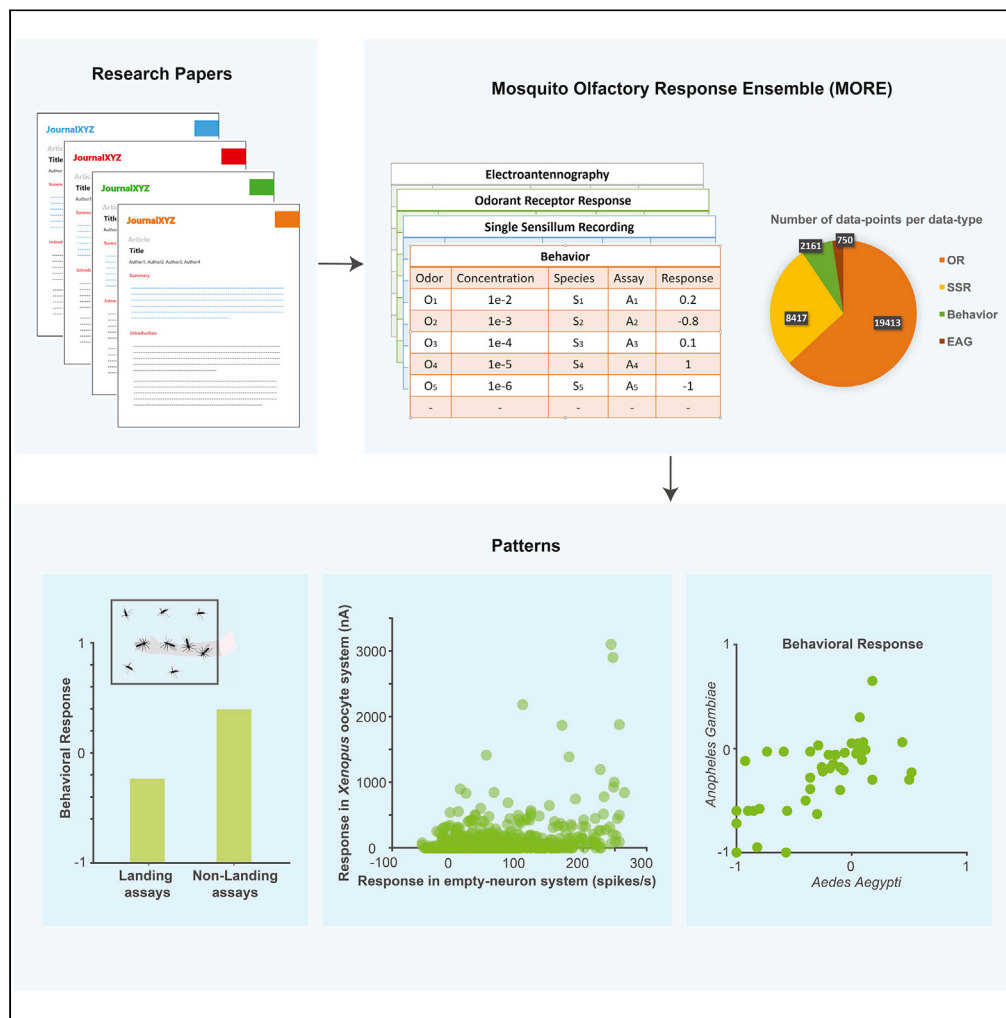


Article

Mosquito Olfactory Response Ensemble enables pattern discovery by curating a behavioral and electrophysiological response database



Abhishek Gupta, Swikriti S. Singh, Aarush M. Mittal, ..., Karthikeyan R. Kannan, Arjit K. Gupta, Nitin Gupta

guptan@iitk.ac.in

Highlights
MORE is a database of behavioral and electrophysiological responses to odors

MORE includes data from 170 studies covering 12 species of mosquitoes along with flies

MORE shows differences in odor preferences measured with different assays

Empty-neuron technique measures responses more sensitively than the oocyte technique

Gupta et al., iScience 25, 103938
March 18, 2022 © 2022 The Author(s).
<https://doi.org/10.1016/j.isci.2022.103938>



Article

Mosquito Olfactory Response Ensemble enables pattern discovery by curating a behavioral and electrophysiological response database

Abhishek Gupta,^{1,2,4} Swikriti S. Singh,¹ Aarush M. Mittal,¹ Pranjul Singh,¹ Shefali Goyal,¹ Karthikeyan R. Kannan,¹ Arjit K. Gupta,¹ and Nitin Gupta^{1,3,5,*}

SUMMARY

Many experimental studies have examined behavioral and electrophysiological responses of mosquitoes to odors. However, the differences across studies in data collection, processing, and reporting make it difficult to perform large-scale analyses combining data from multiple studies. Here we extract and standardize data for 12 mosquito species, along with *Drosophila melanogaster* for comparison, from over 170 studies and curate the Mosquito Olfactory Response Ensemble (MORE), publicly available at <https://neuralsystems.github.io/MORE>. We demonstrate the ability of MORE in generating biological insights by finding patterns across studies. Our analyses reveal that ORs are tuned to specific ranges of several physicochemical properties of odorants; the empty-neuron recording technique for measuring OR responses is more sensitive than the *Xenopus* oocyte technique; there are systematic differences in the behavioral preferences reported by different types of assays; and odorants tend to become less attractive or more aversive at higher concentrations.

INTRODUCTION

Mosquitoes find hosts for blood-feeding using various cues, including odors released by the hosts (Degenaro et al., 2013; McBride, 2016; Vinauger et al., 2019). Odorants are detected by sensory neurons located on the peripheral sensory organs, primarily the antennae and the maxillary palps. These neurons express various receptors for detecting odors, including odorant receptors (ORs), gustatory receptors (GRs), and ionotropic receptors (IRs). Sensory neurons transmit information to the antennal lobe in the brain for further processing (Anton et al., 2003; Su et al., 2009).

