testing. Although machine-learning methods remain constrained by the available training data (**Boyle et al., 2013**; **Chepurwar et al., 2019**; **Caballero-Vidal et al., 2020**), molecular docking offers a scalable complementary approach for comparing potential receptor–ligand interactions across large OR repertoires (**Lyu et al., 2019**; **Comte et al., 2025**). This strategy has previously been applied to mammalian OR repertoires: by employing large-scale molecular docking and OR-binding profile prediction, researchers have revealed differences in olfactory perception among bats with different diets, as well as dynamic trade-offs between vision and olfaction in primates (**Zhang et al., 2024**; **Chi et al., 2025**).

These results suggest that molecular docking can facilitate the exploration of macroevolutionary patterns.

In this study, we characterized large-scale patterns of OR diversification across 115 species spanning 16 insect orders. Using a framework that integrates network analysis, protein structure prediction, and molecular docking, we constructed sequence-similarity (SSN), structural-similarity (StSN), and docking-profile similarity networks (dPSN) to classify ORs from three distinct perspectives. We identified differences in OR sequence-community diversity between holometabolous and non-holometabolous insects, as well as differences in docking-profile community composition among extant lineages whose order-level origins fall on opposite sides of the end-Permian mass extinction (EPME). Across this class-wide sampling, OR repertoire size was associated with adult diet, larval diet, and habitat, but not with circadian rhythm. Moreover, the conserved β-sheet in the extracellular loop 2 (EL2) region of Orco, together with its specialized binding-pocket architecture, is associated with its reduced ligand-binding capacity and represents a prominent structural specialization of the co-receptor lineage.

Results

Characterization of the insect odorant receptor gene repertoire

We compiled all published insect genomes (as of December 2022), selecting the highest-quality genome assembly (contig N50 >1 Mb) from each family as the representative species. Our study includes 115 species, featuring 8 representative species from different insect orders with previously annotated ORs (***Engsontia et al., 2008***; ***Wang et al., 2015***; ***Zhou et al., 2015***; ***Brand et al., 2018***; ***Robertson et al., 2019***), spanning 16 orders and 111 families. This approach ensured a comprehensive analysis across the class Insecta (***Figure 1A***). To increase annotation reliability, we implemented the most stringent protein superfamily member screening strategy to date (***Figure 1—figure supplement 1A***). A total of 8905 structurally intact ORs were retained for downstream analyses . Benchmarking against curated OR repertoires showed that the automated pipeline recovered most previously reported ORs and that stringent structural filtering produced a high-confidence set of structurally intact receptors (Appendix 1, ***Supplementary file 1***, and ***Figure 1—figure supplement 1B–C***). For most species with comparable genome-based references, retained OR counts exceeded 70% of published values. Our analysis revealed substantial variation in OR counts across insect species. Hymenoptera exhibited the highest number of ORs and accounted for eight of the top 10 species with the largest OR repertoires (***Bonasio et al., 2010***). In our OR dataset, Odonata had the fewest intact ORs, with a mean of 4 (N=3). Hymenoptera also showed the widest range in OR counts (min: 38, max: 473, N=23), followed by Coleoptera (min: 10, max: 271, N=15, ***Figure 1B***).

Previous studies have shown that in certain insect orders and specific lineages, OR repertoire characteristics are associated with various ecological traits (***McBride, 2007***; ***Liu et al., 2021***; ***Schrader et al., 2021***; ***Jongepier et al., 2022***; ***Singh et al., 2025***). To determine whether these relationships extend to the class Insecta, we analyzed four ecological factors: adult diet, larval diet, habitat, and circadian rhythm. Adult and larval diets represent stage-specific ecological traits, and our analysis of OR gene numbers at the genomic level aimed to reveal their relationship with long-term evolutionary pressures shaping OR gene family size. Our results revealed a significant correlation between insect OR counts and dietary preferences, both in adults (***Figure 1C***, $p=1.61 \times 10^{-6}$) and larvae (***Figure 1D***, p=0.03143). Habitat was also significantly associated with OR counts (p=0.0354), with terrestrial insects exhibiting more ORs than aquatic insects do (***Figure 1F***). However, there was no significant correlation between circadian rhythm and OR count (p=0.7188, ***Figure 1E***). Overall, our findings suggest that insect OR counts correlate with dietary specialization (in both larvae and adults) and habitat type but are not associated with circadian rhythm. We then repeated the phylogenetic generalized least squares (pGLS) analyses using the larger of our structurally intact OR count and the published OR count, and the main ecological associations were retained (***Figure 1—figure supplement 1D-G***). This indicates that the main OR-count ecological analyses are reasonably robust to moderate underestimation caused by missed ORs or strict structural filtering.

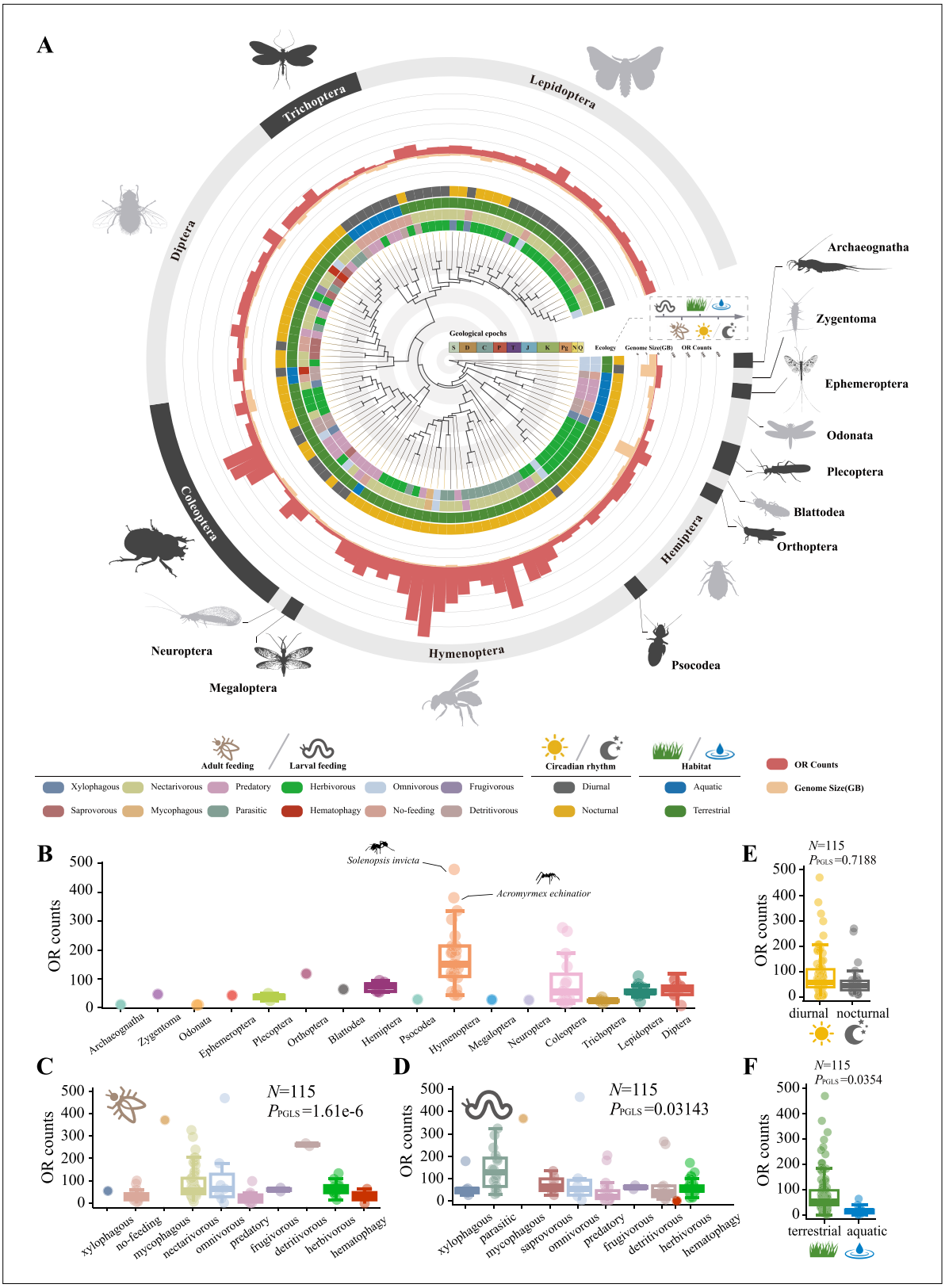


Figure 1. Evolutionary ecology of insect OR repertoires (***Figure 1—source data 1–3***). (**A**) Phylogeny of 115 insect species (see Materials and methodsfor details). Genome size and the number of OR genes for each species are shown as bar plots. Ecological parameters, including larval and adult diets, habitat, and circadian rhythm, are displayed as labels with color coding corresponding to the legend below. Different geological periods are separated by shaded intervals. The tree was visualized using the iTOL (***Letunic and Bork, 2024***). (**B**) Boxplot showing the number of OR genes for

Figure 1 continued on next page

Figure 1 continued

each insect order. (**C–F**) Results of two-sided pGLS tests assessing the relationship between intact OR gene numbers and various ecological parameters: adult diet (**C**), larval diet (**D**), circadian rhythm (**E**), and habitat (**F**). Dots represent OR counts of each species. N represents the counts of species. Animal silhouettes were obtained from PhyloPic.org.

© 2014, Descouens and Keesey. The Rhyacophila dorsalis silhouette is reproduced from PhyloPic, and available under a CC BY-SA 3.0 license.

© 2013, Stemonitis and Keesey. The Petrobius maritimus silhouette is reproduced from PhyloPic, and available under a CC BY-SA 3.0 license.

The online version of this article includes the following source data and figure supplement(s) for figure 1:

Source data 1. Genome information.

Source data 2. Living habits of 115 species.

Source data 3. Insect species tree.

Figure supplement 1. Structural screening strategy and pGLS sensitivity analysis of OR quantity.

OR sequence similarity network

This study includes species representing the full phylogenetic diversity of the class Insecta, and their OR sequences exhibit substantial variation in amino acid similarity (min = 13.7%, max = 97.6%, and mean = 29.6%; ***Figure 2B***). Because of the extensive sequence divergence among ORs from different insect orders, deep relationships are difficult to resolve reliably in a single class-wide phylogeny. We therefore constructed an SSN based on all-vs-all sequence comparisons as a complementary approach to characterize large-scale patterns of OR sequence similarity. In the SSN, each node represents a cluster of ORs sharing at least 60% sequence similarity, and edges are drawn between nodes if the average alignment E value <1e-25 (***Figure 2—figure supplement 1A***). Nodes with a greater number of connections indicate closer evolutionary relationships.

Using the 'trunk–branch' strategy (***Figure 2A***), we identified 37 OR sequence communities (SeqCs), which together account for 94.9% of all ORs . Among them, 22 SeqCs demonstrated clear interaction relationships (***Figure 2C***). The sequence similarity within each SeqC was predominantly between 20% and 30% (***Figure 2—figure supplement 1B***). By mapping previously classified *Acromyrmex echinatior* sequences using phylogenetic methods (***Zhou et al., 2015***), we found that most subfamilies—such as 9-exon, Orco, F, H, B, and A—were independently delineated (***Figure 2—figure supplement 1C***) and exhibited well-defined boundaries (***Figure 2—figure supplement 1E***). Our classification did not disrupt any known subfamilies and instead merged some families with minor differences (***Figure 2—figure supplement 1***, C and D). This comparison indicates that SSN-based communities are broadly consistent with established phylogenetic subfamilies, while capturing sequence-similarity modules rather than strictly monophyletic clades. The SSN is not intended to replace phylogenetic reconstruction, but rather to provide a complementary framework for summarizing sequence relationships among thousands of highly divergent ORs at a class-wide scale. It allows major patterns of OR sequence similarity and diversification to be visualized more clearly across the entire dataset. Phylogenetic analyses can then be used to further resolve evolutionary relationships within specific sequence communities. We further evaluated the SSN using curated lepidopteran pheromone receptors (PRs). To distinguish different PR lineages, experimentally validated classic PRs and the recently reported new PR lineage were analyzed separately (***Bastin-Héline et al., 2019***). Classic PRs were predominantly assigned to SeqC6 (93.9%), whereas the new PR lineage was assigned to SeqC35 (***Figure 2—figure supplement 1F***), supporting the hypothesis of multiple independent origins of PRs (***Bastin-Héline et al., 2019***). These results further support the applicability and robustness of SSN-based community partitioning.

SeqC21 includes the Orco group as well as all ORs from *Thermobia domestica* and *Machilis hrabei* (***Figure 2—figure supplement 1G***), suggesting that ORs in early wingless insects share high sequence similarity with Orco (***Brand et al., 2018***). In addition to SeqC21, most SeqCs are order-specific, indicating substantial sequence divergence among ORs from different insect orders (***Figure 2D***). SeqC26 and SeqC28 are large shared clusters that may contain cross-order orthologous relationships. In SeqC26, we identified 55 orthologous groups, four of which spanned multiple insect orders . OGG54 included ORs from Orthoptera, Blattodea, Trichoptera, and Hymenoptera. This group contains the locust receptors LmigOR5 and LmigOR4. LmigOR5 has been reported to bind geranyl acetone and to be associated with avoidance behavior in the migratory locust (***Chang et al., 2023***). The functions of the corresponding receptors in other insect orders remain to be tested, and we now present these

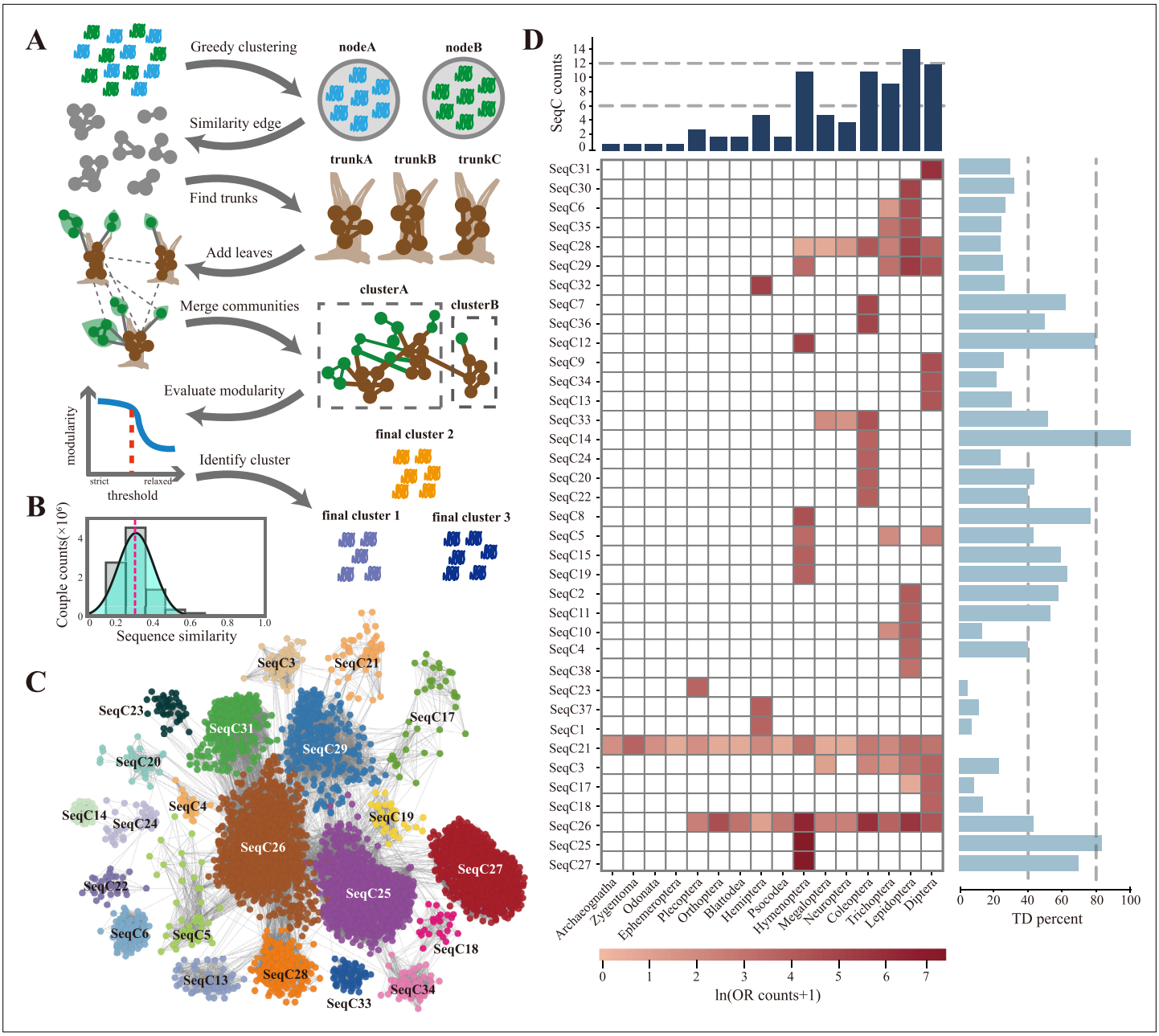


Figure 2. Subfamily classification of ORs based on SSN (***Figure 2—source data 1*** and ***Figure 2—source data 2***). (**A**) Schematic representation of the 'trunk–branch' strategy pipeline for constructing the similarity network and classifying subfamilies. First, all ORs were subjected to greedy clustering based on sequence similarity to generate nodes. The average similarity between nodes was then calculated, and nodes were treated as units to identify core regions (trunks) using a strict similarity threshold. By gradually relaxing the threshold, new core regions and additional nodes (branches) emerged. After a final merging step based on defined criteria, the optimal community structure was determined using modularity as an index (see Materials and methods). (**B**) Distribution of pairwise sequence similarity among all ORs. The red line indicates the mean similarity. (**C**) SSN of ORs constructed using the 'trunk–branch' strategy. Each node represents ORs sharing >60% sequence identity. Edges represent node pairs with an average sequence alignment E-value of $<1 \times 10^{-25}$. Node colors indicate SeqC classifications. (**D**) Heatmap depicting the number of ORs within different SeqCs across various insect orders. Due to substantial differences in OR numbers among orders, the values are ln(count +1)-transformed. The bar plot above the heatmap shows the number of SeqCs present in each order, while the bar plot on the right indicates the proportion of ORs within each SeqC generated by tandem duplication. Two ORs located within 5 kb of each other on the genome were defined as tandem duplicates. TD, tandem duplication.

The online version of this article includes the following source data and figure supplement(s) for figure 2:

Source data 1. Amino acid sequences of insect ORs.

Source data 2. Orthologous groups identified within SeqC26, SeqC28, and StrC23.

Figure supplement 1. SSN intra-subfamily OR sequence similarity and evidence supporting the accuracy of SeqC.

as candidate conserved modules rather than confirmed functional orthologs. In contrast, we did not identify clear cross-order orthologous groups in SeqC28.