Because of their relevance to diseases, the olfactory behaviors of mosquitoes have been studied for a very long time (Kline et al., 1990; Mehr et al., 1990; Price et al., 1979). Researchers have employed several types of behavioral assays, such as Y-tube olfactometers, dual-port assays, arm-in-cage landing assays, wind-tunnels, tip assays, and T-mazes (Afify and Potter, 2020; Geier and Boeckh, 1999; Knaden et al., 2012; Logan et al., 2010; Macwilliam et al., 2018; Pates et al., 2001; Simonnet et al., 2014; Spitzen et al., 2013) to quantify the behavior. In parallel, various electrophysiology techniques such as electroantennography, single-sensillum recordings (both in wild-type animals as well as in heterologous expression systems), and voltage-clamp recordings of receptors have been used to quantify the sensory responses to many different odorants (Carey et al., 2010; de Bruyne et al., 2001; de Fouchier et al., 2017; Hallem and Carlson, 2006; Wang et al., 2010). Recent advancements in the techniques to produce transgenic insects have further boosted the research on the mosquito olfactory system (Afify et al., 2019; Kistler et al., 2015; Raji et al., 2019; Riabinina et al., 2016).

The behavioral and electrophysiological data produced from these studies are currently scattered across hundreds of research articles in an unformatted way. Having all this data in one place, in a structured format, can enable systematic large-scale analyses to discover trends that cannot be seen with individual studies in a variety of animal models (Crauto et al., 2002; Liu et al., 2004, 2011; Marengo et al., 2016; Olender et al., 2013). The Database of Odorant Responses (DoOR) catalogs the OR responses of different odors in *Drosophila melanogaster* (Galizia et al., 2010; Münch and Galizia, 2016) and has proved to be very useful in enabling large-scale computational analyses (Chepurwar et al., 2019; Dasgupta et al., 2018; Saberi and

¹Department of Biological Sciences and Bioengineering, Indian Institute of Technology Kanpur, Kanpur, Uttar Pradesh 208016, India

²Department of Chemistry, Indian Institute of Technology Kanpur, Kanpur, Uttar Pradesh 208016, India

³Mehta Family Center for Engineering in Medicine, Indian Institute of Technology Kanpur, Kanpur, Uttar Pradesh 208016, India

⁴Division of Biological Sciences, University of California San Diego, La Jolla, CA 92093, USA

⁵Lead contact

*Correspondence: guptan@iitk.ac.in

<https://doi.org/10.1016/j.isci.2022.103938>



Seyed-Allaei, 2016; Zwicker et al., 2016). However, no such curated dataset is available for mosquitoes. Further, while DoOR only has OR response data, a curated dataset that puts together different kinds of behavioral and electrophysiological recordings would be more powerful.

Here we have compiled the available behavioral and electrophysiological olfactory responses in many species of mosquitoes from over 170 research papers. This curated dataset brings results from diverse sources into a standard format. We have annotated each data-point with various experimental parameters, such as the concentration of the odorant used or the age and the sex of the animals on which the experiments were performed. We demonstrate how the dataset can be used to gain insights into the olfactory system.

RESULTS

Curating a comprehensive dataset of olfactory responses

We manually collected a large number of research articles that have reported different kinds of olfactory responses in mosquitoes. The responses were sorted into four different data-types: (1) OR: electrophysiological measurements from genetically labeled odorant receptors, using the empty-neuron system (Carey et al., 2010) or other heterologous expression systems; (2) SSR: single-sensillum recordings without genetic identification of the odorant receptors; (3) EAG: electroantennogram recordings; (4) Behavior: measurements of behavioral preferences to odors.

In most of the articles, the data were reported in text or plots, rather than spreadsheets, and thus had to be manually extracted (see STAR methods). Preprocessing was often required to convert the data into standard formats: For example, odor preference results could be reported as preference index, percent repellency, percent attraction, etc.; we converted all of them to a common metric – the preference index, calculated as the number of animals choosing the test odor minus the number of animals choosing the control divided by the sum of the two numbers. Similarly, EAG and OR response datasets were processed wherever required to ensure uniformity in data normalization and background subtraction (see STAR methods).

In total, we collected 30,741 data-points (Figure 1A), where each data-point corresponds to one of the 4 types of responses for an odorant, covering a total of 758 different odorants (Figure S1A). Care was taken to map the entries reported for different synonyms of the same odorant to a standard name, and to convert the odorant concentrations into standard units (see STAR methods). We were able to collect data from 12 different species of mosquitoes: *Anopheles gambiae*, *Aedes aegypti*, *Culex quinquefasciatus*, *Anopheles stephensi*, *Culex pipiens*, *Aedes albopictus*, *Culex nigripalpus*, *Culex tarsalis*, *Anopheles quadrimaculatus*, *Anopheles quadriannulatus*, *Anopheles arabiensis*, and *Anopheles coluzzii*; data from *D. melanogaster* was also included to help with comparative analyses (Figure 1B). The data were sourced from 170 different research papers (Figure 1C), published over a period of more than 4 decades (Figure S1B).

The comprehensive dataset allowed us to note some trends in the experimental preparations. In terms of the number of papers and the number of data points, the three mosquito species with maximum research on the olfactory system are *Anopheles gambiae*, *Aedes aegypti*, and *Culex quinquefasciatus* (Figures 1B and 1C). Among all SSR recordings that are available for mosquitoes, nearly 73% have been performed on the trichoid sensilla (Figure 1D). Among EAG studies, we noted the experiments have been performed in three kinds of preparations: intact animals, isolated heads, and isolated antenna; the isolated head preparation has been used in more than half of the studies (Figure 1E). We also checked the ages of mosquitoes used in behavioral and electrophysiological studies and found that most studies have used animals of age 6–8 days for both kinds of experiments, with only a few using younger or older animals (Figure 1F).