Lepidoptera was found to contain the greatest number of SeqCs, reflecting the highest level of OR sequence divergence. Additionally, we identified SeqC25 and SeqC27 as two exceptionally large, Hymenoptera-specific groups, comprising 39.8% and 31.9% of all Hymenopteran ORs, respectively. SeqC27 corresponds to the previously identified Hymenopteran 9-exon family (***Figure 2—figure supplement 1C and D***), which originated through tandem duplication events (***Engsontia et al., 2015***; ***Zhou et al., 2015***; ***McKenzie et al., 2016***; ***Legan et al., 2021***). Interestingly, we found that SeqC25 exhibited an even higher rate of tandem duplication than did SeqC27, suggesting that SeqC25 also expanded through extensive tandem duplication. Notably, among SeqCs with tandem duplication rates exceeding 60%, five of seven are exclusive to Hymenoptera, indicating that OR lineages in this order have undergone particularly intense tandem duplication expansion compared with other insect orders (***Legan et al., 2021***).

Structural diversity of insect ORs

Insect ORs fall within the 'twilight zone' (10–40% sequence identity) of sequence alignment. However, nearly all ORs remain within the 'safe zone' in structural alignment (***Figure 3A***).

To investigate the structural diversity of insect ORs, we constructed a StSN, where each node represents ORs with Z score >50 (from Dali; ***Holm, 2022***). Edges were drawn between nodes if their average Z score was >42.9 (***Figure 3—figure supplement 1A***). Similar to the SSN, we classified 27 structural communities (StrCs), which together accounted for 87.0% of all ORs . Among these, 13 StrCs exhibited distinct interaction relationships (***Figure 3B***). The boundaries between StrCs were well defined (***Figure 3—figure supplement 1C***).

Most StrCs were also order-specific (***Figure 3D***). StrC1 represents the Orco group, which includes 8 *T. domestica* ORs. These ORs were previously identified as Orco-like genes through sequence analysis (***Brand et al., 2018***). Compared to ORs, Orco is characterized by two distinct structural features: a β-sheet in EL2 and a long loop in intracellular loop 2 (IL2; ***Figure 3C***, ***Figure 3—figure supplement 1B***). Lepidoptera presented the highest diversity of StrC types (***Figure 3D***). We found that the transmembrane regions in different StrCs presented low variability, whereas the loop regions presented distinct variable patterns (***Figure 3D***). In both ORs and Orcos, the IL2 region was identified as the most variable region and the only variable segment within the anchor domain. This high degree of variability may be associated with its role in stabilizing and regulating the anchoring domain (***Bahk and Jones, 2016***; ***Jain et al., 2021***; ***Wang et al., 2024***; ***Zhao et al., 2024***).

StrC23 is the only structural community present in 12 insect orders and accounts for 46.3% of all structurally intact ORs in the dataset. We found that StrC23 has a much larger binding-pocket volume than other ORs and Orco. From a structural perspective, this suggests that StrC23 receptors may have greater potential to accommodate diverse VOCs (***Figure 3—figure supplement 1D***). To further investigate the evolutionary basis of this broadly shared structural feature, we examined the phylogenetic and orthology relationships among StrC23 receptors. StrC23 was divided into 156 orthologous groups, 24 of which contained ORs from more than 10 species. However, only a few of these larger groups spanned multiple insect orders: one included ORs from Hymenoptera, Diptera, and Trichoptera, while two included ORs from both Lepidoptera and Trichoptera. Thus, StrC23 does not represent a single conserved cross-order orthologous lineage but instead comprises multiple phylogenetically distinct OR lineages. Its shared large-pocket architecture therefore cannot be explained solely by conservation of a single orthologous lineage and may reflect either structural convergence or retention of an ancient structural feature followed by substantial sequence divergence.

Notably, StrC17 exhibited a high degree of flexibility in the IL3 region. We observed that a subset of ORs within this group possessed unusually long IL3 regions (***Figure 3E***). These ORs were distributed across 16 families from two suborders of Diptera but were absent from the earliest-diverging families in the Nematocera suborder, including Culicidae, Chironomidae, and Tipulidae (***Figure 3—figure supplement 1G and H***). After manual inspection, annotation errors were ruled out. The construction of a maximum likelihood tree for StrC17 ORs revealed that ORs with long IL3 regions formed a distinct clade, with one subclade identified as orthologous (***Figure 3—figure supplement 1F***). To investigate the evolutionary pressures acting on these ORs, we calculated the dN/dS ratios for different branches (***Figure 3F***). The results revealed that all StrC17 ORs were subject to strong purifying selection (dN/dS

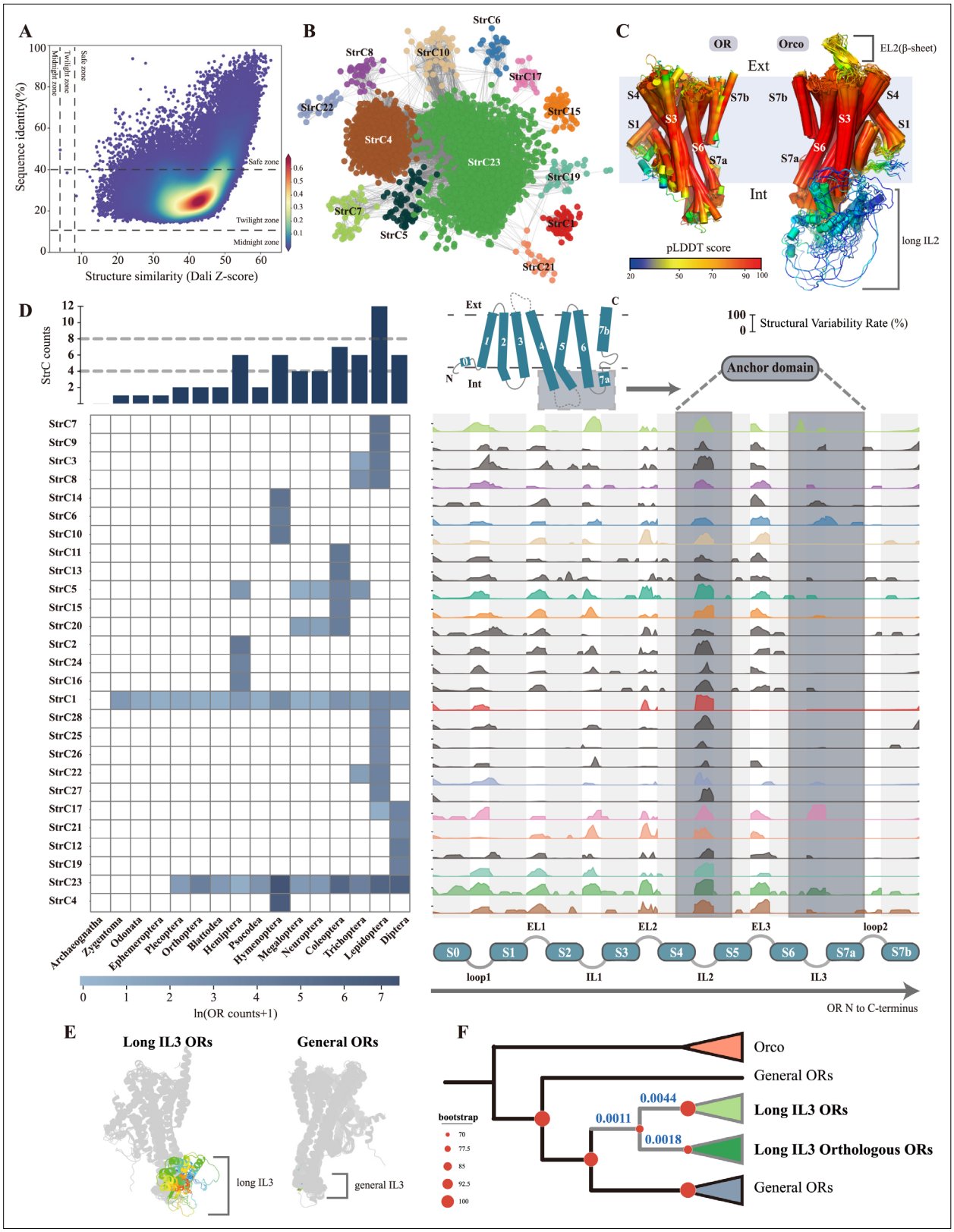


Figure 3. StSN and structural community diversity of ORs (***Figure 3—source data 1***). (**A**) Sequence similarity versus structural similarity of insect ORs, assessed using pairwise comparisons with Dali. Each dot represents comparison between a pair of ORs, with color indicating point density. (**B**) StSN of ORs constructed using the 'trunk–branch' strategy. Nodes represent ORs with pairwise Dali Z-scores >50. Edges connect nodes with an average pairwise Z-score of at least 42.9. Node colors denote StrCs classifications. (**C**) Structural comparison of ORs and Orco. The structures were aligned using US-align,

Figure 3 continued on next page

Figure 3 continued

with multiple structural alignments performed for annotated Orco and OR structures (visualized in PyMol). Colors represent pLDDT scores provided by AlphaFold2, indicating the confidence of structural modeling. (**D**) Visualization of differences among StrCs. The heatmap shows the number of ORs within different StrC across insect orders, with OR counts ln(count +1)-transformed due to large inter-order variations. The bar plot above the heatmap indicates the number of StrC in each order. On the right, a ridge plot illustrates the flexibility rates of different structural regions across StrCs. Flexibility rate = 1 - conservation rate. Light gray represents transmembrane regions, white denotes loop regions, and dark gray marks the anchor domain of ORs. Colors correspond to StrCs, with gray indicating StrCs without interactions. (**E**) Two types of ORs in StrC17—long IL3 ORs and general ORs—with the IL3 regions highlighted in color. (**F**) Phylogenetic analysis of StrC17 ORs based on amino acid sequences. Red circles indicate bootstrap values, with most ORs collapsed based on bootstrap support. Gray branches represent the clade of long IL3 ORs, and the dN/dS values are shown in blue on the corresponding evolutionary branches.

The online version of this article includes the following source data and figure supplement(s) for figure 3:

Source data 1. The likelihood ratio test results for StrC17 ORs.

Figure supplement 1. StrC classification and the phylogenetic tree of StrC17 ORs.

= 0.17). The amino acid substitution rate of ORs with long IL3 regions was two orders of magnitude lower than that of other ORs (dN/dS = 0.001–0.004), indicating intensified purifying selection.

Structural specialization of Orco

Orco evolved from ORs in the common ancestor of Zygentoma and Pterygota (***Brand et al., 2018***; ***Thoma et al., 2019***). The evolutionary drivers underlying Orco specialization remain unresolved. With the currently available early-diverging insect genomes, we could not reliably identify direct orthologous relationships among early ORs, which limits the reconstruction of ancestral OR states. The *T. domestica* ORs TdomOR1-8 are classified within Orco communities (SeqC21 and StrC1) and occupy a transitional phylogenetic position between ORs and characterized Orcos, representing the closest ORs related to Orcos (***Brand et al., 2018***). Because they are the extant ORs most similar to Orco in sequence, structure, and phylogenetic position, we used them as comparative proxies for candidate Orco. Comparisons between TdomOrcoFJ and the other TdomOR1–8 receptors may therefore identify structural features potentially associated with Orco recruitment and specialization in the common ancestor of Zygentoma and Pterygota. We found that TdomOrcoFJ has the smallest, least polar, and most hydrophobic binding pocket among TdomOR1-8 (***Figure 4A***). Molecular docking analysis revealed that TdomOrcoFJ is less suited for VOC binding than other candidates are (***Figure 4—figure supplement 1A***). These findings suggest that binding-pocket specialization may have been one of the structural factors associated with the evolutionary recruitment of Orco.

The reasons behind the loss of Orco function remain unclear, although its conserved binding pocket may act as a molecular sieve (***Pacalon et al., 2023***). We found that *M. hrabei* ORs lack the EL2 β-sheet but have a long IL2 loop region. Similarly, most *T. domestica* ORs also have a long IL2 loop region, whereas only TdomOR1-8 retains the EL2 β-sheet (***Figure 4A***). The long IL2 loop was present in ORs from early-diverging insects as well as in candidate Orco and therefore does not provide a distinguishing structural criterion for candidate Orcos. In contrast, the EL2 β-sheet was consistently present in TdomOR1–8 and highly conserved across characterized Orcos. This distribution suggests that the EL2 β-sheet may have emerged as a key structural innovation during Orco specialization and subsequently become conserved in the Orco lineage.

The formation of the EL2 β-sheet may be associated with the loss of Orco function. We observed that the EL2 region of Orco is distinctly divided into two parts: the β-sheet region and the transmembrane loop region (***Figure 4B***). In the β-sheet region, the proportion of hydrophilic amino acids is 47.8%, whereas it is 14.2% in the transmembrane loop region and 18% in ORs. Overall, across all the insects, the β-sheet region of Orco EL2 presented a significantly greater proportion of hydrophilic amino acids than did the ORs (***Figure 4—figure supplement 1B***), facilitating its stable presence in the extracellular aqueous environment. The extracellular port is the main pathway through which ligands enter the binding pocket (***Renthal and Chen, 2022***; ***Pacalon et al., 2023***). Using molecular dynamics (MD) simulations to analyze VOC movement outside the membrane, we found that the presence of the β-sheet in the EL2 region significantly reduced the likelihood of VOCs reaching the binding pocket (***Figure 4C***, ***Figure 4—figure supplement 1C*** and ***Figure 4—video 1***). Using *Drosophila* OR47a as an example, we demonstrated through calcium imaging experiments that replacing the EL2 region of OR47a with that of Orco resulted in a mutant OR47a (el2. DmOR47a) that exhibited markedly reduced

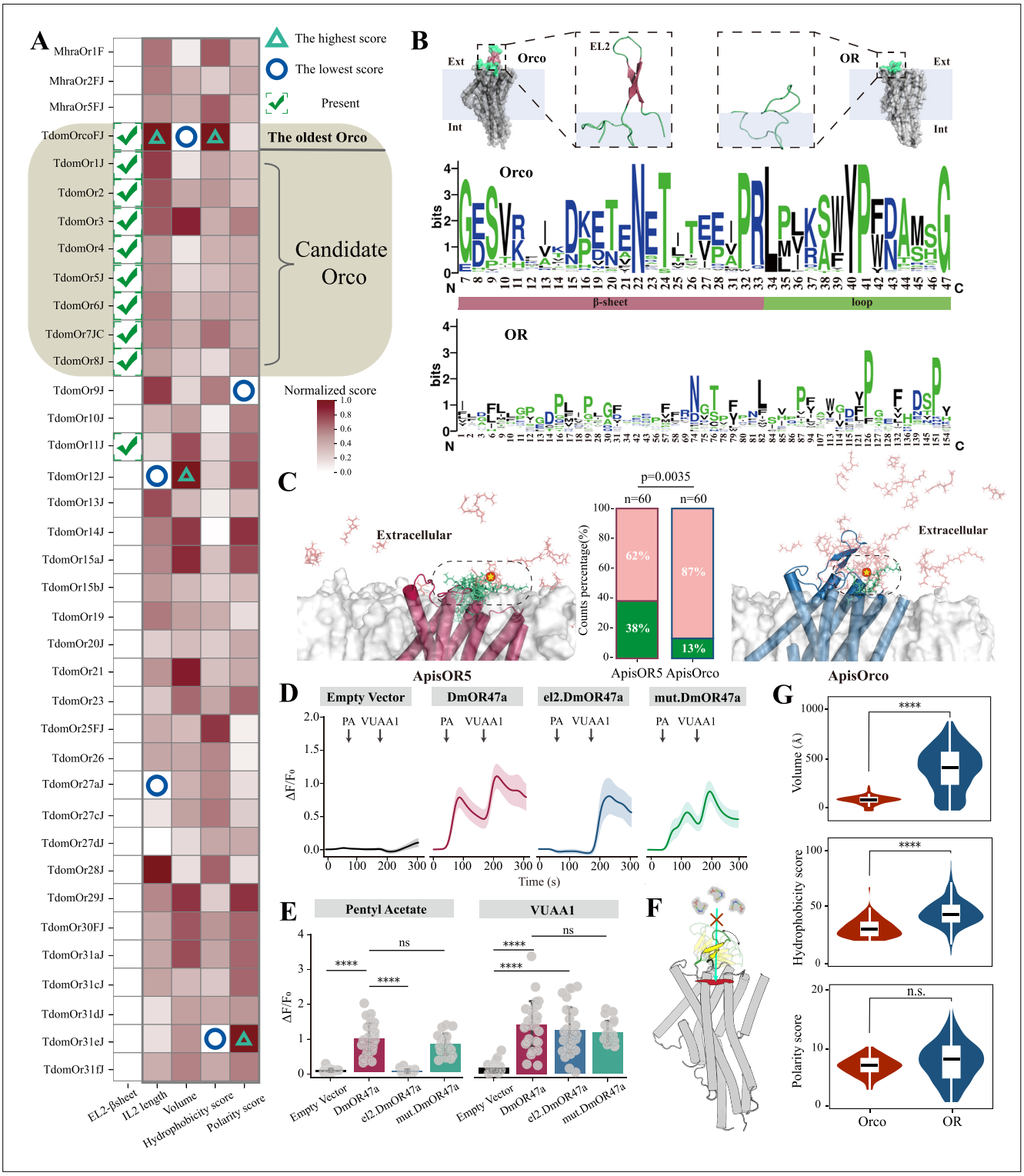


Figure 4. Structural features associated with Orco specialization (***Figure 4—source data 1 and 2***). (**A**) Differences in EL2, IL2, and binding pocket properties between *M. hrabei* and *T. domestica* ORs. Only ORs with closed binding pockets were analyzed, and all scores were normalized. Circles and triangles represent the minimum and maximum scores after column normalization, respectively. The presence of an EL2 β-sheet is indicated with a check mark. (**B**) Differences in amino acid properties of the EL2 region between OR and Orco. Hydrophilic amino acids are shown in blue, neutral amino acids in green, and hydrophobic amino acids in black (visualized by WebLogo ***Crooks et al., 2004***). The x-axis numbers represent amino acid positions from the multiple sequence alignment, with highly gapped positions removed. (**C**) MD simulation-based analysis of the movement trajectory of VOC (geranyl acetate) near Orco and OR. The red protein represents ApisOR5, and the blue protein represents ApisOrco. The green area indicates the probability

Figure 4 continued on next page

Figure 4 continued

of VOC remaining above the binding pocket, while the pink represents its probability of occupying other regions of the protein (see Materials and methods). Statistical significance was evaluated using a Chi-Squared test. The red pentagram indicates the initial position of geranyl acetate in the simulation. (**D**) Changes in relative calcium fluorescence intensity (ΔF/F0) in cells. Cells were stimulated with 100 μM of the OR47a agonist pentyl acetate (PA) at 50 s and 100 μM the Orco agonist VUAA1 at 180 s. el2.DmOR47a: OR47a mutant in which the EL2 region of OR47a was replaced with that of Orco. mut.DmOr47a: the EL2 region of wild-type DmOr47a contains an inserted short peptide. The total number of cells (n=30) was derived from three independent replicate experiments. The shading indicates the 95% confidence interval. (**E**) Relative calcium fluorescence intensity in cells following stimulation with PA and VUAA1, respectively. Error bars represent means ± SEM. n=30. Statistical significance was determined using Tukey's HSD test. (**F**) Schematic illustration of the mechanism by which the β-sheet structure in the EL2 region of Orco influences VOC proximity to the binding pocket. (**G**) Comparative analysis of binding pocket parameters (volume, hydrophobicity score, and polarity score) between Orco and ORs across all studied insect species. Statistical significance was determined using a two-sided t-test. **** p<0.0001; n.s., not significant.

The online version of this article includes the following video, source data, and figure supplement(s) for figure 4:

Source data 1. Calcium fluorescence responses in cells transfected with each plasmid construct.

Source data 2. Structural and binding-pocket parameters of MhORs and TdomORs.

Figure supplement 1. The driving factors underlying the origin of Orco.

Figure 4—video 1. Molecular dynamics trajectories of VOCs above the binding pockets of OR and Orco.
https://elifesciences.org/articles/110877/figures#fig4video1

calcium responses to pentyl acetate (PA; ***Figure 4D and E*** and ***Figure 4—figure supplement 1D***). Additionally, we inserted a non-β-sheet short peptide into the EL2 region of OR47a, and the results showed that its function remained unaffected (***Figure 4D and E*** and ***Figure 4—figure supplement 1D***). This finding suggests that the β-sheet acts as a spatial barrier, limiting VOC access to the binding pocket as a primary physical gate (***Figure 4F***). Furthermore, we observed that Orco has a smaller and less hydrophobic binding pocket than ORs do (***Figure 4G***). These characteristics are unfavorable for VOC entry into the binding pocket, representing the key chemical factors contributing to Orco's loss of natural VOC-binding function.

Classification of ORs based on docking-profile similarity

Currently, molecular docking and virtual screening have been successfully applied to the high-throughput prediction of OR binding profiles in mammals, facilitating the exploration of macroevolutionary patterns in predicted olfactory recognition (***Zhang et al., 2024***; ***Chi et al., 2025***). To better compare the potential binding profiles of insect OR repertoires and classify them accordingly, we collected 14,499 VOCs from four molecular libraries, including pheromones (***Li et al., 2023***), fragrances (thegoodscentscompany.com), food odors (***Chi et al., 2025***), and theoretical odors (***Mayhew et al., 2022***). These VOCs were categorized into nine functional groups (***Figure 5A***). Although terpenoids are important in insect chemical ecology, they are distributed across multiple regions of odor space (***Figure 5—figure supplement 1C***). To reduce the docking instability that could result from excessive VOC subdivision, we did not treat terpenoids as a category parallel to the nine basic functional-group categories. We used this comprehensive VOC library to systematically evaluate the docking-derived binding profiles of ORs, aiming to characterize the potential ligand-binding diversity encoded by insect OR repertoires, even though some of these compounds may never be encountered by insects in nature.

We performed molecular docking for all insect ORs using these VOCs to generate their docking-derived binding profiles (***Figure 5—figure supplement 1A***). A long-standing challenge in molecular docking is how to predict binding likelihood on the basis of ranking scores. To address this issue, we modeled a theoretical 'hit rate' curve using docking scores and experimental datasets from *Drosophila* (***Hallem and Carlson, 2006***), mosquitoes (***Carey et al., 2010***), and moths (***de Fouchier et al., 2017***; ***Guo et al., 2021***; ***Figure 5B***). Using a Bayesian framework with prior probability distributions, we determined an informative docking-score range for insect ORs. A threshold score of –8 kcal/mol was established to maximize data retention while maintaining a reliable hit rate (***Figure 5B***). In addition, our single-taxon hit-rate sensitivity analysis showed similar trends for *Drosophila*, *Anopheles*, and moths (***Figure 5—figure supplement 1D–F***). Because the locust OR repertoire had an extremely low positive rate, we did not include the locust experimental dataset in the final threshold assessment (***Figure 5—figure supplement 1G and H***, see Appendix 2 for details). Although this threshold enriches experimentally responsive OR–VOC pairs, the resulting hit rate is not sufficient to support

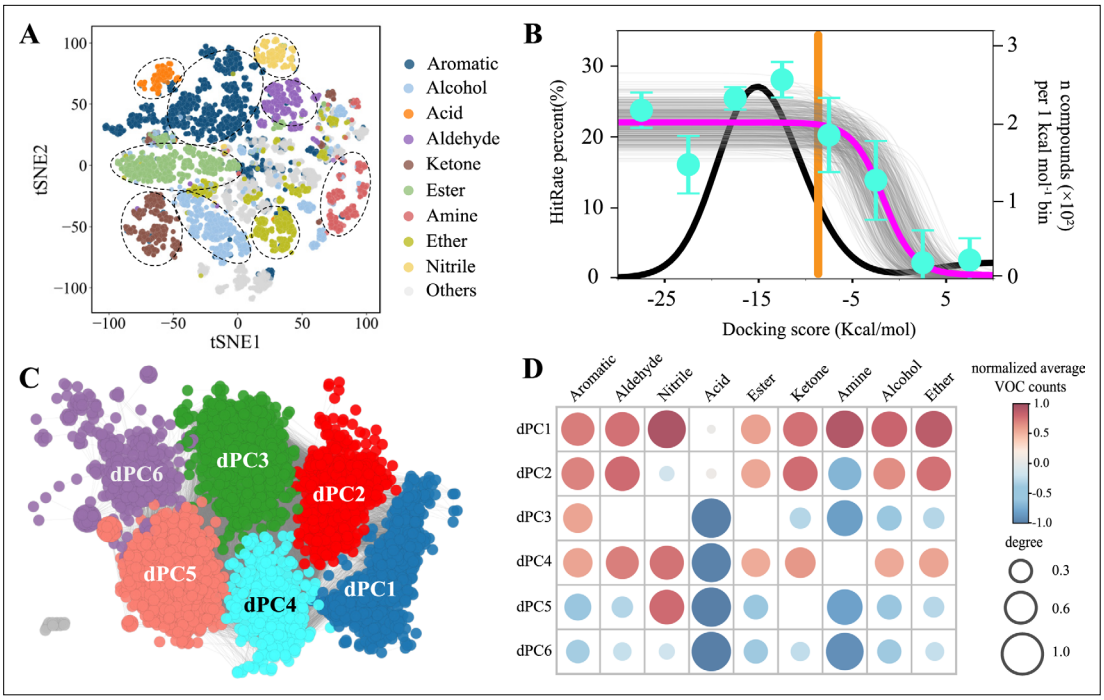


Figure 5. Classification of insect OR docking-profile communities (***Figure 5—source data 1 and 2***). (**A**) 2D odor space based on molecule physicochemical properties. Each dot represents an odor molecule, with different colors indicating different functional groups. (**B**) Relationship between hit rate and docking score, inferred using Bayesian statistics. The black curve shows the distribution of hit rates and docking scores across all OR–VOC docking results. The top plateau has a hit rate of 22%, while the bottom plateau has a hit rate of 0%. Cyan dots represent the mean hit rate ± SEM for each docking score interval. n = 3.The orange line indicates the peak hit rate at a docking score of –8 kcal/mol. The gray curve represents posterior distributions from Bayesian inference, with n=500. (**C**) dPSN and dPC of ORs. Nodes represent ORs with pairwise PCCs >0.6 based on docking scores. Edges represent node pairs with an average PCC ≥0.48. (**D**) Labels for the recognition of different functional group small molecules by each dPC. Red indicates a tendency to recognize a higher number of odor molecules from a particular functional group, while blue indicates a tendency to recognize fewer. Circle size represents the degree of this tendency.

The online version of this article includes the following source data and figure supplement(s) for figure 5:

Source data 1. The physicochemical properties and functional group labels of the 14,499 VOCs.

Source data 2. Matrix of all OR–VOC docking scores.

Figure supplement 1. Insect OR–VOC binding profiles, dPC labeling accuracy, and distribution of docking-profile communities across insect orders.

Figure supplement 2. Bayesian prior modeling to optimize VOC selection by balancing information acquisition and ligand discovery.

deterministic inference for individual receptor-ligand pairs. Therefore, subsequent analyses were interpreted at the OR repertoire level as docking-derived comparative patterns rather than as experimentally validated OR response spectra.

We constructed a dPSN for ORs (***Figure 5C***). In this network, each node represents ORs with a docking-profile Pearson correlation coefficient (PCC) greater than 0.6, and nodes with an average PCC greater than 0.48 are connected by edges. We identified six docking-profile communities (dPCs), representing 72.5% of the ORs (***Figure 5—figure supplement 1B***). It should be noted that dPCs represent docking-profile communities rather than experimentally confirmed functional classes. We subsequently quantified the number of VOCs from different functional groups meeting the docking threshold for each dPC (***Figure 5—figure supplement 1I***) and assigned docking-derived binding labels on the basis of their relative enrichment compared with the background (***Figure 5D***). These labels summarize the predicted breadth of VOC functional-group binding for each dPC. To evaluate the extent to which these docking-derived labels were consistent with available experimental data, we calculated the area under the receiver operating characteristic curve (AUC) by comparing experimental OR response data with the corresponding docking-derived labels. For most functional group predictions, an AUC exceeding 0.75 was achieved, suggesting that these labels capture broad directional tendencies in available experimental data (***Figure 5—figure supplement 1J***). dPC1 showed a relatively broad docking profile across VOC functional groups, whereas dPC6 showed a narrower

docking profile (***Figure 5D***). Additionally, dPC3 showed relatively stronger predicted binding toward aromatics, whereas dPC5 showed relatively stronger predicted binding toward nitriles.

Unlike SeqC and StrC, almost all insect orders contain all dPCs without exhibiting specificity (***Figure 5—figure supplement 1K***). dPC1, characterized by a relatively broad docking profile, accounted for the largest proportion of ORs in most orders, particularly Psocodea, Diptera, and Lepidoptera (***Figure 5—figure supplement 1L***). In contrast, dPC6, characterized by a relatively narrow docking profile, represented only a small proportion of ORs across the sampled species.

Comparative patterns of insect OR diversification across evolutionary periods

Although the sequence evolution of the insect OR gene family and functional diversity among different orders and species have been reported multiple times (***Brand and Ramírez, 2017***; ***McKenzie and Kronauer, 2018***; ***Legan et al., 2021***; ***Chang et al., 2023***), these studies reveal only a partial view rather than the broader patterns of insect OR sequence, structural, and docking-profile diversity. To characterize large-scale diversification patterns of insect ORs, we integrated SeqCs, StrCs, and dPCs. The results revealed strong congruence between SeqCs and StrCs, with both communities including nearly all dPCs (***Figure 6***, ***Figure 6—figure supplement 1***). This pattern suggests that sequence and structural diversification are more closely coupled, whereas docking-profile variation shows a weaker correspondence with either level.

Additionally, extant holometabolous insects showed higher OR sequence-community diversity than non-holometabolous insects (***Figure 6—figure supplement 2***). Because total SeqC counts may be affected by unequal species sampling and OR numbers, we added normalized analyses. First, we calculated SeqC numbers at the species level and found that holometabolous insects contained more SeqCs on average (***Figure 6—figure supplement 3A***). Second, because OR number was significantly correlated with SeqC count (p=0.001035), we performed equal-OR random sampling between holometabolous and non-holometabolous insects. Under the same OR sampling depth, holometabolous insects still covered more SeqCs, indicating higher OR sequence-community diversity after controlling for OR number (***Figure 6—figure supplement 3B***). Therefore, this result supports a genuine difference in SeqC diversity rather than a difference driven by unequal species sampling or OR number. The higher SeqC diversity observed in Holometabola was retained when the comparison was restricted to Neoptera, indicating that this pattern is not simply a restatement of previously reported Paleoptera-Neoptera differences (***Figure 6—figure supplement 4***).

The number of SeqCs increased with deeper divergence among insect orders, suggesting gradual accumulation of OR sequence variation during insect evolution (***Figure 6—figure supplement 3C***). Despite the limited SeqC diversity in non-holometabolous insects, all six dPC types were represented in these extant lineages, suggesting that the major docking-profile community types may have arisen relatively early in insect evolution.

The EPME was a significant geological event that resulted in the extinction of most species and caused drastic environmental changes (***Close et al., 2020***; ***Jouault et al., 2022***). In the extant species sampled here, we detected differences in dPC composition between lineages whose order-level origins fall before and after the EPME (***Figure 6—figure supplement 3D***). Lineages originating after the EPME showed a greater proportion of dPC1, which was characterized by a broader docking profile, and a lower proportion of dPC2 (***Figure 6—figure supplement 3E***). This pattern is consistent with previously reported data that the proportion of broad-tuned ORs in *Locusta* is much smaller than that in *Drosophila* and *Anopheles* (***Chang et al., 2023***). These observations should be interpreted as a docking-derived comparative pattern among extant lineages rather than direct evidence of ancestral functional changes across the EPME.

Discussion

In this study, we integrated sequence, structural, and docking-profile analyses to characterize large-scale patterns of insect OR diversification. Across the sampled insect species, OR repertoire size was associated with several ecological traits. Sequence and structural communities showed a close correspondence, whereas docking-profile communities were less closely associated with either classification. Differences in dPC composition among extant lineages whose order-level origins predate

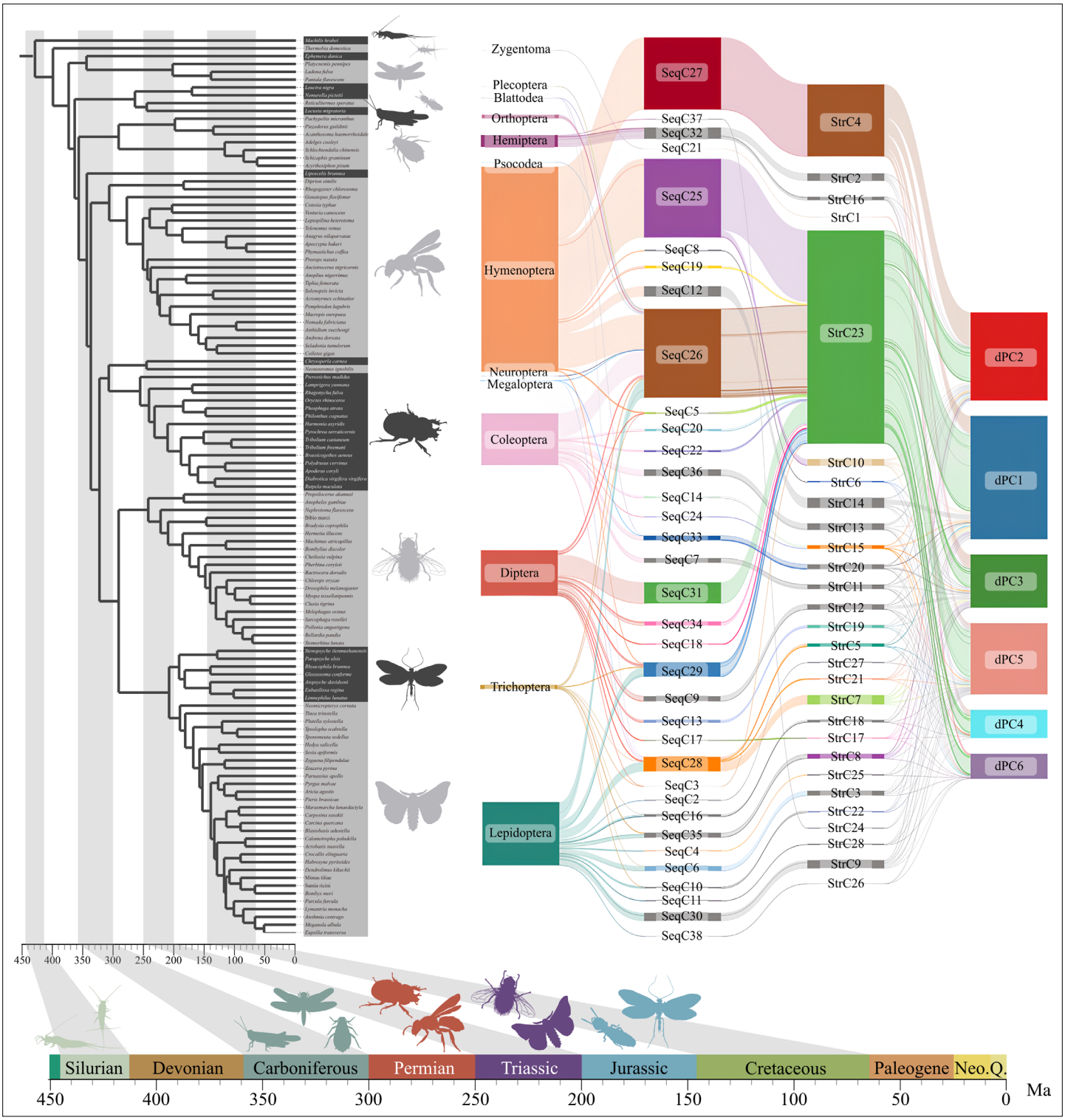


Figure 6. Relationships between OR sequence, structure, and docking-profile communities across insect orders (***Figure 6—source data 1***). The phylogenetic tree follows the structure from ***Figure 1***. On the right, the mapping relationships between all insect OR SeqCs, StrCs, and dPCs are shown, considering only ORs classified into communities. Time nodes for different geological periods are based on the previous study (***Misof et al., 2014***) and are separated by gray shading. The timeline beneath the tree marks the origin of various insect orders. Animal silhouettes were obtained from PhyloPic.org.

© 2014, Descouens and Keesey. The Rhyacophila dorsalis silhouette is reproduced from PhyloPic, and available under a CC BY-SA 3.0 license.

© 2013, Stemonitis and Keesey. The Petrobius maritimus silhouette is reproduced from PhyloPic, and available under a CC BY-SA 3.0 license.

Figure 6 continued on next page

Figure 6 continued

The online version of this article includes the following source data and figure supplement(s) for figure 6:

Source data 1. Community annotation of insect ORs.

Figure supplement 1. Relationships between OR sequence, structure, and docking-profile communities across insect orders (all ORs).

Figure supplement 2. Sequence, structural, and docking-profile community associations of ORs in holometabolous and non-holometabolous insects.

Figure supplement 3. Changes in insect OR SeqC types and dPC proportions across different geological periods.

Figure supplement 4. Comparison of OR SeqC diversity between Holometabola and non-holometabolous Neoptera.

or postdate the EPME raise the hypothesis that these differences may be associated with ecological changes surrounding this event. Within the 'conserved chassis-diversified sensors' model, we further examined Orco specialization and identified the conserved EL2 β-sheet and distinctive binding-pocket properties as structural features that may contribute to reduced odorant-binding capacity.

The long-IL3 OR orthologous group includes *Drosophila melanogaster* DmOr13a and *Bactrocera dorsalis* BdorOR13a, both of which respond to 1-octen-3-ol and belong to a group of dipteran fly OR homologs associated with oviposition behavior (***Kreher et al., 2008***; ***Liu et al., 2023***; ***Xu et al., 2023***). Modeling of these homologous ORs showed that they also contain a long IL3 region. We therefore propose that long-IL3 ORs correspond to these fly homologs that recognize 1-octen-3-ol. However, 1-octen-3-ol has different behavioral functions in different species: it acts as an attractant in blood-feeding mosquitoes and tsetse flies (***Hall et al., 1984***; ***Kline et al., 2007***), whereas in flies it is usually associated with oviposition regulation (***Kreher et al., 2008***; ***Liu et al., 2023***). Mosquito receptors known to detect 1-octen-3-ol, such as AgOR8, AaOR8, TaOR8, and CquiOR118b (***Lu et al., 2007***; ***Xu et al., 2015***; ***Dekel et al., 2016***; ***Frunze et al., 2024***), do not contain a long IL3 region (***Figure 3—figure supplement 1E***). This suggests that the long IL3 structure in fly homologs may be related to the oviposition-associated role of 1-octen-3-ol in flies, which requires future experimental validation.