Web interface for accessing the dataset

We have made the whole dataset available freely through a website: <http://neuralsystems.github.io/MORE>. The website is organized into 4 sections, corresponding to the 4 data types. The data are displayed in a tabulated format, which can be sorted in the increasing or the decreasing order of any selected feature (Figure S2). Each row provides one data-point for an odor, along with the corresponding experimental details (such as odor concentration or the species used) and the reference. The row can be expanded to see additional details about the experimental conditions. A search box allows users to enter a term, so that only the rows containing the terms are displayed; this is particularly useful if a user wants to find data for a particular odor, a particular species, or a particular experimental condition, among the thousands of data points. The results displayed on any screen can be downloaded as an Excel spreadsheet.

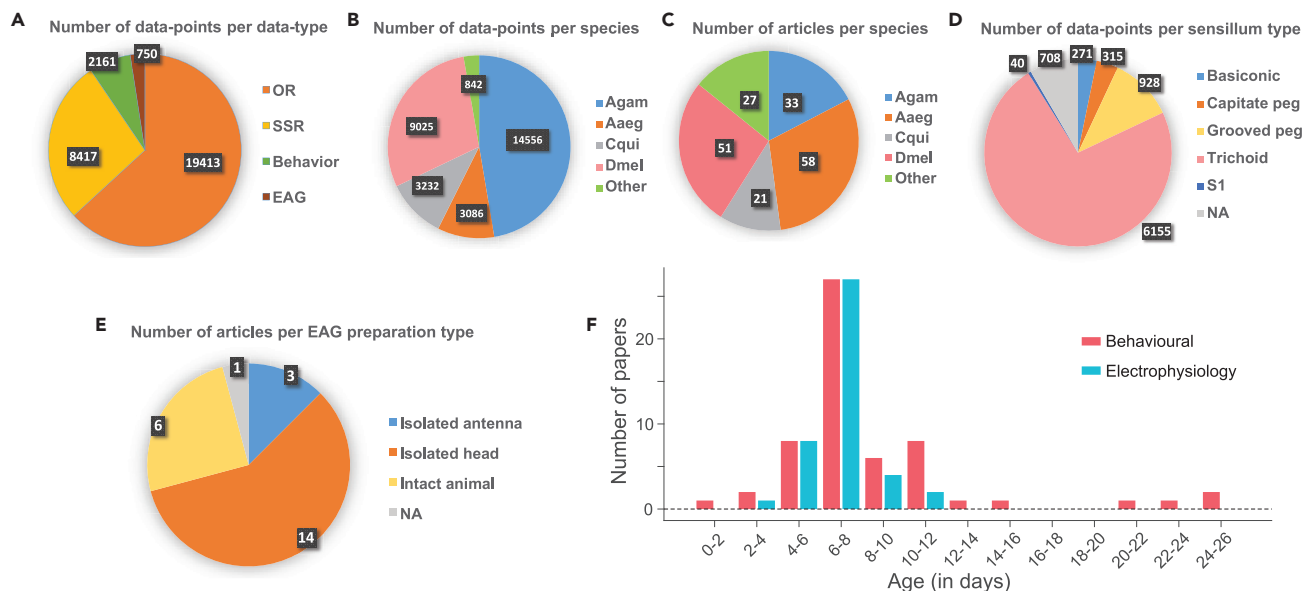


Figure 1. An overview of the dataset

(A) Pie-chart showing the number of data points corresponding to each data type. Each data point corresponds to the response to an odor, and a data type represents the type of response.

(B) Pie-chart indicates the number of data points per species.

(C) Pie-chart showing the number of studies from which data were extracted for each species. The total number of studies was 170, but some studies included data for multiple species.

(D) Pie-chart indicates the number of odor-responses per sensillum type in the Single Sensillum Recording dataset.

(E) Pie-chart showing the number of studies that use a particular animal preparation type for Electroantennography recordings.

(F) Bars indicate the number of studies versus the age group of the animal for behavioural (red) and electrophysiology (blue) experiments.

In D, E, 'NA' category represents not available.

Abbreviations: Agam, *Anopheles gambiae*; Aaeg, *Aedes aegypti*; Cqui, *Culex quinquefasciatus*; Dmel, *Drosophila melanogaster*; OR, Odorant Receptor response; SSR, Single Sensillum Recordings; Behavior, Behavioral preference; EAG, Electroantennography.

See also [Figures S1](#) and [S2](#).

The whole dataset can also be downloaded as an Excel file with a single click, without requiring any registration or permissions. This will enable other researchers to use this large and structured dataset, possibly in combination with new or other kinds of data, to conduct new analyses.

Relationship between OR responses and physicochemical properties of odorant molecules

An analysis of OR responses in *Drosophila* previously suggested that ORs tend to respond more strongly to odorant molecules whose volumes are in a specific range (Saber and Seyed-Allaei, 2016). Our dataset allowed us an opportunity to systematically examine such relationships between different physicochemical properties and OR responses, and further check if they are conserved between *Drosophila* and mosquitoes. We retrieved the physicochemical properties of odorants from PubChem and analyzed the correspondence between 13 different properties and OR responses (see [STAR methods](#)).

We found that the mosquito ORs responded most strongly to odorants with molecular volumes around 100 \AA^3 , in a bell-shaped tuning curve; interestingly, the tuning was largely overlapping between mosquitoes and *Drosophila* (Figure 2A). To quantify the tuning, we fitted the distribution to a Gaussian and estimated the standard deviation (σ); a smaller value of σ indicates a sharper tuning. To determine the statistical reliability of the observed tuning (with a null hypothesis of no tuning), we compared the observed σ with the values of σ obtained after shuffling the mapping between the responses and the molecular volume (see [STAR methods](#)). This analysis confirmed that the tuning observed for molecular volume was statistically reliable ($p < 0.001$).

Similar tuning curves were also observed for molecular weight (Figure 2B), octanol-water partition coefficient (Figure 2C), and molecular complexity (Figure 2D) in *A. gambiae* and *D. melanogaster*. In total, out

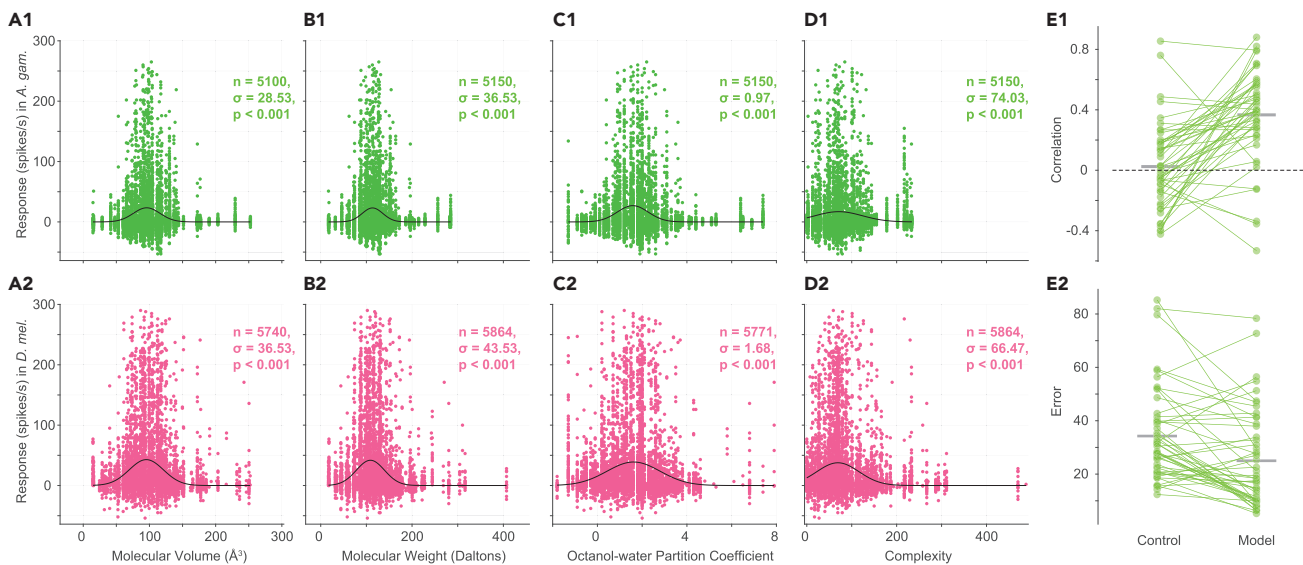


Figure 2. Relationship between OR responses and physicochemical properties of odorant molecules

(A–D) Scatterplots of OR responses (in spikes/second) to odors and the physicochemical properties of those odors, including Volume 3D (A), Molecular weight (B), Octanol-water partition coefficient (C), and Complexity parameter (D), in *Anopheles* (A1, B1, C1, D1) and *Drosophila* (A2, B2, C2, D2). Each point corresponds to an OR-odor pair. In the plots, n is the number of OR-odor pairs, σ is the standard deviation of the fitted gaussian, and p is the p -value. In all plots, the black curve corresponds to the fitted gaussian line. The value of n in different plots could differ as the values of all physico-chemical properties were not available for some of the odors.

(E) Comparison of the performances of the neural network model and the control model on the test dataset. E1 shows the correlations between the prediction and the actual odorant responses for each OR; E2 shows the error (absolute difference between actual and predicted responses) averaged over all test odors. In both plots, each point corresponds to an OR ($N = 50$).

See also [Figure S3](#).

of the 13 properties examined, we found statistically significant response tuning for 12 properties ([Figure S3](#)); the only exception was “Conformer Count 3D” (the number of different conformers of the molecule; [Figure S3A](#)). Overall, this analysis suggests that insect olfactory systems have evolved to respond preferentially to molecules whose various physicochemical properties lie in certain ranges.

Next, we checked whether a model could be trained to predict the OR responses in *A. gambiae* using the physicochemical properties of odorants. For this analysis, we used a larger set of 295 properties (see [STAR methods](#)). We trained a feedforward neural network model for each OR using 70% of the odors for training, 15% for model validation and keeping 15% for test (see [STAR methods](#)). We found that the responses predicted by the model for the test odors showed higher correlations ($R = 0.37 \pm 0.30$, mean \pm s.d.; $N = 50$ ORs) with the actual responses for the same odors with in an OR, compared the predictions of a control model ($R = 0.02 \pm 0.28$; $N = 50$ ORs; see [Methods](#)); the difference was statistically significant ($P = 8.73 \times 10^{-7}$, sign-rank test; $N = 50$ ORs; [Figure 2E1](#)). We also checked the magnitude of the errors in the predictions, quantified as the average of the absolute differences between the predicted and the actual responses for the test odors ([Figure 2E2](#)): the error for the model predictions (24.99 ± 17.96 , $N = 50$ ORs) was smaller than the error for the control predictions (34.26 ± 17.24 , $N = 50$ ORs) by 9.27 spikes/s ($P = 1.98 \times 10^{-4}$; $N = 50$ ORs). These results suggest that machine learning-based models can be used with physicochemical properties of odorants to predict OR responses to novel odorants.