Despite substantial differences in OR sequence and structural community composition, nearly all sampled insect orders contained representatives of all major docking-profile communities. SeqCs and StrCs showed a closer correspondence with each other than either did with dPCs, suggesting that sequence and structural diversification are more tightly coupled, whereas variation in docking profiles is less directly associated with either level. Thus, distinct sequence and structural repertoires can converge on broadly similar distributions of docking-profile communities across insect orders. This pattern suggests that diversification in sequence and structure does not necessarily translate into equivalent divergence in predicted ligand-binding profiles. We also detected differences in dPC composition between extant lineages whose order-level origins predate or postdate the EPME. In particular, lineages whose order-level origins postdate the EPME had a higher proportion of dPC1, a community characterized by a relatively broad docking profile. The Permian–Triassic transition was accompanied by extensive ecological disruption and major restructuring of terrestrial ecosystems, including substantial changes in plant communities and insect–plant interactions (***Grauvogel-Stamm and Ash, 2005***; ***Benton, 2016***; ***Vajda et al., 2020***; ***Lu et al., 2021***; ***Jouault et al., 2022***). These environmental changes raise the hypothesis that ecological turnover surrounding the EPME may have contributed to differences in OR docking-profile composition among insect lineages. However, this interpretation remains hypothetical. Because rapid OR evolution and limited cross-order orthology preclude reliable reconstruction of deep ancestral repertoires, the EPME-related pattern should be viewed as a descriptive comparison among extant lineages. Further validation will require broader taxon sampling, ancestral reconstruction, and experimental testing of receptor tuning breadth.

Previous studies have proposed an evolutionary transition from ancestral chemoreceptors to the insect OR–Orco system, and our results provide structural insight into the specialization of Orco within this framework (***Brand et al., 2018***; ***Thoma et al., 2019***). The ancestral insect recruited GRs to perform olfactory functions, giving rise to the earliest ORs (***Robertson et al., 2003***; ***Frank et al., 2024***; ***Mitchell et al., 2026***). These early ORs retained the homotetrameric architecture of GRs (***Del Mármol et al., 2021***). However, the sensitivity requirements of olfaction and gustation differ (***de Fouchier et al., 2017***; ***Pask et al., 2017***; ***Slone et al., 2017***; ***Dweck and Carlson, 2023***; ***Wang et al., 2024***; ***Zhao et al., 2024***). This issue appears to have been first addressed in the ancestor of Zygentoma and Pterygota by reducing the number of functional subunits while maintaining the tetrameric structure, thereby increasing sensitivity (***Brand et al., 2018***; ***Thoma et al., 2019***; ***Wang et al., 2024***;

Zhao et al., 2024). Two structural features are consistently conserved among Orcos: the EL2 β-sheet and the highly specialized binding pocket. The β-sheet physically prevents VOCs from accessing the binding pocket, whereas the chemically restrictive nature of the binding pocket prevents VOCs from penetrating deeply. Orco exhibits strong sequence conservation, ensuring that these chemical and physical barriers are effectively preserved. Sequence similarity is much lower in the insect OR family than in the GR family, which may be related to the emergence of Orco. This division between a conserved co-receptor and diversified ligand-selective subunits has been proposed to provide a stable structural framework within which OR diversification can occur (*Zhao et al., 2024*). However, we also identified several ligand-selective ORs containing an Orco-like EL2 β-sheet. Whether this structural similarity reflects shared ancestral features, convergent structural evolution, or other evolutionary processes remains unresolved (*Figure 4—figure supplement 1E and F*).

Overall, we characterized the relationships among sequence, structure, and docking profiles in insect ORs, identified structural features potentially associated with Orco specialization, and detected class-wide associations between OR repertoire size and ecological traits. The resulting dataset and similarity-based community framework provide a resource for insect OR research, while the docking-profile patterns offer testable hypotheses for future experimental and evolutionary studies.

Materials and methods

Genome data

We compiled data from all published insect genomes available in the NCBI public database as of December 2022. Genomes with a contig N50>1 Mb were considered reliable for OR annotation. When multiple high-quality genomes were available within a single insect family, we selected the genome with the highest assembly quality as the representative. After filtering, 107 insect genomes were included: Odonata (2), Plecoptera (2), Blattodea (1), Hemiptera (6), Psocodea (1), Hymenoptera (22), Megaloptera (1), Neuroptera (1), Coleoptera (14), Trichoptera (7), Lepidoptera (30), and Diptera (20). Detailed genome information is provided in the *Figure 1—source data 1*. To maximize phylogenetic diversity, we added eight representative species from different orders with annotated ORs: *M. hrabei* (*Brand et al., 2018*), *T. domestica* (*Brand et al., 2018*), *Ladona fulva* (*Brand et al., 2018*), *Ephemera danica* (*Brand et al., 2018*), *Locusta migratoria* (*Wang et al., 2015*), *Acyrthosiphon pisum* (*Robertson et al., 2019*), *Acromyrmex echinatior* (*Zhou et al., 2015*), and *Tribolium castaneum* (RRID:SCR_010440) (*Engsontia et al., 2008*). In total, our study includes 115 insect species, covering 16 orders and 111 families.

Species phylogenies

To construct the phylogenetic tree used in this study, we merged existing phylogenies across different insect orders (*Misof et al., 2014*). For relationships among orders within Insecta, we referred to previously published phylogenetic frameworks. To ensure broad taxonomic representation, we included 16 insect orders in our analysis. Intra-order relationships were obtained from insectphylo.org and other published species trees covering various insect taxa (*Chesters, 2017*). For species with available genomes that were not included in existing phylogenies, we inferred their phylogenetic placements based on family-level information. Divergence times among orders were obtained from TimeTree.org (*Kumar et al., 2022*) and calibrated using MCMCtree (*Rannala and Yang, 2007*). The resulting phylogeny includes 115 insect species spanning 16 orders.

Ecological trait data of insect species

Given the vast diversity and ecological complexity of insects, and the lack of a comprehensive authoritative database, we manually retrieved ecological trait data from published literature for each species. We characterized insect ecology based on four aspects: larval diet, adult diet, habitat, and circadian rhythm. For dietary traits, we classified diets into 12 types: herbivorous, xylophagous, hematophagy, mycophagous, non-feeding, omnivorous, predatory, saprovorous, frugivorous, nectarivorous, detritivorous, and parasitic. For habitat, following previous classification criteria, species were assigned as either terrestrial or aquatic (*Dijkstra et al., 2014*). Insects that live entirely or have at least one life stage in water were considered aquatic; those that are strictly land-dwelling throughout their life cycle were considered terrestrial. Circadian rhythm was classified based on adult activity periods as

either diurnal or nocturnal. The ecological classification and corresponding literature sources for each species are provided in the ***Figure 1—source data 2***.

OR gene mining and structural filtering

All OR and Orco protein sequences were downloaded from NCBI, supplemented with ORs reported in the paper, yielding a total of 16,176 ORs. Redundancy and data cleaning were then performed using CD-HIT (***Li and Godzik, 2006***; ***Fu et al., 2012***) v4.8.1 to remove highly similar sequences with >90% identity, resulting in 6029 non-redundant ORs used as queries. We used Exonerate (***Slater and Birney, 2005***) v2.4.0 to search whole-genome data and identify potential exon regions. To improve annotation efficiency, genome contigs were split into 20 Mb segments, with 20 kb flanks added at breakpoints to avoid truncating ORs. Next, InsectOR (***Karpe et al., 2021***) was used for OR gene annotation.

Given that multi-exon gene annotation from genome can be error-prone, we applied the strictest structural filtering strategy to enhance annotation reliability (***Figure 1—figure supplement 1A***). All annotated ORs were modeled using AlphaFold2 (***Jumper et al., 2021***). Structural alignment against the cryo-EM structure of MhOR5 (PDB: 7LIC) was performed using Dali (***Holm, 2022***), and ORs with Z-scores <20 were excluded. We further manually removed sequences lacking at least seven transmembrane domains. In total, 8905 ORs with complete and credible structural models were retained for downstream analyses. This rigorous approach eliminates structurally compromised ORs that may appear intact at the sequence level, thereby increasing the likelihood of identifying ORs with functional physiological and biochemical properties.

Phylogenetic comparative analyses

The function pGLS of the R package caper was used to perform all phylogenetic generalized least squares models presented in this study. Phylogenetic manipulations were conducted using the R packages phytools (RRID:SCR_015502) and ape. Phylogenetic trees were visualized by iTOL (***Letunic and Bork, 2024***; https://itol.embl.de/). Animal silhouettes used in this study were retrieved from http://phylopic.org/.

'Trunk–Branch' strategy

Most existing insect ORs have arisen through gene duplication events (***McKenzie and Kronauer, 2018***; ***Legan et al., 2021***). Initially, these duplicated OR sequences were highly similar, forming the 'trunk' of core communities. Over time, mutations led ORs to diverge from their original sequences and functional features, forming the 'branches' of the community (***Figure 2A***). Based on this 'trunk–branch' concept, we developed a strategy to construct similarity networks and define communities and the specific implementation process is as follows:

1. Communities with more than 20 nodes in the initial-threshold network are defined as core communities (trunks). The initial thresholds were set at e-value=1e-68 for the SSN and Z-score=45.9 for the StSN.
2. Thresholds were gradually relaxed to extend the branches of each core community. Newly appearing nodes were assigned to the community with which they shared the highest number of edges. The SSN threshold was relaxed in logarithmic steps (e.g. 1e-68, 1e-67,...), while the StSN threshold was relaxed in steps of 0.1 Z-score (e.g. 45.9, 45.8,...).
3. Communities were merged based on their inter-community connectivity. If the proportion of edges connecting two communities exceeded 50% for either community, they were merged. This process was repeated until all inter-community connectivities fell below the 50% threshold.
4. New communities with more than 20 unassigned nodes were identified and included.
5. Steps 2–4 were repeated until the network reached a stage of edge explosion and massive community merging.

Previously, final visualization thresholds for sequence similarity networks were often chosen based on known functional annotations. However, such information is lacking for most ORs. We found that under different thresholds, the modularity of community structures defined by the 'trunk–branch' strategy showed a sharp inflection point when compared with theoretical modularity derived from the Louvain algorithm. Here, we propose a quantifiable approach for determining the final community

display threshold by identifying the inflection point. Validation against known data confirmed that the resulting community partitions are highly reliable.

Sequence similarity network

First, we performed pairwise sequence alignment of all ORs using MMseqs2 (***Steinegger and Söding, 2017***) to obtain sequence similarities and e-values. To improve network readability, we applied pre-clustering with CD-HIT (***Li and Godzik, 2006***; ***Fu et al., 2012***) v4.8.1 ('-c 0.6 n 4') to reduce redundant edges between highly similar nodes, resulting in a total of 4947 nodes. The edge weight between nodes was defined as the average e-value of all OR pairs they represent. We then constructed the sequence similarity network based on the 'trunk–branch' strategy (see 'Trunk–branch strategy'). An e-value threshold of 1e-25 was ultimately selected as the cutoff for defining the final sequence similarity network and OR subfamilies. At this threshold, 37 OR communities were identified, comprising 8,457 ORs. The network was visualized using the organic layout algorithm in Cytoscape (***Shannon et al., 2003***).

Structure similarity network

We performed structural modeling of all ORs using AlphaFold2 (***Jumper et al., 2021***). Pairwise structural comparisons were then conducted using Dali (***Holm, 2022***) to obtain Z-scores. Following an approach similar to that used in SSN, we applied a greedy clustering algorithm to pre-cluster structurally similar ORs at a Z-score threshold of 50, resulting in 3854 nodes. The edge weight between nodes was defined as the average Z-score of all OR pairs they represent. We then constructed the structure similarity network using the 'trunk–branch' strategy. A Z-score threshold of 42.9 was selected as the final cutoff for defining the StSN and OR structural subfamilies. At this threshold, 27 OR structural communities were identified, comprising 7750 ORs. The network was visualized using the organic layout algorithm in Cytoscape (***Shannon et al., 2003***).

Phylogenetic tree and selection pressure analysis

To infer the amino acid phylogeny of ORs within StrC17, we performed multiple sequence alignment using MAFFT (***Katoh and Standley, 2013***) v7.505 with the parameters '--globalpair `--maxiterate` 1000'. The alignments were trimmed using trimAl (***Capella-Gutiérrez et al., 2009***) v1.4.rev22 with the parameter '-automated1'. A maximum likelihood tree was then constructed using IQ-TREE (***Minh et al., 2013***; ***Nguyen et al., 2015***) v 2.2.0 with the parameters '-m MFP -B 1000 `--bnni`'. To estimate dN/dS ratios, we used the codeml program from the PAML package (RRID:SCR_014932; ***Yang, 1997***; ***Yang, 2007***). The codeml analysis employed the branch model with the control file parameters set to 'model = 2', 'NSsites = 0'. We used a custom script to extract log-likelihood values (lnL) from the output files and perform likelihood ratio tests (LRTs).

Evaluation of OR binding pocket properties

We used fpocket to identify OR binding pockets and assess their properties, with parameters set to '-m 3.1 -i 10'. To locate the centroid region of the binding pocket, we aligned each OR structure to MhOR5 using USalign (***Zhang et al., 2022***) v 20220924. Among all candidate pockets identified by fpocket (***Le Guilloux et al., 2009***) 4.0, the one closest to the centroid was designated as the binding pocket. Due to the stochastic nature of OR modeling, some ORs exhibited open binding pockets, making it difficult to define cavity-related parameters. To ensure consistency in the quantitative analysis, we only included ORs with closed binding pockets. Specifically, we retained only binding pockets with Number of Alpha Spheres <140 and Volume <800, excluding any that were open to the exterior.

Molecular dynamics simulation of Orco β-sheet structural function

We used GROMACS (***Abraham et al., 2015***) to simulate the interaction between the EL2 regions of OR and Orco with VOCs. The proteins used were the cryo-EM structures of ApOrco and ApOR5 (***Wang et al., 2024***; PDB: 8z9z), and the selected VOC was geranyl acetate (CID: 1549026). The membrane system was rapidly constructed using the QwikMD plugin in VMD (***Humphrey et al., 1996***). After default minimization and annealing steps, an optimized membrane-protein system was obtained. The Caver (***Jurcik et al., 2018***) tool was used to identify the pocket tunnels in the cryo-EM structures of ApisOR5 and ApisOrco. Surface residues were identified using FindSurfaceResidues

(https://pymolwiki.org/index.php/FindSurfaceResidues), helping define the residues surrounding the binding pocket. Geranyl acetate was positioned above the pocket, and 5 ns MD simulations were performed to evaluate interactions.

A binding state was considered potential if the following criteria were met:

- The minimum distance between geranyl acetate atoms and pocket residue atoms was <12 Å (cutoff).
- The spatial distribution of the ligand satisfied the following exclusion conditions:
- Not located on the lateral side of the protein
- Not embedded in the phospholipid bilayer
- Not sterically blocked by the EL2 of ApisOrco

Each simulation was repeated 60 times for both ApOrco and ApOR5.

It should be noted that the MD simulations were designed to examine the early trajectories of VOCs near the extracellular entrance of the binding pocket, rather than to assess the conformational stability of the receptor complex. A duration of 5 ns is sufficient for small molecules to respond to local steric constraints.

Cell lines

HEK293T cells (RRID:CVCL_0063) were purchased from Wuhan Procell Biotechnology, China. Cell line identity was authenticated by STR profiling performed by Wuhan Procell. Cells were tested for mycoplasma contamination every three months and were negative. Cells were maintained in DMEM (BasalMedia, China) supplemented with 10% fetal bovine serum (ExPhos, China) and 1% penicillin–streptomycin (Sangon Biotech, China) at 37 °C in a humidified atmosphere containing 5% CO_2.

Heterologous expression and calcium imaging

pDmelOR-mRho.V5.mER.hOr47a (RRID:Addgene_126478), a high-copy number bidirectional expression vector by inserting *D. melanogaster* Or47a (DmOr47a) and Orco in the pCMV-BI vector (RRID:Addgene_126475) backbone (***Miazzi et al., 2019***), was a gift from Bill Hansson (Addgene plasmid # 126478), then the V5 tag was replaced with the Flag tag, named pDmelOR-OR47a. The EL2 of OR47a (GHAEPELPFPCLFPWNIHI) in pDmelOR-OR47a was replaced by the EL2 of Orco (GDSVKMVVDHET NSSIPVEIPRLPIKSFYPWNASHG) to construct the new plasmid of pDmelOR-el2.OR47a. The sequences were confirmed by Sanger sequencing (Sangon Biotech, China).

At 70–80% confluency, HEK293T cells were transfected with plasmids using LipoFiter transfection reagent (Hanbio, China). At 6–8 hr after transfection, cells were seeded at a density of 6×10^5 cells per well into 35 mm glass-bottom culture dishes (NEST, China). Cells were harvested for calcium imaging approximately 36 hr post-transfection. The culture medium was removed, and cells were rinsed with HBSS solution (BasalMedia, China). Subsequently, cells were incubated with Fluo-8 AM (5 μM, AAT Bioquest, USA) diluted in HBSS + 0.04% pluronic acid (F127, Sangon Biotech, China) in a CO_2 cell culture incubator for 30 min at 37 °C. Following three washes with HBSS solution, the Ca^{2+}-dependent fluorescence was recorded with a fluorescence microscope (DeltaVision Ultra, GE) with the emission wavelength set at 525 nm and the excitation wavelength at 490 nm. Experiments lasted 5 min with a sampling interval of 5 s. The stimuli—PA (100 μM; Aladdin, China), an OR47a agonist, and VUAA1 (100 μM; Sigma-Aldrich, USA), an Orco agonist—were gently added to the culture dish at 50 s and 180 s, respectively, after the start of the imaging protocol. The recorded signals were analyzed by Fiji (RRID:SCR_002285) ImageJ v.1.54 (RRID:SCR_003070) (***Schindelin et al., 2012***). Levels of Ca^{2+} were expressed as ΔF/F0 ratios. The average fluorescence intensity of the cells during the first 30 seconds was recorded as the baseline (F0). Each assay was conducted in triplicate to ensure reproducibility. For calcium imaging experiments, data acquisition and initial ΔF/F0 ratio calculations were performed by researchers who were blinded to the treatment groups during the analysis phase. Sample sizes for calcium imaging (n=30) were determined based on common standards in the field and prior pilot experiments to ensure sufficient statistical power for detecting changes in fluorescence intensity.

Virtual docking data of ORs and VOCs

We constructed an insect-relevant odorant molecule library from four sources: pheromones (***Li et al., 2023***), food odors (***Chi et al., 2025***), fragrances (thegoodscentscompany.com), and theoretical odors (***Mayhew et al., 2022***). This approach ensured ecological relevance and chemical diversity. After

deduplication, a total of 14,499 VOCs were retained for downstream analysis. We used Dragon (v.5.4) (**Haddad et al., 2008**) to calculate 32 representative molecular descriptors for each VOC, and the ChemmineR (**Cao et al., 2008**) R package to identify functional groups. VOCs were assigned functional group labels based on group priority rules. Dimensionality reduction of the 32 descriptors was performed using t-SNE, and the resulting two-dimensional odor space was visualized using Matplotlib (**Hunter, 2007**). All VOC SMILES were converted to db2 format using the online resource tldr.docking.org. OR structures were aligned to the cryo-EM structure of MhOR5 (PDB:7LIC) to estimate the approximate position of the binding pocket. Virtual docking was then performed using DOCK 3.8 (RRID:SCR_000128; **Coleman et al., 2013**), with batch processing handled via custom scripts ' prepare_dock_file.sh' and 'run_dock_paraFly.sh'. In total, we obtained 130 million docking records between ORs and VOCs.