Differences in techniques for measuring OR responses

Next, we compared different methods for calculating OR responses to odors. In *A. gambiae*, responses of many ORs to a large panel of odors have been measured using two different methods: (1) The empty neuron system ([Carey et al., 2010](#)) in which the OR of interest is expressed in an accessible sensillum of *Drosophila* in place of the native OR, and its response is then measured in the units of spikes/s using the SSR technique; (2) the *Xenopus* oocyte expression system ([Wang et al., 2010](#)), in which the OR of interest is expressed in an oocyte, and the response is then measured in the units of nano-Amperes using two-electrode voltage clamp.

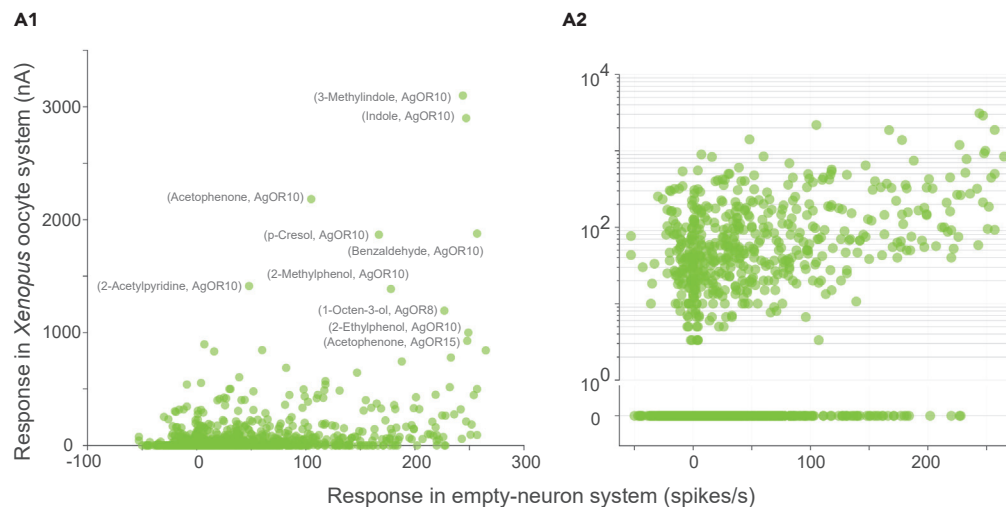


Figure 3. Differences in techniques for measuring OR responses

Scatter plots of OR responses in *Xenopus* oocyte expression system (in units of nanoampere) and empty-neuron system (in units of spikes/second) in linear (A1) and logarithmic (A2) y-scale ($n = 2423$ OR-odor measurements). In panel A1, responses greater than 900 in the oocyte system are labeled with the name of the odorant and the OR.

By comparing the responses in these two datasets for the same OR-odor combinations, we found that a large fraction (1738 out of 2423; 71.7%) of combinations show non-zero responses in the empty-neuron system and zero responses in the oocyte-recording technique. However, very few combinations (13 out of 2423; 0.5%) show the reverse trend of zero responses in empty-neuron and non-zero responses in oocyte-recording (Figure 3). To understand the reason for this surprising abundance of zero values in the oocyte recordings, we checked if these mainly correspond to OR-odor combinations that generate an inhibitory (negative) response in the empty-neuron recordings. We found that the zero responses in oocyte-recordings are not limited to cases where the empty-neuron response is negative: In fact, out of 1889 cases with zero responses in oocyte recordings, 918 (48.6%) have a positive response in the empty-neuron recordings, 151 (8%) have zero response and only 820 (43.4%) have a negative response. Moreover, among the combinations that have negative empty-neuron responses, the responses of these latter 820 combinations with zero oocyte responses are no more negative (-8.55 ± 8.64 spikes/s, mean \pm s.d.) than the responses of the 99 combinations with non-zero oocyte responses (-11.32 ± 10.56). Thus, the zero responses in oocyte recordings do not necessarily correspond to inhibitory responses; rather, our analysis of these two datasets suggests that the oocyte recording technique is less sensitive than the empty-neuron technique at detecting OR responses for the same set of odors.

Differences in behavioral assays

Many different assays have been devised and used by different laboratories to measure the behavioral attractiveness or aversiveness of an odor. Our data collection from multiple studies offers an opportunity to see the most frequently used assays and their relative abundance.

In mosquitoes, these assays belonged to five broad categories: Y-tube (Geier et al., 1996), dual-port (Pates et al., 2001), wind-tunnel (Healy and Copland, 2000), tip (Afify and Potter, 2020), and landing (or arm-in-cage) assays (Ali et al., 2017; Logan et al., 2010) (Figure 4A1). Although all these assays quantify the preference of mosquitoes to the tested odor, they can differ in the odorant exposure profile and the specific motor actions used by the mosquitoes: For example, a Y-tube assay involves a choice between two alternatives in a confined chamber, a wind-tunnel involves free flight movement toward an odor source in a large chamber, and a landing assay involves the termination of flight followed by landing very close to an odor source. For adult mosquitoes, Y-tube assay was the most abundant (33.3%), followed by dual-port (28.6%) and landing assays (28.6%), and wind-tunnel (7.9%) and tip assay (1.6%) were the least abundant. In *Drosophila*, the assays could be grouped into three categories: Y-maze (Charro and Alcorta, 1994), T-maze (Helfand and Carlson, 1989), and dual-port (or trap) assays (Knaden et al., 2012) (Figure 4A2). Among these, T-maze assays were the most common (54.6%) in our dataset, followed by Y-maze (24.2%) and dual-port (21.2%) assays.