Hit rate curve prediction and maximum hit estimation

We first compiled published functional datasets containing explicit OR–VOC response relationships. These datasets included ORs from *D. melanogaster*, *Anopheles gambiae*, *Helicoverpa armigera*, *Spodoptera littoralis,* and *L. migratoria*. For each dataset, OR–VOC combinations were assigned response or non-response labels according to the original experimental results. Each OR–VOC pair was then docked using the same structural modeling, docking, and scoring workflow used in the global OR–VOC analysis. Pairs were grouped into docking-score intervals, and the hit rate for each interval was defined as the proportion of experimentally positive pairs among all pairs in that interval. Each score interval was sampled three times. To determine the docking score range at which VOCs are likely to exhibit binding activity in electrophysiological assays, we applied Bayesian statistical modeling using prior probability distributions to define the top, midpoint, random hit rate, and slope of the hit rate curve. We assumed that the relationship between docking score (energy, ei) and hit rate follows a dose-response-like curve (**Figure 5—figure supplement 2A**; **Lyu et al., 2019**).

$$HitRate\left(e_i\right) = \frac{top - bottom}{1 + e^{-S\cdot\left(e_i - dock_{50}\right)}} + bottom$$

This function is defined by four parameters: (1) top: the maximum hit rate; (2) dock50: the docking score at which the hit rate is half of the top value (in kcal mol^{-1}); (3) S: the slope of the curve, calculated as S=slope × 4/top, where slope is the percentage change in hit rate at dock50 per kcal mol^{-1}; (4) bottom: the minimum hit rate, fixed at 0. To define prior distributions, we analyzed OR hit rates for *D. melanogaster* (**Hallem and Carlson, 2006**), *A. gambiae* (**Carey et al., 2010**), and *S. littorali* (**de Fouchier et al., 2017**) across eight docking energy intervals and used this empirical data to derive posterior distributions of the model parameters. We specified independent Bayesian priors for each parameter: P(top)~beta (α=20, β=80), P(dock50)~normal (μ = –10, σ=15), and P(S)~normal (μ=–0.6, σ=0.1; **Figure 5—figure supplement 2B to D**). Sampling from the prior distributions was performed using Hamiltonian Monte Carlo implemented in Stan, with parameters: 'num_samples = 100000', 'num_chains = 4', 'num_warmup = 50000', and 'max_depth = 12'.

Docking-profile similarity network construction and community labeling

We first calculated the Pearson correlation coefficients (PCCs) for docking energies between all OR pairs across the VOC library. ORs were then ranked by the number of VOCs meeting the docking-score threshold, and a greedy clustering algorithm was applied to pre-cluster ORs with high docking-profile similarity (PCCs ≥0.6), resulting in 6299 nodes. The edge weight between nodes was defined as the average PCC of all OR pairs. The docking-profile similarity network was then constructed using the 'trunk–branch' strategy. Due to the complexity of OR functional divergence, our strategy yielded a modularity value that deviated by 0.1 from the theoretical Louvain modularity. To ensure accurate community assignment, we adopted the Louvain algorithm for final community detection. A PCC threshold of 0.48 was selected as the cutoff for defining the final docking-profile similarity network and OR docking-profile communities. The network was visualized using the organic layout algorithm in Cytoscape (**Shannon et al., 2003**).

Then, we assigned a VOC docking breadth label to each docking-profile community. First, we counted the number of VOCs with each functional group that were with docking scores below −8 kcal/mol in each docking-profile community. These counts were then compared against the total number

of ORs recognizing those VOCs across all ORs, using Bonferroni-corrected post hoc tests. If a community showed a significantly higher or lower recognition count, it was considered to have a higher or lower docking-derived VOC count for that functional group. The results were normalized by the average VOC count to derive a docking-profile breadth score for each dPC. *D. melanogaster* ORs were mapped to dPCs to evaluate the consistency between dPC labels and experimental response data. For each VOC functional group, ORs were assigned positive or negative labels according to their dPC labels, and these labels were compared with response counts from the *Drosophila* functional matrix to generate ROC curves and AUC values.

Acknowledgements

We thank Yuan Huang from Shaanxi Normal University and Chenzhu Wang from the Institute of Zoology, Chinese Academy of Sciences, for their contributions to the early planning of the project. We thank Wenyu Zhang from Northwestern Polytechnical University and Yang Liu from the Chinese Academy of Agricultural Sciences for their suggestions on improving the conceptual framework during the mid-stage of the project. We also thank Hongbo Jiang and Fu Cao from Southwest University for their guidance on the experiments. We sincerely thank the reviewers for their constructive and thoughtful comments, which helped us improve the clarity, rigor, and presentation of this manuscript. We acknowledge the financial support for this work from: National Natural Science Foundation of China grant nos. 32270525 (HL) National Natural Science Foundation of China grant nos. 32470445 (GL) Innovation Foundation for Doctor Dissertation of Northwestern Polytechnical University grant no. CX2024082 (TZ) Natural Science Basic Research Program of Shaanxi grant no. 2025JC-QYCX-026 (GL).

Additional information

Funding

Funder	Grant reference number	Author
National Natural Science Foundation of China	32270525	Huimeng Lu
National Natural Science Foundation of China	32470445	Gang Li
Innovation Foundation for Doctor Dissertation of Northwestern Polytechnical University	CX2024082	Tianmin Zhang
Natural Science Basic Research Program of Shaanxi Province	2025JC-QYCX-026	Gang Li

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Author contributions

Tianmin Zhang, Conceptualization, Resources, Data curation, Formal analysis, Funding acquisition, Validation, Investigation, Visualization, Methodology, Writing – original draft, Project administration, Writing – review and editing; Xuanxiao Yang, Data curation, Investigation, Visualization, Methodology; Yufang Fu, Data curation, Validation, Methodology; Wei Xue, Data curation, Investigation; Yifeng Zhang, Suyang Duan, Yangming Yin, Yi Guo, Chenxi Gao, Investigation; Yang Liu, Chang Xu, Supervision, Methodology; Gang Li, Supervision, Funding acquisition; Huimeng Lu, Conceptualization, Supervision, Funding acquisition, Investigation, Writing – original draft, Project administration

Author ORCIDs

Tianmin Zhang https://orcid.org/0009-0003-7172-847X
Yangming Yin https://orcid.org/0009-0004-2185-2656
Yang Liu https://orcid.org/0000-0002-9258-7237

Gang Li https://orcid.org/0000-0001-7400-5486
Chang Xu https://orcid.org/0000-0002-5103-8450
Huimeng Lu https://orcid.org/0000-0002-5334-1402

Peer review material
Version of Record:
Reviewer #1 (Public review): https://doi.org/10.7554/eLife.110877.4.sa1
Reviewer #2 (Public review): https://doi.org/10.7554/eLife.110877.4.sa2
Author response https://doi.org/10.7554/eLife.110877.4.sa3

Reviewed preprint v3
Reviewer #1 (Public review): https://doi.org/10.7554/eLife.110877.3.sa2
Reviewer #2 (Public review): https://doi.org/10.7554/eLife.110877.3.sa1
Author response: https://doi.org/10.7554/eLife.110877.3.sa0
eLife Assessment https://doi.org/10.7554/eLife.110877.3.sa3

Reviewed preprint v2
Reviewer #1 (Public review): https://doi.org/10.7554/eLife.110877.2.sa2
Reviewer #2 (Public review): https://doi.org/10.7554/eLife.110877.2.sa1
Author response: https://doi.org/10.7554/eLife.110877.2.sa0
eLife Assessment https://doi.org/10.7554/eLife.110877.2.sa3

Reviewed preprint v1
Reviewer #1 (Public review): https://doi.org/10.7554/eLife.110877.1.sa1
Reviewer #2 (Public review): https://doi.org/10.7554/eLife.110877.1.sa0
eLife Assessment https://doi.org/10.7554/eLife.110877.1.sa2

Additional files

Supplementary files

Supplementary file 1. Benchmarking of insect odorant receptor annotation. Tables A-C assess the completeness and reliability of the odorant receptor annotation and structural-filtering workflow. Table A compares receptor counts with published repertoires. Table B summarizes correspondence between predicted receptors and curated reference annotations. Table C examines the genomic locations of selected *Bombyx mori* candidates to assess potential exon-merging artifacts in tandem receptor regions.

MDAR checklist

Data availability

The data supporting the findings of this study are available for download from the figshare repository at https://doi.org/10.6084/m9.figshare.30541190. All custom scripts and R code used for bioinformatics filtering and similarity network construction are available at GitHub (https://github.com/zhangtm0309/Code copy archived at ***Zhang, 2026***) and archived on Zenodo (DOI: https://doi.org/10.5281/zenodo.18459276).

The following datasets were generated:

Author(s)	Year	Dataset title	Dataset URL	Database and Identifier
Lu H	2025	Class-wide sequence, structural, and docking-profile diversity of insect odorant receptors and structural specialization of Orco	https://doi.org/10.6084/m9.figshare.30541190.v3	figshare, 10.6084/m9.figshare.30541190

Continued on next page

Continued

Author(s)	Year	Dataset title	Dataset URL	Database and Identifier
Lu H	2026	code and data	https://doi.org/10.6084/m9.figshare.31230736	figshare, 10.6084/m9.figshare.31230736
Zhang T	2026	zhangtm0309/Code: code and Calcium imaging rawdata	https://doi.org/10.5281/zenodo.18459276	Zenodo, 10.5281/zenodo.18459276

The following previously published datasets were used:

Author(s)	Year	Dataset title	Dataset URL	Database and Identifier
The i5k Initiative	2018	Genome assembly Mhra_1.0	https://www.ncbi.nlm.nih.gov/assembly/GCA_003456935.1	NCBI Assembly, GCA_003456935.1
Brand P, Robertson HM, Lin W, Pothula R, Klingeman WE, Jurat-Fuentes JL, Johnson BR	2019	Data from: The origin of the odorant receptor gene family in insects	https://doi.org/10.5061/dryad.p2t8170	Dryad Digital Repository, 10.5061/dryad.p2t8170
BCM-HGSC	2017	Genome assembly Lful_2.0	https://www.ncbi.nlm.nih.gov/assembly/GCA_000376725.2	NCBI Assembly, GCA_000376725.2
Chinese Academy of Agricultural Sciences	2021	Genome assembly CAAS_Pfla_1.0	https://www.ncbi.nlm.nih.gov/assembly/GCA_020796165.1	NCBI Assembly, GCA_020796165.1
WELLCOME SANGER INSTITUTE	2022	Genome assembly ioPlaPenn1.1	https://www.ncbi.nlm.nih.gov/assembly/GCA_933228895.1	NCBI Assembly, GCA_933228895.1
The i5k Initiative	2017	Genome assembly Edan_2.0	https://www.ncbi.nlm.nih.gov/assembly/GCA_000507165.2	NCBI Assembly, GCA_000507165.2
WELLCOME SANGER INSTITUTE	2021	Genome assembly ipNemPict2.1	https://www.ncbi.nlm.nih.gov/assembly/GCA_921293315.1	NCBI Assembly, GCA_921293315.1
WELLCOME SANGER INSTITUTE	2022	Genome assembly ipLeuNigr2.1	https://www.ncbi.nlm.nih.gov/assembly/GCA_934045905.1	NCBI Assembly, GCA_934045905.1
China Agricultural University	2022	Genome assembly CAU_Lmig_1.0	https://www.ncbi.nlm.nih.gov/assembly/GCA_026315105.1	NCBI Assembly, GCA_026315105.1
Texas A&M University	2021	Genome assembly TAMU-RSper-2021	https://www.ncbi.nlm.nih.gov/assembly/GCA_021186555.1	NCBI Assembly, GCA_021186555.1
WELLCOME SANGER INSTITUTE	2022	Genome assembly ihAcaHaem1.1	https://www.ncbi.nlm.nih.gov/assembly/GCA_930367205.1	NCBI Assembly, GCA_930367205.1
University of Georgia	2022	Genome assembly UGA_ACOO_1.1	https://www.ncbi.nlm.nih.gov/assembly/GCA_023614345.1	NCBI Assembly, GCA_023614345.1
The University of Texas at Austin	2019	Genome assembly pea_aphid_22Mar2018_4r6ur_v2	https://www.ncbi.nlm.nih.gov/assembly/GCF_005508785.2	NCBI Assembly, GCF_005508785.2
China Agricultural University	2021	Genome assembly ASM2088223v1	https://www.ncbi.nlm.nih.gov/assembly/GCA_020882235.1	NCBI Assembly, GCA_020882235.1

Continued on next page

Continued

Author(s)	Year	Dataset title	Dataset URL	Database and Identifier
Key Laboratory of Forest Disaster Warning and Control of Yunnan Province	2021	Genome assembly ASM2046615v1	https://www.ncbi.nlm.nih.gov/assembly/GCA_020466155.1	NCBI Assembly, GCA_020466155.1
Zhejiang University	2021	Genome assembly Schi1.0	https://www.ncbi.nlm.nih.gov/assembly/GCA_019022885.1	NCBI Assembly, GCA_019022885.1
USDA-ARS	2022	Genome assembly USDA_Piezodorus_guildinii_Assembly_v1.1	https://www.ncbi.nlm.nih.gov/assembly/GCA_023052935.1	NCBI Assembly, GCA_023052935.1
China Agricutural University	2022	Genome assembly Lbru_1.0	https://www.ncbi.nlm.nih.gov/assembly/GCA_023512825.1	NCBI Assembly, GCA_023512825.1
WELLCOME SANGER INSTITUTE	2021	Genome assembly iyAncNigr1.1	https://www.ncbi.nlm.nih.gov/assembly/GCA_916049575.1	NCBI Assembly, GCA_916049575.1
University of Copenhagen	2022	Genome assembly ASM2471352v1	https://www.ncbi.nlm.nih.gov/assembly/GCA_024713525.1	NCBI Assembly, GCA_024713525.1
Zhejiang university	2021	Genome assembly Gfla_zju_1.0	https://www.ncbi.nlm.nih.gov/assembly/GCA_018340375.1	NCBI Assembly, GCA_018340375.1
WELLCOME SANGER INSTITUTE	2022	Genome assembly iyTipFemo1.1	https://www.ncbi.nlm.nih.gov/assembly/GCA_944319695.1	NCBI Assembly, GCA_944319695.1
WELLCOME SANGER INSTITUTE	2022	Genome assembly iyRhoChlo1.1	https://www.ncbi.nlm.nih.gov/assembly/GCA_944452935.1	NCBI Assembly, GCA_944452935.1
Zhejiang University	2021	Genome assembly ZJU_Trem_1.0	https://www.ncbi.nlm.nih.gov/assembly/GCA_020615435.1	NCBI Assembly, GCA_020615435.1
Nankai University	2021	Genome assembly ASM1890698v1	https://www.ncbi.nlm.nih.gov/assembly/GCA_018906985.1	NCBI Assembly, GCA_018906985.1
WELLCOME SANGER INSTITUTE	2021	Genome assembly iyAnoNige1.1	https://www.ncbi.nlm.nih.gov/assembly/GCA_914767735.1	NCBI Assembly, GCA_914767735.1
Sun Yat-sen University	2022	Genome assembly ASM2257598v1	https://www.ncbi.nlm.nih.gov/assembly/GCA_022575985.1	NCBI Assembly, GCA_022575985.1
WELLCOME SANGER INSTITUTE	2022	Genome assembly iyMacEuro1.2	https://www.ncbi.nlm.nih.gov/assembly/GCA_916610135.2	NCBI Assembly, GCA_916610135.2
Institute of Zoology, Chinese Academy of Sciences	2022	Genome assembly ASM2240512v1	https://www.ncbi.nlm.nih.gov/assembly/GCA_022405125.1	NCBI Assembly, GCA_022405125.1
USDA-ARS	2021	Genome assembly ASM1945775v1	https://www.ncbi.nlm.nih.gov/assembly/GCA_019457755.1	NCBI Assembly, GCA_019457755.1
WELLCOME SANGER INSTITUTE	2022	Genome assembly iySelTumu1.3	https://www.ncbi.nlm.nih.gov/assembly/GCA_913789895.3	NCBI Assembly, GCA_913789895.3

Continued on next page

Continued

Author(s)	Year	Dataset title	Dataset URL	Database and Identifier
University of Lausanne	2021	Genome assembly UNIL_Sinv_3.0	https://www.ncbi.nlm.nih.gov/assembly/GCA_016802725.1	NCBI Assembly, GCA_016802725.1
Institute of Plant Physiology and Ecology, SIBS, CAS	2020	Genome assembly ASM1547642v1	https://www.ncbi.nlm.nih.gov/assembly/GCA_015476425.1	NCBI Assembly, GCA_015476425.1
National Coffee Research Center - CENICAFE	2022	Genome assembly Cenicafe_Pcof_1.1	https://www.ncbi.nlm.nih.gov/assembly/GCA_024137745.1	NCBI Assembly, GCA_024137745.1
United States Department of Agriculture	2021	Genome assembly iyDipSimi1.1	https://www.ncbi.nlm.nih.gov/assembly/GCA_021155765.1	NCBI Assembly, GCA_021155765.1
WELLCOME SANGER INSTITUTE	2022	Genome assembly iyPemLugu1.1	https://www.ncbi.nlm.nih.gov/assembly/GCA_933228935.1	NCBI Assembly, GCA_933228935.1
Institute of Zoology, Chinese Academy of Sciences	2020	Genome assembly ASM1312311v1	https://www.ncbi.nlm.nih.gov/assembly/GCA_013123115.1	NCBI Assembly, GCA_013123115.1
IRBI-UMR 7261 CNRS - University of Tours	2021	Genome assembly CoteBio_Ctyphae	https://www.ncbi.nlm.nih.gov/assembly/GCA_013202065.2	NCBI Assembly, GCA_013202065.2
National Coffee Research Center - CENICAFE	2022	Genome assembly Cenicafe_Pnas_1.1	https://www.ncbi.nlm.nih.gov/assembly/GCA_024137725.1	NCBI Assembly, GCA_024137725.1
WELLCOME SANGER INSTITUTE	2021	Genome assembly iyNomFabr1.1	https://www.ncbi.nlm.nih.gov/assembly/GCA_907165295.1	NCBI Assembly, GCA_907165295.1
WELLCOME SANGER INSTITUTE	2022	Genome assembly iyAndDors1.1	https://www.ncbi.nlm.nih.gov/assembly/GCA_929108735.1	NCBI Assembly, GCA_929108735.1
Zhejiang University	2022	Genome assembly NIG_curated_v1.5.1	https://www.ncbi.nlm.nih.gov/assembly/GCA_024320075.1	NCBI Assembly, GCA_024320075.1
WELLCOME SANGER INSTITUTE	2021	Genome assembly inChrCarn1.1	https://www.ncbi.nlm.nih.gov/assembly/GCA_905475395.1	NCBI Assembly, GCA_905475395.1
Rudjer Boskovic Institute	2022	Genome assembly Tfree1.0	https://www.ncbi.nlm.nih.gov/assembly/GCA_939628115.1	NCBI Assembly, GCA_939628115.1
USDA-ARS Ag100Pest	2023	Genome assembly icTriCast1.1	https://www.ncbi.nlm.nih.gov/assembly/GCA_031307605.1	NCBI Assembly, GCA_031307605.1
WELLCOME SANGER INSTITUTE	2022	Genome assembly icPhiCogn1.1	https://www.ncbi.nlm.nih.gov/assembly/GCA_932526585.1	NCBI Assembly, GCA_932526585.1
WELLCOME SANGER INSTITUTE	2022	Genome assembly icSilAtra2.1	https://www.ncbi.nlm.nih.gov/assembly/GCA_944588485.1	NCBI Assembly, GCA_944588485.1
The University of Queensland	2021	Genome assembly S4-7k-2v3	https://www.ncbi.nlm.nih.gov/assembly/GCA_020654165.1	NCBI Assembly, GCA_020654165.1
WELLCOME SANGER INSTITUTE	2022	Genome assembly icPyrSerr1.2	https://www.ncbi.nlm.nih.gov/assembly/GCA_905333025.2	NCBI Assembly, GCA_905333025.2

Continued on next page

Continued

Author(s)	Year	Dataset title	Dataset URL	Database and Identifier
ROTHAMSTED RESEARCH	2022	Genome assembly PGI_MELIAE_v1	https://www.ncbi.nlm.nih.gov/assembly/GCA_921294245.1	NCBI Assembly, GCA_921294245.1
Northwestern Polytechnical University	2020	Genome assembly ASM1336807v1	https://www.ncbi.nlm.nih.gov/assembly/GCA_013368075.1	NCBI Assembly, GCA_013368075.1
WELLCOME SANGER INSTITUTE	2022	Genome assembly icPolCerv1.1	https://www.ncbi.nlm.nih.gov/assembly/GCA_935413205.1	NCBI Assembly, GCA_935413205.1
WELLCOME SANGER INSTITUTE	2021	Genome assembly icHarAxyr1.1	https://www.ncbi.nlm.nih.gov/assembly/GCA_914767665.1	NCBI Assembly, GCA_914767665.1
ROTHAMSTED RESEARCH	2022	Genome assembly PGI_DIABVI_V1	https://www.ncbi.nlm.nih.gov/assembly/GCA_917563875.1	NCBI Assembly, GCA_917563875.1
WELLCOME SANGER INSTITUTE	2022	Genome assembly icLepMacu1.1	https://www.ncbi.nlm.nih.gov/assembly/GCA_936432065.1	NCBI Assembly, GCA_936432065.1
WELLCOME SANGER INSTITUTE	2022	Genome assembly icPteMadi1.2	https://www.ncbi.nlm.nih.gov/assembly/GCA_911728475.2	NCBI Assembly, GCA_911728475.2
WELLCOME SANGER INSTITUTE	2021	Genome assembly icRhaFulv1.1	https://www.ncbi.nlm.nih.gov/assembly/GCA_905340355.1	NCBI Assembly, GCA_905340355.1
WELLCOME SANGER INSTITUTE	2021	Genome assembly icApoCory1.1	https://www.ncbi.nlm.nih.gov/assembly/GCA_911728435.1	NCBI Assembly, GCA_911728435.1
China Agricultural University	2019	Genome assembly ASM897352v1	https://www.ncbi.nlm.nih.gov/assembly/GCA_008973525.1	NCBI Assembly, GCA_008973525.1
WELLCOME SANGER INSTITUTE	2022	Genome assembly iiLimLuna2.2	https://www.ncbi.nlm.nih.gov/assembly/GCA_917563855.2	NCBI Assembly, GCA_917563855.2
Florida Museum of Natural History	2022	Genome assembly ASM2284056v1	https://www.ncbi.nlm.nih.gov/assembly/GCA_022840565.1	NCBI Assembly, GCA_022840565.1
LOEWE Centre for Translational Biodiversity Genomics	2022	Genome assembly ASM2265174v1	https://www.ncbi.nlm.nih.gov/assembly/GCA_022651745.1	NCBI Assembly, GCA_022651745.1
LOEWE Centre for Translational Biodiversity Genomics	2022	Genome assembly ASM2260657v1	https://www.ncbi.nlm.nih.gov/assembly/GCA_022606575.1	NCBI Assembly, GCA_022606575.1
LOEWE Centre for Translational Biodiversity Genomics	2022	Rhyacophila brunnea genome assembly ASM2249500v1	https://www.ncbi.nlm.nih.gov/assembly/GCA_022495005.1	NCBI Assembly, GCA_022495005.1
Florida Museum of Natural History	2022	Genome assembly ASM2211383v1	https://www.ncbi.nlm.nih.gov/assembly/GCA_022113835.1	NCBI Assembly, GCA_022113835.1
WELLCOME SANGER INSTITUTE	2021	Genome assembly ilZygFili1.2	https://www.ncbi.nlm.nih.gov/assembly/GCA_907165275.2	NCBI Assembly, GCA_907165275.2

Continued on next page

Continued

Author(s)	Year	Dataset title	Dataset URL	Database and Identifier
WELLCOME SANGER INSTITUTE	2021	Genome assembly ilYpsScab1.1	https://www.ncbi.nlm.nih.gov/assembly/GCA_910592155.1	NCBI Assembly, GCA_910592155.1
WELLCOME SANGER INSTITUTE	2022	Genome assembly ilYpoSedl1.1	https://www.ncbi.nlm.nih.gov/assembly/GCA_934045075.1	NCBI Assembly, GCA_934045075.1
WELLCOME SANGER INSTITUTE	2021	Genome assembly ilHedSali1.2	https://www.ncbi.nlm.nih.gov/assembly/GCA_905404275.2	NCBI Assembly, GCA_905404275.2
WELLCOME SANGER INSTITUTE	2021	Genome assembly ilTinTrin1.1	https://www.ncbi.nlm.nih.gov/assembly/GCA_905220615.1	NCBI Assembly, GCA_905220615.1
WELLCOME SANGER INSTITUTE	2021	Genome assembly ilMimTili1.1	https://www.ncbi.nlm.nih.gov/assembly/GCA_905332985.1	NCBI Assembly, GCA_905332985.1
WELLCOME SANGER INSTITUTE	2021	Genome assembly ilSesApif2.1	https://www.ncbi.nlm.nih.gov/assembly/GCA_914767545.1	NCBI Assembly, GCA_914767545.1
Gakushuin University	2020	Genome assembly Sr_UT_JL_v1.0	https://www.ncbi.nlm.nih.gov/assembly/GCA_014132275.1	NCBI Assembly, GCA_014132275.1
WELLCOME SANGER INSTITUTE	2022	Genome assembly ilAcrSuav1.1	https://www.ncbi.nlm.nih.gov/assembly/GCA_943193695.1	NCBI Assembly, GCA_943193695.1
WELLCOME SANGER INSTITUTE	2021	Genome assembly ilMarLuna1.1	https://www.ncbi.nlm.nih.gov/assembly/GCA_923062675.1	NCBI Assembly, GCA_923062675.1
University of Adelaide	2021	Genome assembly PxLV.1	https://www.ncbi.nlm.nih.gov/assembly/GCA_019096205.1	NCBI Assembly, GCA_019096205.1
Wellcome Sanger Institute	2021	Genome assembly ilPieBrab1.1	https://www.ncbi.nlm.nih.gov/assembly/GCA_905147105.1	NCBI Assembly, GCA_905147105.1
Stockholm University	2021	Genome assembly Parnassius_apollo	https://www.ncbi.nlm.nih.gov/assembly/GCA_907164705.1	NCBI Assembly, GCA_907164705.1
WELLCOME SANGER INSTITUTE	2021	Genome assembly ilFurFurc1.1	https://www.ncbi.nlm.nih.gov/assembly/GCA_911728495.1	NCBI Assembly, GCA_911728495.1
WELLCOME SANGER INSTITUTE	2022	Genome assembly ilMegAlbu1.1	https://www.ncbi.nlm.nih.gov/assembly/GCA_936450015.1	NCBI Assembly, GCA_936450015.1
WELLCOME SANGER INSTITUTE	2022	Genome assembly ilAteCent1.3	https://www.ncbi.nlm.nih.gov/assembly/GCA_905333075.3	NCBI Assembly, GCA_905333075.3
University of Florida	2021	Genome assembly ASM2038319v1	https://www.ncbi.nlm.nih.gov/assembly/GCA_020383195.1	NCBI Assembly, GCA_020383195.1
Wellcome Sanger Institute	2021	Genome assembly ilAriAges1.1	https://www.ncbi.nlm.nih.gov/assembly/GCA_905147365.1	NCBI Assembly, GCA_905147365.1
Southwest Forestry University	2021	Genome assembly ASM1992509v1	https://www.ncbi.nlm.nih.gov/assembly/GCA_019925095.1	NCBI Assembly, GCA_019925095.1

Continued on next page

Continued

Author(s)	Year	Dataset title	Dataset URL	Database and Identifier
WELLCOME SANGER INSTITUTE	2022	Genome assembly ilPyrMalv3.2 alternate haplotype	https://www.ncbi.nlm.nih.gov/assembly/GCA_911387725.2	NCBI Assembly, GCA_911387725.2
WELLCOME SANGER INSTITUTE	2021	Genome assembly ilCroElin1.1	https://www.ncbi.nlm.nih.gov/assembly/GCA_907269065.1	NCBI Assembly, GCA_907269065.1
Wellcome Sanger Institute	2022	Genome assembly ilLymMona1.2	https://www.ncbi.nlm.nih.gov/assembly/GCA_905163515.2	NCBI Assembly, GCA_905163515.2
WELLCOME SANGER INSTITUTE	2021	Genome assembly ilHabPyri1.1	https://www.ncbi.nlm.nih.gov/assembly/GCA_907165245.1	NCBI Assembly, GCA_907165245.1
WELLCOME SANGER INSTITUTE	2022	Genome assembly ilCarQuer1.2	https://www.ncbi.nlm.nih.gov/assembly/GCA_910589575.2	NCBI Assembly, GCA_910589575.2
WELLCOME SANGER INSTITUTE	2022	Genome assembly ilCalPalu1.1	https://www.ncbi.nlm.nih.gov/assembly/GCA_927399485.1	NCBI Assembly, GCA_927399485.1
WELLCOME SANGER INSTITUTE	2021	Genome assembly ilZeuPyri1.1	https://www.ncbi.nlm.nih.gov/assembly/GCA_907165235.1	NCBI Assembly, GCA_907165235.1
Beijing Academy of Agriculture and Forestry Sciences	2020	Genome assembly ASM1460749v2	https://www.ncbi.nlm.nih.gov/assembly/GCA_014607495.2	NCBI Assembly, GCA_014607495.2
Graduate School of Agricultural and Life Sciences, The University of Tokyo	2020	Genome assembly Bmori_2016v1.0	https://www.ncbi.nlm.nih.gov/assembly/GCA_014905235.2	NCBI Assembly, GCA_014905235.2
WELLCOME SANGER INSTITUTE	2021	Genome assembly ilBlaAdus2.1	https://www.ncbi.nlm.nih.gov/assembly/GCA_907269095.1	NCBI Assembly, GCA_907269095.1
WELLCOME SANGER INSTITUTE	2021	Genome assembly ilEupTran1.1	https://www.ncbi.nlm.nih.gov/assembly/GCA_914767815.1	NCBI Assembly, GCA_914767815.1
Chinese Academy of Inspection and Quarantine	2021	Genome assembly ASM2028386v1	https://www.ncbi.nlm.nih.gov/assembly/GCA_020283865.1	NCBI Assembly, GCA_020283865.1
WELLCOME SANGER INSTITUTE	2022	Genome assembly idNepFlae1.1	https://www.ncbi.nlm.nih.gov/assembly/GCA_932526605.1	NCBI Assembly, GCA_932526605.1
WELLCOME SANGER INSTITUTE	2021	Genome assembly idCheVulp2.1	https://www.ncbi.nlm.nih.gov/assembly/GCA_916610125.1	NCBI Assembly, GCA_916610125.1
CAM	2020	Genome assembly iHerIll2.2.curated.20191125	https://www.ncbi.nlm.nih.gov/assembly/GCA_905115235.1	NCBI Assembly, GCA_905115235.1
WELLCOME SANGER INSTITUTE	2022	Genome assembly idPheCory1.1	https://www.ncbi.nlm.nih.gov/assembly/GCA_943735915.1	NCBI Assembly, GCA_943735915.1
Brown University	2020	Genome assembly BU_Bcop_v1	https://www.ncbi.nlm.nih.gov/assembly/GCA_014529535.1	NCBI Assembly, GCA_014529535.1

Continued on next page

Continued

Author(s)	Year	Dataset title	Dataset URL	Database and Identifier
WELLCOME SANGER INSTITUTE	2022	Genome assembly idSarRose1.1	https://www.ncbi.nlm.nih.gov/assembly/GCA_930367235.1	NCBI Assembly, GCA_930367235.1
WELLCOME SANGER INSTITUTE	2022	Genome assembly idStoLuna1.1	https://www.ncbi.nlm.nih.gov/assembly/GCA_933228675.1	NCBI Assembly, GCA_933228675.1
WELLCOME SANGER INSTITUTE	2022	Genome assembly idPolAngu1.1	https://www.ncbi.nlm.nih.gov/assembly/GCA_930367215.1	NCBI Assembly, GCA_930367215.1
Institute of Zoology	2022	Genome assembly ASM2308958v1	https://www.ncbi.nlm.nih.gov/assembly/GCA_023089585.1	NCBI Assembly, GCA_023089585.1
Stanford University	2021	Genome assembly ASM1890436v1	https://www.ncbi.nlm.nih.gov/assembly/GCA_018904365.1	NCBI Assembly, GCA_018904365.1
WELLCOME SANGER INSTITUTE	2022	Genome assembly idAnoGambNW_F1_1	https://www.ncbi.nlm.nih.gov/assembly/GCA_943734735.1	NCBI Assembly, GCA_943734735.1
WELLCOME SANGER INSTITUTE	2022	Genome assembly idMyoTess1.1	https://www.ncbi.nlm.nih.gov/assembly/GCA_943737955.1	NCBI Assembly, GCA_943737955.1
WELLCOME SANGER INSTITUTE	2021	Genome assembly idCluTigr1.1	https://www.ncbi.nlm.nih.gov/assembly/GCA_920105625.1	NCBI Assembly, GCA_920105625.1
Hunan Agricultural University	2021	Genome assembly Cory_1.0	https://www.ncbi.nlm.nih.gov/assembly/GCA_020466095.1	NCBI Assembly, GCA_020466095.1
Tianjin Normal University	2021	Genome assembly ASM1839793v1	https://www.ncbi.nlm.nih.gov/assembly/GCA_018397935.1	NCBI Assembly, GCA_018397935.1
WELLCOME SANGER INSTITUTE	2022	Genome assembly idBelPand1.2	https://www.ncbi.nlm.nih.gov/assembly/GCA_916048285.2	NCBI Assembly, GCA_916048285.2
WELLCOME SANGER INSTITUTE	2022	Genome assembly idBomDisc1.1	https://www.ncbi.nlm.nih.gov/assembly/GCA_939192795.1	NCBI Assembly, GCA_939192795.1
WELLCOME SANGER INSTITUTE	2021	Genome assembly idBibMarc1.1	https://www.ncbi.nlm.nih.gov/assembly/GCA_910594885.1	NCBI Assembly, GCA_910594885.1
WELLCOME SANGER INSTITUTE	2022	Genome assembly idMacAtri3.1	https://www.ncbi.nlm.nih.gov/assembly/GCA_933228815.1	NCBI Assembly, GCA_933228815.1

References

Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, Lindahl E. 2015. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1–2**:19–25. DOI: https://doi.org/10.1016/j.softx.2015.06.001

Auer TO, Khallaf MA, Silbering AF, Zappia G, Ellis K, Álvarez-Ocaña R, Arguello JR, Hansson BS, Jefferis GSXE, Caron SJC, Knaden M, Benton R. 2020. Olfactory receptor and circuit evolution promote host specialization. *Nature* **579**:402–408. DOI: https://doi.org/10.1038/s41586-020-2073-7, PMID: 32132713

Bahk S, Jones WD. 2016. Insect odorant receptor trafficking requires calmodulin. *BMC Biology* **14**:83. DOI: https://doi.org/10.1186/s12915-016-0306-x, PMID: 27686128

Bargmann CI. 2006. Comparative chemosensation from receptors to ecology. *Nature* **444**:295–301. DOI: https://doi.org/10.1038/nature05402, PMID: 17108953

Bastin-Héline L, de Fouchier A, Cao S, Koutroumpa F, Caballero-Vidal G, Robakiewicz S, Monsempes C, François MC, Ribeyre T, Maria A, Chertemps T, de Cian A, Walker WB III, Wang G, Jacquin-Joly E, Montagné N.