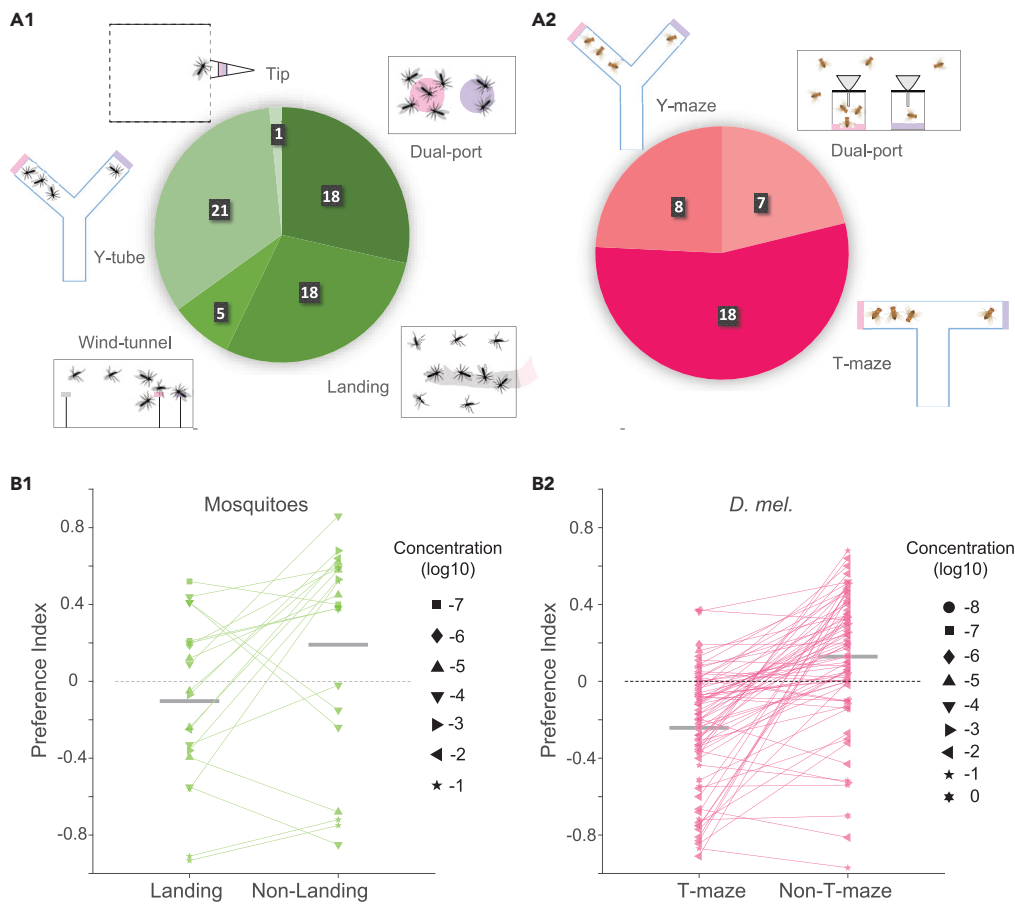


Figure 4. Differences in behavioral assays

(A1 and A2) Pie-charts showing the number of studies from which data were extracted for different behavioral assays in mosquitoes (A1) and *Drosophila* (A2).

(B1 and B2) Comparison of the preference indices in landing and non-landing (Dual-port and Y-tube) assays in mosquitoes (B1), and T-maze and non-T-maze (Dual-port and Y-maze) assays in *Drosophila* (B2). Each point represents an odor-concentration pair ($n = 20$ for mosquitoes and $n = 74$ for *Drosophila*). The size of each dots indicates the odor concentration (see legend). Horizontal gray lines represent means.

The behavioral attraction or aversion for an odor, estimated from these assays, is often reported as a preference index. Although some variability is expected in the preference indices measured in different studies, because of experimental noise or minor differences in the experimental conditions, it is not known if there are systematic biases in the preference indices reported by different types of assays. We used our large dataset to explore the possibility of such systematic differences. In mosquitoes, we noticed that the preference indices obtained in the landing assays were often smaller (or more negative) than the indices for the same odors in other types of assays ($P = 1.68 \times 10^{-2}$, $N = 20$ pairs of data points from *Aedes aegypti* and *Culex quinquefasciatus*; Figure 4B1). In *Drosophila*, we found that T-maze assays reported odors to be more aversive than Y-maze or dual-port assays ($P = 6.03 \times 10^{-11}$, $N = 74$ pairs of odors; signrank test; Figure 4B2). These results highlight the need for caution when comparing behavioral preferences of odors across different studies.

Relationship between the preference index and the oviposition index

Behavioral preferences are governed by the internal states of the animals (Sayin et al., 2018). There are examples where the attraction or aversion to an odor during foraging behavior is different from that during egg-laying behavior. For example, *D. melanogaster* show avoidance to acetic acid in odor choice assays during foraging, but attraction to acetic acid during egg-laying (Joseph et al., 2009). Another study found that valencene, b-caryophyllene, b-caryophyllene oxide, and limonene oxide had very different preference indices (during foraging) and oviposition indices in *D. melanogaster* (Dweck et al., 2013).