2019. A novel lineage of candidate pheromone receptors for sex communication in moths. *eLife* **8**:e49826. DOI: https://doi.org/10.7554/eLife.49826, PMID: 31818368

Benton MJ. 2016. The Triassic. *Current Biology* **26**:R1214–R1218. DOI: https://doi.org/10.1016/j.cub.2016.10.060, PMID: 27923124

Benton R, Himmel NJ. 2023. Structural screens identify candidate human homologs of insect chemoreceptors and cryptic *Drosophila* gustatory receptor-like proteins. *eLife* **12**:e85537. DOI: https://doi.org/10.7554/eLife.85537, PMID: 36803935

Bonasio R, Zhang G, Ye C, Mutti NS, Fang X, Qin N, Donahue G, Yang P, Li Q, Li C, Zhang P, Huang Z, Berger SL, Reinberg D, Wang J, Liebig J. 2010. Genomic comparison of the ants *Camponotus floridanus* and *Harpegnathos saltator*. *Science* **329**:1068–1071. DOI: https://doi.org/10.1126/science.1192428, PMID: 20798317

Boyle SM, McInally S, Ray A. 2013. Expanding the olfactory code by in silico decoding of odor-receptor chemical space. *eLife* **2**:e01120. DOI: https://doi.org/10.7554/eLife.01120, PMID: 24137542

Brand P, Ramírez SR. 2017. The evolutionary dynamics of the odorant receptor gene family in corbiculate bees. *Genome Biology and Evolution* **9**:2023–2036. DOI: https://doi.org/10.1093/gbe/evx149, PMID: 28854688

Brand P, Robertson HM, Lin W, Pothula R, Klingeman WE, Jurat-Fuentes JL, Johnson BR. 2018. The origin of the odorant receptor gene family in insects. *eLife* **7**:38340. DOI: https://doi.org/10.7554/eLife.38340

Caballero-Vidal G, Bouysset C, Grunig H, Fiorucci S, Montagné N, Golebiowski J, Jacquin-Joly E. 2020. Machine learning decodes chemical features to identify novel agonists of a moth odorant receptor. *Scientific Reports* **10**:1655. DOI: https://doi.org/10.1038/s41598-020-58564-9, PMID: 32015393

Cao Y, Charisi A, Cheng LC, Jiang T, Girke T. 2008. ChemmineR: a compound mining framework for R. *Bioinformatics* **24**:1733–1734. DOI: https://doi.org/10.1093/bioinformatics/btn307

Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**:1972–1973. DOI: https://doi.org/10.1093/bioinformatics/btp348, PMID: 19505945

Carey AF, Wang G, Su CY, Zwiebel LJ, Carlson JR. 2010. Odorant reception in the malaria mosquito Anopheles gambiae. *Nature* **464**:66–71. DOI: https://doi.org/10.1038/nature08834, PMID: 20130575

Chang H, Unni AP, Tom MT, Cao Q, Liu Y, Wang G, Llorca LC, Brase S, Bucks S, Weniger K, Bisch-Knaden S, Hansson BS, Knaden M. 2023. Odorant detection in a locust exhibits unusually low redundancy. *Current Biology* **33**:5427–5438. DOI: https://doi.org/10.1016/j.cub.2023.11.017, PMID: 38070506

Chepurwar S, Gupta A, Haddad R, Gupta N. 2019. Sequence-based prediction of olfactory receptor responses. *Chemical Senses* **44**:693–703. DOI: https://doi.org/10.1093/chemse/bjz059, PMID: 31665762

Chesters D. 2017. Construction of a species-level tree of life for the insects and utility in taxonomic profiling. *Systematic Biology* **66**:426–439. DOI: https://doi.org/10.1093/sysbio/syw099, PMID: 27798407

Chi H, Wan J, Melin AD, DeCasien AR, Wang S, Zhang Y, Cui Y, Guo X, Zhao L, Williamson J, Zhang T, Li Q, Zhan Y, Li N, Guo J, Xu Z, Hou W, Cao Y, Yuan J, Zheng J, et al. 2025. Genomic and phenotypic evidence support visual and olfactory shifts in primate evolution. *Nature Ecology & Evolution* **9**:721–733. DOI: https://doi.org/10.1038/s41559-025-02651-5, PMID: 40021902

Close RA, Benson RBJ, Saupe EE, Clapham ME, Butler RJ. 2020. The spatial structure of Phanerozoic marine animal diversity. *Science* **368**:420–424. DOI: https://doi.org/10.1126/science.aay8309, PMID: 32327597

Clyne PJ, Warr CG, Freeman MR, Lessing D, Kim J, Carlson JR. 1999. A novel family of divergent seven-transmembrane proteins: candidate odorant receptors in *Drosophila*. *Neuron* **22**:327–338. DOI: https://doi.org/10.1016/s0896-6273(00)81093-4, PMID: 10069338

Coleman RG, Carchia M, Sterling T, Irwin JJ, Shoichet BK. 2013. Ligand pose and orientational sampling in molecular docking. *PLOS ONE* **8**:e75992. DOI: https://doi.org/10.1371/journal.pone.0075992

Comte A, Lalis M, Brajon L, Moracci R, Montagné N, Topin J, Joly EJ, Fiorucci S. 2025. Accelerating ligand discovery for insect odorant receptors. *International Journal of Biological Sciences* **21**:2101–2117. DOI: https://doi.org/10.7150/ijbs.105648

Crooks GE, Hon G, Chandonia JM, Brenner SE. 2004. WebLogo: a sequence logo generator: Figure 1. *Genome Research* **14**:1188–1190. DOI: https://doi.org/10.1101/gr.849004

de Fouchier A, Walker WB III, Montagné N, Steiner C, Binyameen M, Schlyter F, Chertemps T, Maria A, François MC, Monsempes C, Anderson P, Hansson BS, Larsson MC, Jacquin-Joly E. 2017. Functional evolution of Lepidoptera olfactory receptors revealed by deorphanization of a moth repertoire. *Nature Communications* **8**:15709. DOI: https://doi.org/10.1038/ncomms15709, PMID: 28580965

Dekel A, Pitts RJ, Yakir E, Bohbot JD. 2016. Evolutionarily conserved odorant receptor function questions ecological context of octenol role in mosquitoes. *Scientific Reports* **6**:37330. DOI: https://doi.org/10.1038/srep37330, PMID: 27849027

Del Mármol J, Yedlin MA, Ruta V. 2021. The structural basis of odorant recognition in insect olfactory receptors. *Nature* **597**:126–131. DOI: https://doi.org/10.1038/s41586-021-03794-8, PMID: 34349260

Depetris-Chauvin A, Galagovsky D, Keesey IW, Hansson BS, Sachse S, Knaden M. 2023. Evolution at multiple processing levels underlies odor-guided behavior in the genus *Drosophila*. *Current Biology* **33**:4771–4785. DOI: https://doi.org/10.1016/j.cub.2023.09.039, PMID: 37804828

Dijkstra K-DB, Monaghan MT, Pauls SU. 2014. Freshwater biodiversity and aquatic insect diversification. *Annual Review of Entomology* **59**:143–163. DOI: https://doi.org/10.1146/annurev-ento-011613-161958, PMID: 24160433

Dweck HKM, Carlson JR. 2023. Diverse mechanisms of taste coding in *Drosophila*. *Science Advances* **9**:eadj7032. DOI: https://doi.org/10.1126/sciadv.adj7032, PMID: 37976361

Engsontia P, Sanderson AP, Cobb M, Walden KKO, Robertson HM, Brown S. 2008. The red flour beetle's large nose: an expanded odorant receptor gene family in *Tribolium castaneum*. *Insect Biochemistry and Molecular Biology* **38**:387–397. DOI: https://doi.org/10.1016/j.ibmb.2007.10.005, PMID: 18342245

Engsontia P, Sangket U, Robertson HM, Satasook C. 2015. Diversification of the ant odorant receptor gene family and positive selection on candidate cuticular hydrocarbon receptors. *BMC Research Notes* **8**:380. DOI: https://doi.org/10.1186/s13104-015-1371-x, PMID: 26306879

Frank HM, Walujkar S, Walsh RM, Laursen WJ, Theobald DL, Garrity PA, Gaudet R. 2024. Structural basis of ligand specificity and channel activation in an insect gustatory receptor. *Cell Reports* **43**:114035. DOI: https://doi.org/10.1016/j.celrep.2024.114035, PMID: 38573859

Frunze O, Lee D, Lee S, Kwon HW. 2024. A single mutation in the mosquito (Aedes aegypti) olfactory receptor 8 causes loss of function to 1-octen-3-ol. *Insect Biochemistry and Molecular Biology* **167**:104069. DOI: https://doi.org/10.1016/j.ibmb.2023.104069

Fu L, Niu B, Zhu Z, Wu S, Li W. 2012. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**:3150–3152. DOI: https://doi.org/10.1093/bioinformatics/bts565

Grauvogel-Stamm L, Ash SR. 2005. Recovery of the Triassic land flora from the end-Permian life crisis. *Comptes Rendus Palevol* **4**:593–608. DOI: https://doi.org/10.1016/j.crpv.2005.07.002

Guo M, Du L, Chen Q, Feng Y, Zhang J, Zhang X, Tian K, Cao S, Huang T, Jacquin-Joly E, Wang G, Liu Y. 2021. Odorant receptors for detecting flowering plant cues are functionally conserved across moths and butterflies. *Molecular Biology and Evolution* **38**:1413–1427. DOI: https://doi.org/10.1093/molbev/msaa300

Haddad R, Khan R, Takahashi YK, Mori K, Harel D, Sobel N. 2008. A metric for odorant comparison. *Nature Methods* **5**:425–429. DOI: https://doi.org/10.1038/nmeth.1197, PMID: 18376403

Hall DR, Beevor PS, Cork A, Nesbitt BF, Vale GA. 1984. 1-Octen-3-ol. *International Journal of Tropical Insect Science* **5**:335–339. DOI: https://doi.org/10.1017/S1742758400008626

Hallem EA, Carlson JR. 2006. Coding of odors by a receptor repertoire. *Cell* **125**:143–160. DOI: https://doi.org/10.1016/j.cell.2006.01.050

Hansson BS, Stensmyr MC. 2011. Evolution of insect olfaction. *Neuron* **72**:698–711. DOI: https://doi.org/10.1016/j.neuron.2011.11.003, PMID: 22153368

Himmel NJ, Moi D, Benton R. 2023. Remote homolog detection places insect chemoreceptors in a cryptic protein superfamily spanning the tree of life. *Current Biology* **33**:5023–5033. DOI: https://doi.org/10.1016/j.cub.2023.10.008, PMID: 37913770

Holm L. 2022. Dali server: structural unification of protein families. *Nucleic Acids Research* **50**:W210–W215. DOI: https://doi.org/10.1093/nar/gkac387

Hughes GM, Boston ESM, Finarelli JA, Murphy WJ, Higgins DG, Teeling EC. 2018. The birth and death of olfactory receptor gene families in mammalian Niche adaptation. *Molecular Biology and Evolution* **35**:1390–1406. DOI: https://doi.org/10.1093/molbev/msy028, PMID: 29562344

Humphrey W, Dalke A, Schulten K. 1996. VMD: visual molecular dynamics. *Journal of Molecular Graphics* **14**:33–38. DOI: https://doi.org/10.1016/0263-7855(96)00018-5, PMID: 8744570

Hunter JD. 2007. Matplotlib: a 2D graphics environment. *Computing in Science & Engineering* **9**:90–95. DOI: https://doi.org/10.1109/MCSE.2007.55

Jain K, Lavista-Llanos S, Grabe V, Hansson BS, Wicher D. 2021. Calmodulin regulates the olfactory performance in *Drosophila melanogaster*. *Scientific Reports* **11**:3747. DOI: https://doi.org/10.1038/s41598-021-83296-9, PMID: 33580172

Jiang X, Dimitriou E, Grabe V, Sun R, Chang H, Zhang Y, Gershenzon J, Rybak J, Hansson BS, Sachse S. 2024. Ring-shaped odor coding in the antennal lobe of migratory locusts. *Cell* **187**:3973–3991. DOI: https://doi.org/10.1016/j.cell.2024.05.036, PMID: 38897195

Jones WD, Nguyen TAT, Kloss B, Lee KJ, Vosshall LB. 2005. Functional conservation of an insect odorant receptor gene across 250 million years of evolution. *Current Biology* **15**:R119–R121. DOI: https://doi.org/10.1016/j.cub.2005.02.007, PMID: 15723778

Jones PL, Pask GM, Rinker DC, Zwiebel LJ. 2011. Functional agonism of insect odorant receptor ion channels. *PNAS* **108**:8821–8825. DOI: https://doi.org/10.1073/pnas.1102425108, PMID: 21555561

Jongepier E, Séguret A, Labutin A, Feldmeyer B, Gstöttl C, Foitzik S, Heinze J, Bornberg-Bauer E. 2022. Convergent loss of chemoreceptors across independent origins of slave-making in ants. *Molecular Biology and Evolution* **39**:msab305. DOI: https://doi.org/10.1093/molbev/msab305, PMID: 34668533

Jouault C, Nel A, Perrichot V, Legendre F, Condamine FL. 2022. Multiple drivers and lineage-specific insect extinctions during the Permo-Triassic. *Nature Communications* **13**:7512. DOI: https://doi.org/10.1038/s41467-022-35284-4, PMID: 36473862

Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, Bridgland A, Meyer C, Kohl SAA, Ballard AJ, Cowie A, Romera-Paredes B, Nikolov S, Jain R, Adler J, Back T, et al. 2021. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**:583–589. DOI: https://doi.org/10.1038/s41586-021-03819-2

Jurcik A, Bednar D, Byska J, Marques SM, Furmanova K, Daniel L, Kokkonen P, Brezovsky J, Strnad O, Stourac J, Pavelka A, Manak M, Damborsky J, Kozlikova B. 2018. CAVER Analyst 2.0: analysis and visualization of channels and tunnels in protein structures and molecular dynamics trajectories. *Bioinformatics* **34**:3586–3588. DOI: https://doi.org/10.1093/bioinformatics/bty386, PMID: 29741570

Karpe SD, Tiwari V, Ramanathan S. 2021. InsectOR-Webserver for sensitive identification of insect olfactory receptor genes from non-model genomes. *PLOS ONE* **16**:e0245324. DOI: https://doi.org/10.1371/journal.pone.0245324, PMID: 33465132

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* **30**:772–780. DOI: https://doi.org/10.1093/molbev/mst010, PMID: 23329690

Kaupp UB. 2010. Olfactory signalling in vertebrates and insects: differences and commonalities. *Nature Reviews. Neuroscience* **11**:188–200. DOI: https://doi.org/10.1038/nrn2789, PMID: 20145624

Kline DL, Allan SA, Bernier UR, Welch CH. 2007. Evaluation of the enantiomers of 1-octen-3-ol and 1-octyn-3-ol as attractants for mosquitoes associated with a freshwater swamp in Florida, U.S.A. *Medical and Veterinary Entomology* **21**:323–331. DOI: https://doi.org/10.1111/j.1365-2915.2007.00697.x, PMID: 18092970

Kreher SA, Mathew D, Kim J, Carlson JR. 2008. Translation of sensory input into behavioral output via an olfactory system. *Neuron* **59**:110–124. DOI: https://doi.org/10.1016/j.neuron.2008.06.010, PMID: 18614033

Kumar S, Suleski M, Craig JM, Kasprowicz AE, Sanderford M, Li M, Stecher G, Hedges SB. 2022. TimeTree 5: an expanded resource for species divergence times. *Molecular Biology and Evolution* **39**:msac174. DOI: https://doi.org/10.1093/molbev/msac174, PMID: 35932227

Larsson MC, Domingos AI, Jones WD, Chiappe ME, Amrein H, Vosshall LB. 2004. Or83b encodes a broadly expressed odorant receptor essential for *Drosophila olfaction*. *Neuron* **43**:703–714. DOI: https://doi.org/10.1016/j.neuron.2004.08.019, PMID: 15339651

Legan AW, Jernigan CM, Miller SE, Fuchs MF, Sheehan MJ. 2021. Expansion and accelerated evolution of 9-Exon odorant receptors in polistes paper wasps. *Molecular Biology and Evolution* **38**:3832–3846. DOI: https://doi.org/10.1093/molbev/msab023, PMID: 34151983

Le Guilloux V, Schmidtke P, Tuffery P. 2009. Fpocket: an open source platform for ligand pocket detection. *BMC Bioinformatics* **10**:168. DOI: https://doi.org/10.1186/1471-2105-10-168, PMID: 19486540

Letunic I, Bork P. 2024. Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. *Nucleic Acids Research* **52**:W78–W82. DOI: https://doi.org/10.1093/nar/gkae268, PMID: 38613393

Li W, Godzik A. 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**:1658–1659. DOI: https://doi.org/10.1093/bioinformatics/btl158, PMID: 16731699

Li Q, Zhang YF, Zhang TM, Wan JH, Zhang YD, Yang H, Huang Y, Xu C, Li G, Lu HM. 2023. iORbase: a database for the prediction of the structures and functions of insect olfactory receptors. *Insect Science* **30**:1245–1254. DOI: https://doi.org/10.1111/1744-7917.13162, PMID: 36519267

Liu H, Chen C, Lv M, Liu N, Hu Y, Zhang H, Enbody ED, Gao Z, Andersson L, Wang W. 2021. A chromosome-level assembly of blunt snout bream (*Megalobrama amblycephala*) genome reveals an expansion of olfactory receptor genes in freshwater fish. *Molecular Biology and Evolution* **38**:4238–4251. DOI: https://doi.org/10.1093/molbev/msab152, PMID: 34003267

Liu WB, Li HM, Wang GR, Cao HQ, Wang B. 2023. Conserved Odorant Receptor, EcorOR4, mediates attraction of mated female *Eupeodes corollae* to 1-Octen-3-ol. *Journal of Agricultural and Food Chemistry* **71**:1837–1844. DOI: https://doi.org/10.1021/acs.jafc.2c06132, PMID: 36682010

Lu T, Qiu YT, Wang G, Kwon JY, Rutzler M, Kwon H-W, Pitts RJ, van Loon JJA, Takken W, Carlson JR, Zwiebel LJ. 2007. Odor coding in the maxillary palp of the malaria vector mosquito *Anopheles gambiae*. *Current Biology* **17**:1533–1544. DOI: https://doi.org/10.1016/j.cub.2007.07.062

Lu J, Zhang P, Dal Corso J, Yang M, Wignall PB, Greene SE, Shao L, Lyu D, Hilton J. 2021. Volcanically driven lacustrine ecosystem changes during the Carnian Pluvial Episode (Late Triassic). *PNAS* **118**:2109895118. DOI: https://doi.org/10.1073/pnas.2109895118

Lyu J, Wang S, Balius TE, Singh I, Levit A, Moroz YS, O'Meara MJ, Che T, Algaa E, Tolmachova K, Tolmachev AA, Shoichet BK, Roth BL, Irwin JJ. 2019. Ultra-large library docking for discovering new chemotypes. *Nature* **566**:224–229. DOI: https://doi.org/10.1038/s41586-019-0917-9, PMID: 30728502

Mayhew EJ, Arayata CJ, Gerkin RC, Lee BK, Magill JM, Snyder LL, Little KA, Yu CW, Mainland JD. 2022. Transport features predict if a molecule is odorous. *PNAS* **119**:e2116576119. DOI: https://doi.org/10.1073/pnas.2116576119, PMID: 35377807

McBride CS. 2007. Rapid evolution of smell and taste receptor genes during host specialization in *Drosophila sechellia*. *PNAS* **104**:4996–5001. DOI: https://doi.org/10.1073/pnas.0608424104, PMID: 17360391

McBride CS, Baier F, Omondi AB, Spitzer SA, Lutomiah J, Sang R, Ignell R, Vosshall LB. 2014. Evolution of mosquito preference for humans linked to an odorant receptor. *Nature* **515**:222–227. DOI: https://doi.org/10.1038/nature13964, PMID: 25391959

McKenzie SK, Fetter-Pruneda I, Ruta V, Kronauer DJC. 2016. Transcriptomics and neuroanatomy of the clonal raider ant implicate an expanded clade of odorant receptors in chemical communication. *PNAS* **113**:14091–14096. DOI: https://doi.org/10.1073/pnas.1610800113, PMID: 27911792

McKenzie SK, Kronauer DJC. 2018. The genomic architecture and molecular evolution of ant odorant receptors. *Genome Research* **28**:1757–1765. DOI: https://doi.org/10.1101/gr.237123.118, PMID: 30249741

Miazzi F, Hoyer C, Sachse S, Knaden M, Wicher D, Hansson BS, Lavista-Llanos S. 2019. Optimization of insect odorant receptor trafficking and functional expression via transient transfection in HEK293 cells. *Chemical Senses* **44**:673–682. DOI: https://doi.org/10.1093/chemse/bjz062, PMID: 31504297

Minh BQ, Nguyen MAT, von Haeseler A. 2013. Ultrafast approximation for phylogenetic bootstrap. *Molecular Biology and Evolution* **30**:1188–1195. DOI: https://doi.org/10.1093/molbev/mst024

Misof B, Liu S, Meusemann K, Peters RS, Donath A, Mayer C, Frandsen PB, Ware J, Flouri T, Beutel RG, Niehuis O, Petersen M, Izquierdo-Carrasco F, Wappler T, Rust J, Aberer AJ, Aspöck U, Aspöck H, Bartel D, Blanke A, et al. 2014. Phylogenomics resolves the timing and pattern of insect evolution. *Science* **346**:763–767. DOI: https://doi.org/10.1126/science.1257570

Mitchell RF, Schneider TM, Schwartz AM, Andersson MN, McKenna DD. 2020. The diversity and evolution of odorant receptors in beetles (Coleoptera). *Insect Molecular Biology* **29**:77–91. DOI: https://doi.org/10.1111/imb.12611, PMID: 31381201

Mitchell RF, Moran KM, McKenna DD. 2026. Early hexapod genomes reveal the deep origin of insect gustatory and odorant receptors. *Nature Communications* **17**:6881. DOI: https://doi.org/10.1038/s41467-026-73718-5, PMID: 42204179