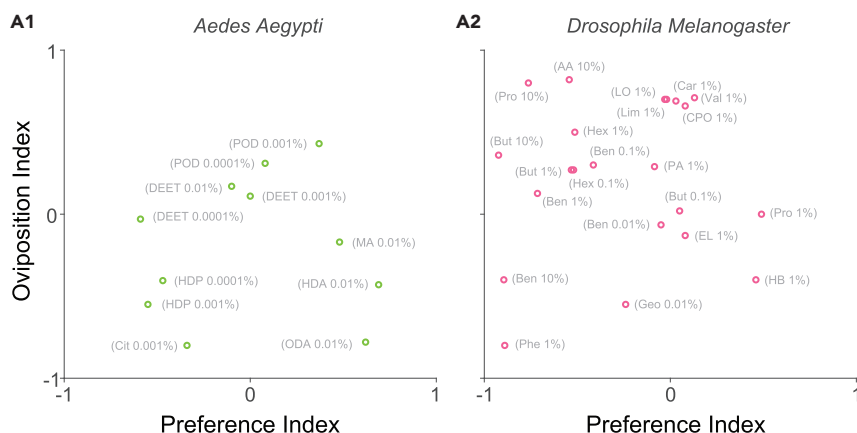


Figure 5. Relationship between the preference index and the oviposition index

(A1 and A2), Scatter plots between the preference index and the oviposition index in *Aedes aegypti* (A1) and *Drosophila* (A2). Each point represents an odorant at a specific concentration ($n = 11$ for mosquitoes and $n = 22$ for *Drosophila*). Odorant acronyms: DEET, N,N-diethyl-meta-toluamide; POD, Propyl Octadecanoate; MA, Myristic Acid; HDP, Hexadecyl Pentanoate; Cit, Citronellal; HDA, Hexadecanoic Acid; ODA, Octadecanoic Acid; Pro, 1-Propanol; AA, Acetic Acid; But, 1-Butanol; EL, Ethyl Lactate; Ben, Benzaldehyde; Hex, (E)-2-Hexenal; Phe, Phenol; Geo, Geosmin; HB, Hexylbutyrate; PA, Pentanoic Acid; CPO, (-)-Caryophyllene Oxide; LO, Limonene Oxide; Lim, Limonene; Val, Valencene; Car, B-Caryophyllene.

See also [Figure S4](#).

Our large collection of behavioral data allowed us to systematically examine whether the odor preferences during oviposition are independent of odor preferences during foraging or host-seeking. We selected odors for which both the oviposition index and the preference index were available in the dataset (see [STAR methods](#)), and then compared the two values ([Figure 5](#)). We found that the oviposition index was not correlated with the preference index during host-seeking in *A. aegypti* ($R = 0.06$, $P = 0.86$, $N = 11$; [Figure 5A1](#)) or the preference index during foraging in *D. melanogaster* ($R = 0.06$, $P = 0.78$, $N = 22$; [Figure 5A2](#)). Thus, the preferences during foraging or host-seeking appear completely independent of the preferences during oviposition.

Comparison of behavioral preference across mosquito species

Our dataset including multiple species of mosquitoes provided an opportunity to check how similar or different are the preferences indices of odors between the species. For each pair of species, we selected odors that were tested in both the species at similar concentrations (see [STAR methods](#)), and calculated the correlation coefficient between their preference indices in the two species. We found weak to moderate correlations in different pairs of species, with the Pearson correlation coefficient varying between 0.12 and 0.82 ([Figure S4](#)). The correlation between *A. gambiae* and *A. aegypti* was 0.68 ($n = 44$), while it was 0.59 between *C. quinquefasciatus* and *A. aegypti* ($n = 117$) and 0.46 between *A. aegypti* and *A. albopictus* ($n = 24$). We note that the correlation values are affected by the exact identities of the available common odors, which differed for different pairs.

Dependence of behavioral preference on odor concentration

The concentration of an odor can affect the behavioral preference. In *D. melanogaster*, there are examples where a 10-fold change in concentration can result in either an increase or a decrease in the preference index. [Figure 6A1](#) shows the preference indices of benzaldehyde with concentrations varying over 5 orders of magnitude in T-maze assays: Higher concentrations typically show more aversion. [Figure 6A2](#) shows the preference indices of ethanol with concentrations varying over 4 orders of magnitude in Y-maze assays: here, the preference increases from 10^{-3} to 10^{-1} , but decreases if the concentration is further increased.

To check if there may be a general pattern in how the preference index varies in response to increasing odor concentration, we collected pairs of preference indices at concentrations separated by a factor of 10 for the same odor, using the same type of assay in the same species ([Figure S6](#); see [STAR methods](#)). In 33 such pairs available in our dataset, we checked the difference between the preference indices at the higher concentration