Mora C, Tittensor DP, Adl S, Simpson AGB, Worm B. 2011. How many species are there on earth and in the ocean? *PLOS Biology* **9**:e1001127. DOI: https://doi.org/10.1371/journal.pbio.1001127

Nei M, Niimura Y, Nozawa M. 2008. The evolution of animal chemosensory receptor gene repertoires: roles of chance and necessity. *Nature Reviews. Genetics* **9**:951–963. DOI: https://doi.org/10.1038/nrg2480, PMID: 19002141

Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution* **32**:268–274. DOI: https://doi.org/10.1093/molbev/msu300

Niimura Y, Nei M. 2005. Evolutionary dynamics of olfactory receptor genes in fishes and tetrapods. *PNAS* **102**:6039–6044. DOI: https://doi.org/10.1073/pnas.0501922102

Pacalon J, Audic G, Magnat J, Philip M, Golebiowski J, Moreau CJ, Topin J. 2023. Elucidation of the structural basis for ligand binding and translocation in conserved insect odorant receptor co-receptors. *Nature Communications* **14**:8182. DOI: https://doi.org/10.1038/s41467-023-44058-5, PMID: 38081900

Pask GM, Slone JD, Millar JG, Das P, Moreira JA, Zhou X, Bello J, Berger SL, Bonasio R, Desplan C, Reinberg D, Liebig J, Zwiebel LJ, Ray A. 2017. Specialized odorant receptors in social insects that detect cuticular hydrocarbon cues and candidate pheromones. *Nature Communications* **8**:297. DOI: https://doi.org/10.1038/s41467-017-00099-1, PMID: 28819196

Policarpo M, Baldwin MW, Casane D, Salzburger W. 2024. Diversity and evolution of the vertebrate chemoreceptor gene repertoire. *Nature Communications* **15**:1421. DOI: https://doi.org/10.1038/s41467-024-45500-y, PMID: 38360851

Rannala B, Yang Z. 2007. Inferring speciation times under an episodic molecular clock. *Systematic Biology* **56**:453–466. DOI: https://doi.org/10.1080/10635150701420643

Renthal R, Chen LY. 2022. Tunnel connects lipid bilayer to occluded odorant-binding site of insect olfactory receptor. *Biophysical Chemistry* **289**:106862. DOI: https://doi.org/10.1016/j.bpc.2022.106862, PMID: 35933834

Robertson HM, Warr CG, Carlson JR. 2003. Molecular evolution of the insect chemoreceptor gene superfamily in *Drosophila melanogaster*. *PNAS* **100 Suppl 2**:14537–14542. DOI: https://doi.org/10.1073/pnas.2335847100, PMID: 14608037

Robertson HM, Wanner KW. 2006. The chemoreceptor superfamily in the honey bee, *Apis mellifera*: expansion of the odorant, but not gustatory, receptor family. *Genome Research* **16**:1395–1403. DOI: https://doi.org/10.1101/gr.5057506, PMID: 17065611

Robertson HM. 2019. Molecular evolution of the major arthropod chemoreceptor gene families. *Annual Review of Entomology* **64**:227–242. DOI: https://doi.org/10.1146/annurev-ento-020117-043322, PMID: 30312552

Robertson HM, Robertson ECN, Walden KKO, Enders LS, Miller NJ. 2019. The chemoreceptors and odorant binding proteins of the soybean and pea aphids. *Insect Biochemistry and Molecular Biology* **105**:69–78. DOI: https://doi.org/10.1016/j.ibmb.2019.01.005, PMID: 30654011

Sato K, Pellegrino M, Nakagawa T, Nakagawa T, Vosshall LB, Touhara K. 2008. Insect olfactory receptors are heteromeric ligand-gated ion channels. *Nature* **452**:1002–1006. DOI: https://doi.org/10.1038/nature06850, PMID: 18408712

Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B, Tinevez J-Y, White DJ, Hartenstein V, Eliceiri K, Tomancak P, Cardona A. 2012. Fiji: an open-source platform for biological-image analysis. *Nature Methods* **9**:676–682. DOI: https://doi.org/10.1038/nmeth.2019

Schrader L, Pan H, Bollazzi M, Schiøtt M, Larabee FJ, Bi X, Deng Y, Zhang G, Boomsma JJ, Rabeling C. 2021. Relaxed selection underlies genome erosion in socially parasitic ant species. *Nature Communications* **12**:2918. DOI: https://doi.org/10.1038/s41467-021-23178-w

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research* **13**:2498–2504. DOI: https://doi.org/10.1101/gr.1239303

Singh A, Pope NS, López-Uribe MM. 2025. Shifts in bee diet breadths are associated with gene gains and losses and positive selection across olfactory receptors. *G3* **15**:jkaf105. DOI: https://doi.org/10.1093/g3journal/jkaf105, PMID: 40643014

Slater GSC, Birney E. 2005. Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics* **6**:31. DOI: https://doi.org/10.1186/1471-2105-6-31, PMID: 15713233

Slone JD, Pask GM, Ferguson ST, Millar JG, Berger SL, Reinberg D, Liebig J, Ray A, Zwiebel LJ. 2017. Functional characterization of odorant receptors in the ponerine ant, *Harpegnathos saltator PNAS* **114**:8586–8591. DOI: https://doi.org/10.1073/pnas.1704647114, PMID: 28696298

Steinegger M, Söding J. 2017. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nature Biotechnology* **35**:1026–1028. DOI: https://doi.org/10.1038/nbt.3988, PMID: 29035372

Thoma M, Missbach C, Jordan MD, Grosse-Wilde E, Newcomb RD, Hansson BS. 2019. Transcriptome surveys in silverfish suggest a multistep origin of the insect odorant receptor gene family. *Frontiers in Ecology and Evolution* **7**:e00281. DOI: https://doi.org/10.3389/fevo.2019.00281
Tian J, Dewer Y, Hu H, Li F, Yang S, Luo C. 2022. Diversity and molecular evolution of odorant receptor in hemipteran insects. *Insects* **13**:214. DOI: https://doi.org/10.3390/insects13020214
Vajda V, McLoughlin S, Mays C, Frank TD, Fielding CR, Tevyaw A, Lehsten V, Bocking M, Nicoll RS. 2020. End-Permian (252 Mya) deforestation, wildfires and flooding—An ancient biotic crisis with lessons for the present. *Earth and Planetary Science Letters* **529**:115875. DOI: https://doi.org/10.1016/j.epsl.2019.115875
Vosshall LB, Amrein H, Morozov PS, Rzhetsky A, Axel R. 1999. A spatial map of olfactory receptor expression in the *Drosophila* antenna. *Cell* **96**:725–736. DOI: https://doi.org/10.1016/s0092-8674(00)80582-6, PMID: 10089887
Wang Z, Yang P, Chen D, Jiang F, Li Y, Wang X, Kang L. 2015. Identification and functional analysis of olfactory receptor family reveal unusual characteristics of the olfactory system in the migratory locust. *Cellular and Molecular Life Sciences* **72**:4429–4443. DOI: https://doi.org/10.1007/s00018-015-2009-9
Wang Y, Qiu L, Wang B, Guan Z, Dong Z, Zhang J, Cao S, Yang L, Wang B, Gong Z, Zhang L, Ma W, Liu Z, Zhang D, Wang G, Yin P. 2024. Structural basis for odorant recognition of the insect odorant receptor OR-Orco heterocomplex. *Science* **384**:1453–1460. DOI: https://doi.org/10.1126/science.adn6881
Wicher D, Miazzi F. 2021. Functional properties of insect olfactory receptors: ionotropic receptors and odorant receptors. *Cell and Tissue Research* **383**:7–19. DOI: https://doi.org/10.1007/s00441-020-03363-x
Xu P, Zhu F, Buss GK, Leal WS. 2015. 1-Octen-3-ol - the attractant that repels. *F1000Research* **4**:156. DOI: https://doi.org/10.12688/f1000research.6646.1, PMID: 26543554
Xu L, Jiang HB, Yu JL, Pan D, Tao Y, Lei Q, Chen Y, Liu Z, Wang JJ. 2023. Two odorant receptors regulate 1-octen-3-ol induced oviposition behavior in the oriental fruit fly. *Communications Biology* **6**:176. DOI: https://doi.org/10.1038/s42003-023-04551-5
Yang Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. *Computer Applications in the Biosciences* **13**:555–556. DOI: https://doi.org/10.1093/bioinformatics/13.5.555, PMID: 9367129
Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution* **24**:1586–1591. DOI: https://doi.org/10.1093/molbev/msm088, PMID: 17483113
Yin N, Xiao H, Yang A, Wu C, Liu N. 2022. Genome-wide analysis of odorant and gustatory receptors in six *Papilio* butterflies (Lepidoptera: Papilionidae). *Insects* **13**:779. DOI: https://doi.org/10.3390/insects13090779, PMID: 36135480
Zhang C, Shine M, Pyle AM, Zhang Y. 2022. US-align: universal structure alignments of proteins, nucleic acids, and macromolecular complexes. *Nature Methods* **19**:1109–1115. DOI: https://doi.org/10.1038/s41592-022-01585-1
Zhang T, Jing H, Wang J, Zhao L, Liu Y, Rossiter SJ, Lu H, Li G. 2024. Evolution of olfactory receptor superfamily in bats based on high throughput molecular modelling. *Molecular Ecology Resources* **24**:e13958. DOI: https://doi.org/10.1111/1755-0998.13958, PMID: 38567648
Zhang T. 2026. Code. swh:1:rev:fc7e86f4e72b6eaa81e9cf14beb3069437d09c3c. Software Heritage. https://archive.softwareheritage.org/swh:1:dir:1429e076f2b2e64d297af06ff1eb62235a0ffadb;origin=https://github.com/zhangtm0309/Code;visit=swh:1:snp:adb8d299485e71a1e1184846a4339ad33c0bcf30;anchor=swh:1:rev:fc7e86f4e72b6eaa81e9cf14beb3069437d09c3c
Zhao J, Chen AQ, Ryu J, Del Mármol J. 2024. Structural basis of odor sensing by insect heteromeric odorant receptors. *Science* **384**:1460–1467. DOI: https://doi.org/10.1126/science.adn6384, PMID: 38870275
Zhou X, Slone JD, Rokas A, Berger SL, Liebig J, Ray A, Reinberg D, Zwiebel LJ. 2012. Phylogenetic and transcriptomic analysis of chemosensory receptors in a pair of divergent ant species reveals sex-specific signatures of odor coding. *PLOS Genetics* **8**:e1002930. DOI: https://doi.org/10.1371/journal.pgen.1002930, PMID: 22952454
Zhou X, Rokas A, Berger SL, Liebig J, Ray A, Zwiebel LJ. 2015. Chemoreceptor evolution in hymenoptera and its implications for the evolution of eusociality. *Genome Biology and Evolution* **7**:2407–2416. DOI: https://doi.org/10.1093/gbe/evv149, PMID: 26272716

Appendix 1

Benchmarking of OR annotation completeness and structural filtering

To evaluate the completeness and reliability of the OR annotation workflow, we benchmarked the automated annotation and structural-filtering pipeline against curated OR repertoires from three well-studied insect species: *Drosophila melanogaster*, *Anopheles gambiae*, and *Bombyx mori*. These species were selected because their OR repertoires have been extensively annotated and are among the best curated datasets available for comparison. The benchmark was designed to assess three aspects of the workflow: (i) recovery of previously reported ORs by the automated annotation pipeline, (ii) the effect of structural filtering on OR model quality, and (iii) whether candidate ORs in tandem-rich regions could reflect exon-merging artifacts.

For each benchmark species, we recorded the number of ORs recovered before structural filtering, the number retained after structural filtering, and the correspondence between retained candidates and literature-annotated ORs. Candidate ORs were matched to published ORs based on sequence similarity (***Supplementary file 1B***). When discrepancies occurred, the corresponding sequences and predicted structures were manually inspected. For *Bombyx mori* candidates that did not correspond one-to-one to literature-annotated ORs, we further examined genomic locations to evaluate whether they may have resulted from exon merging among adjacent tandem ORs (***Supplementary file 1C***). Candidates located on different chromosomes or outside the same tandem region as their closest literature ORs were not considered likely exon-merging artifacts.

For *D. melanogaster*, the accepted repertoire contains 60 OR genes and 65 protein products, including splice variants. At the gene level, the automated workflow recovered most benchmark ORs before structural filtering, but Or47b and Or98b were not annotated. After structural filtering, fragmented or structurally incomplete ORs were removed, leaving 56 ORs that otherwise correspond one-to-one with the benchmark ORs. Two ORs that had been annotated before filtering, DmOr59a and DmOr85e, were removed by the structural filter because DmOr59a was fragmented and DmOr85e lacked TM7 in the predicted structural model.

The published *A. gambiae* OR repertoire contains 79 OR genes. The automated annotation workflow initially recovered 77 ORs, including 72 previously reported ORs. After removing fragmented ORs and five literature-annotated full-length ORs, 67 structurally intact ORs were retained. The retained ORs otherwise corresponded to literature annotations. Because five published full-length ORs were excluded by the structural filter, we further examined their predicted structures. AgOr52, AgOr47, and AgOr64 lacked one complete transmembrane region. AgOr58 and AgOr6 contained all expected transmembrane regions, but showed local deletions in part of the transmembrane regions. These results indicate that structural filtering can remove clearly incomplete ORs, but may also exclude a small number of genuine ORs with atypical or locally incomplete transmembrane-region features.

The published *B. mori* OR repertoire contains 66 OR genes, including two pseudogenes. The automated workflow initially annotated 103 OR candidates, including 60 literature-annotated ORs. After structural filtering, fragmented ORs and seven literature-annotated ORs were removed, and 64 structurally intact ORs were retained. Among these, 53 corresponded to literature-annotated ORs. Some retained candidates showed high sequence similarity to literature ORs, whereas others showed lower similarity and did not clearly correspond to previously reported ORs. To assess whether such candidates could result from exon-merging errors in tandem OR regions, we examined their genomic positions. Most candidates did not occur on the same chromosome or in the same tandem region as their closest literature-annotated ORs. Therefore, these candidates are unlikely to be simple exon-merging artifacts involving adjacent OR genes, and may represent previously unannotated OR candidates. Nevertheless, transcriptomic evidence or manual gene-model curation will be needed to confirm these candidates.

Because previously known OR sequences from some benchmark species were included in the query database, we next tested whether the presence of conspecific queries systematically influenced the final inferred OR repertoire size. We divided the benchmark species into two groups: those with conspecific OR sequences represented in the query database and those without conspecific queries. We detected no significant difference in OR annotation performance between the two groups (p=0.8, ***Figure 1—figure supplement 1B***). Because *Anagrus nilaparvatae* showed an

unusually large difference in OR number between the two studies, we further performed a sensitivity analysis excluding this species, and the conclusion remained unchanged (p=0.4, ***Figure 1—figure supplement 1C***). Therefore, among species for which comparable genome-based reference repertoires are currently available, we found no evidence that the presence or absence of conspecific ORs in the query database systematically affects the final inferred OR repertoire size.

To extend the benchmarking beyond these three detailed case studies, we compared the final numbers of structurally intact ORs with published repertoire counts from additional insect species (***Supplementary file 1A***). For most species with comparable genome-based reference repertoires, the number of ORs retained in this study exceeded 70% of the published count and showed generally good agreement in repertoire size. *Diabrotica virgifera virgifera* was an exception: a previous manual annotation reported 124 complete ORs, whereas our pipeline retained 78 structurally intact ORs, approximately 63% of the published count. This discrepancy likely reflects the challenges posed by the large, repeat-rich genome of *D. virgifera virgifera*, in which repetitive sequences and complex OR loci can complicate automated gene prediction and locus reconstruction. The previous species-specific annotation incorporated transcriptomic evidence and locus-by-locus manual curation, which can help resolve gene models in complex genomic regions. Therefore, transcriptome-assisted manual annotation may be required to obtain more comprehensive OR repertoires for species with large, repeat-rich, or otherwise complex genomes.

Overall, the benchmarking analysis indicates that the automated annotation workflow recovers most known OR repertoires in well-annotated species and that structural filtering improves dataset reliability for downstream structural comparisons and docking-profile analyses. The final dataset was deliberately designed as a conservative, high-confidence set of structurally intact ORs rather than the broadest possible species-level gene catalog. Accordingly, lower counts relative to some published repertoires are an expected consequence of uniformly excluding fragments, pseudogenes, and candidates that do not meet the structural inclusion criteria. To evaluate whether this conservative filtering influenced the ecological analyses, we repeated the phylogenetic comparative analyses using the larger of our structurally intact OR count and the corresponding published count. The main associations between OR repertoire size and ecological traits remained unchanged (***Figure 1—figure supplement 1D–G***), indicating that the filtering strategy did not affect our conclusions.

Appendix 2

Hit-rate benchmark for docking-score threshold selection

To evaluate whether the docking-score threshold was driven by any single functional dataset, we performed hit-rate analyses using experimentally characterized OR–VOC pairs from multiple insect groups.

When *D. melanogaster*, *Anopheles gambiae*, and Lepidoptera datasets were analyzed separately, they showed similar docking score-hit rate relationships. In all three datasets, experimentally responsive OR–VOC pairs were enriched as docking scores became more negative. The hit-rate plateau for these datasets was approximately 23%, and the corresponding docking-score threshold was around –8 kcal/mol. This result is consistent with the threshold used in the main analysis and indicates that the –8 kcal/mol cutoff was not driven by a single species or dataset.

We also included *Locusta migratoria* OR functional data as an independent sensitivity analysis. The locust dataset differed from the other datasets in two important ways. First, the experimental positive rate was low, approximately 5.2%. Second, previously characterized locust ORs are mostly narrowly tuned receptors.

Consistent with these properties, the locust dataset showed lower overall hit rates than the *Drosophila*, *Anopheles*, and Lepidoptera datasets. In most docking-score intervals, the locust hit rate remained below 10% (***Figure 5—figure supplement 1G***). When the locust dataset was combined with the other three datasets, the overall hit-rate plateau decreased to approximately 18%, and the corresponding docking-score threshold shifted to approximately –11 kcal/mol (***Figure 5—figure supplement 1H***).

The locust result indicates that OR tuning breadth and the experimental positive rate of a dataset can influence the relationship between docking score and hit rate. For datasets dominated by narrowly tuned ORs and low experimental positive rates, a stricter docking-score threshold may be required to achieve stronger enrichment of experimentally responsive OR–VOC pairs.

However, the goal of the present study was to establish a unified empirical threshold for large-scale comparison across thousands of ORs and a broad VOC library, rather than to optimize separate thresholds for individual lineages. Therefore, the –8 kcal/mol threshold derived from the *Drosophila*, *Anopheles*, and Lepidoptera datasets was retained for the main analysis, whereas the locust result was treated as a sensitivity analysis and used to define the limitation of applying a single cutoff across all insect groups.

The hit-rate benchmark supports the use of docking scores for repertoire-level enrichment and comparative analysis. Nevertheless, the plateau hit rate remains limited and is not sufficient to assign definitive ligands or experimentally validated response spectra to individual ORs.

Therefore, docking results in this study should be interpreted as docking-derived binding potential. They provide a high-throughput, internally comparable framework for identifying repertoire-level trends and candidate OR–VOC relationships for future testing, but specific receptor-ligand pairs require validation by heterologous expression, electrophysiology, calcium imaging, genetic perturbation, or behavioral assays.