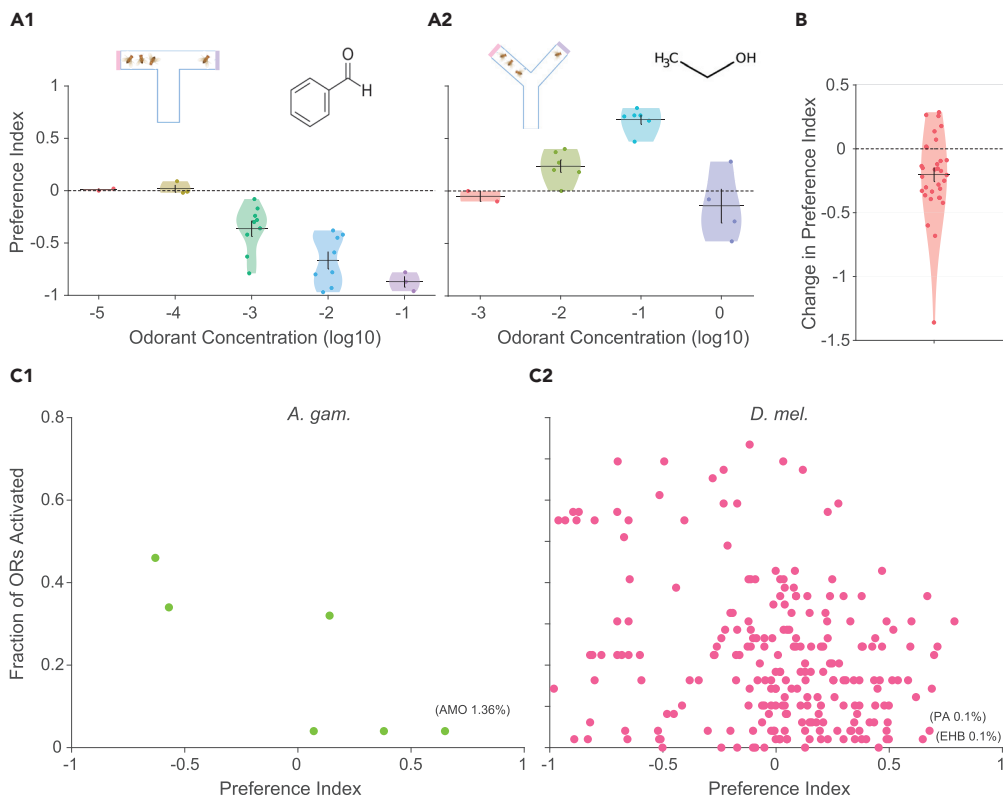


Figure 6. Dependence of behavioral preference on odor concentration

(A1 and A2) Violin plots show the preference indices for different concentrations of benzaldehyde (A1) and ethanol (A2) in *Drosophila*. Each point within a violin represents the preference index obtained from a different study.

(B) The plot shows the change (mostly reduction) in the preference index on increasing the concentration by 10-folds. Each point represents the preference index at a particular concentration minus the preference index at a 10-fold lower concentration of the same odor, in the same type of assay and same species. The plot includes data for $n = 33$ concentration pairs from *A. aegypti*, *A. albopictus*, *A. gambiae*, *C. pipiens*, and *D. melanogaster*. Odorant acronyms: AMO, Ammonia; PA, Propanoic Acid; EHB, Ethyl-3-hydroxybutyrate.

(C1 and C2) Scatter plots show the negative correlation between the number of ORs activated by an odor and the preference index of the odor, in *A. gambiae* (C1) and *D. melanogaster* (C2).

In all plots, the horizontal line and the error bar indicate the mean and the SEM, respectively.

See also [Figures S5](#) and [S6](#).

and at the lower concentration ([Figure 6B](#)). We found that increasing the odor concentration 10-folds decreases the preference index, on average by 0.2 ($P = 4.01 \times 10^{-4}$, $N = 33$; *signed rank test*). The same trend was observed even when we limited the analysis to only those pairs where the preference index at the lower concentration was negative (*mean change* = -0.14 , $P = 0.04$, $N = 11$; [Figure S5A1](#)) or positive (*mean change* = -0.23 , $P = 0.005$, $N = 22$; [Figure S5A2](#)). Thus, our analysis using this large dataset suggests that an increase in odor concentration tends to make aversive odors more aversive and attractive odors less attractive.

To probe this further, we checked the relationship between the number of ORs activated (increase of at least 10 spikes) by an odor and the preference index of the odor. This analysis revealed a negative correlation between the two parameters in both *A. gambiae* ($R = -0.83$, $P = 0.039$, $N = 6$; [Figure 6C1](#)) and *D. melanogaster* ($R = -0.27$, $P = 3.7 \times 10^{-5}$, $N = 230$; [Figure 6C2](#)). As higher concentrations are more likely to activate non-specific ORs, this provides a possible explanation for why higher concentrations tend to be more aversive. We further highlight the odors that activated a small fraction of the ORs and were highly attractive. In *A. gambiae*, ammonia 1.36% activates AgOR46 and AgOR50 and has a preference index of 0.65. In *D. melanogaster*, propanoic acid 0.1% activates OR24a and OR42a and has a preference index of 0.68, and ethyl-3-hydroxybutyrate 0.1% activates OR85a and has a preference index of 0.65.

DISCUSSION

In summary, we have curated MORE, a large dataset of behavioral and electrophysiological responses in 12 species of mosquitoes along with *D. melanogaster*. The dataset includes 30,741 data-points for 758 odorants collected from 170 research articles.

Bringing this scattered information into a well-structured database involved several challenges, because of the different preprocessing steps or the different units or metrics used by the studies. Some studies normalized the EAG responses using the responses to a reference odor, while others reported the raw values. Some studies reported the OR electrophysiological responses after subtracting the background activity, while others reported without this step. In MORE, we processed the data to include uniform normalization and background subtraction. Odor preferences in behavioral studies were reported using a variety of metrics, such as preference index (Yu et al., 2015), percent attraction (Geier et al., 1996), percent repellency (Islam et al., 2017b), protective efficacy (Logan et al., 2010), and so on. In MORE, we converted the reported preferences in all papers into the common metric of preference index. The odor concentrations were reported in a variety of units, which had to be standardized before their inclusion in MORE. Different studies referred to the same odorants using different names. For example, isoamyl alcohol, isopentyl alcohol, and isopentanol, all are common names of 3-methyl-1-butanol. In MORE, we combined all such data-points using a single standard name for each odorant. In many studies, the data-points were not available in an accessible format, and had to be obtained either by requesting the original authors or by extracting from the figures using a computer script.

The structured format of MORE makes the data amenable to large-scale analyses of patterns across the datasets, as we have demonstrated here. MORE can also facilitate the application of machine learning methods that are particularly dependent on large and structured datasets. We have created an interactive website for browsing the data, while also providing an easy option for downloading the entire dataset for offline analyses.

We found that in mosquitoes as well as in flies, the sensory responses were tuned to specific ranges of various physicochemical properties of the odorant molecules. The knowledge of these ranges may be useful in designing synthetic agonists for the ORs. We observed reliable tuning for 12 of the 13 physicochemical properties we tested. One of these properties, the octanol-water partition coefficient, is known to be related to the air-mucus odorant partition coefficient (Scott et al., 2014). Another property, molecular complexity, has been reported to be the determinant of the number of olfactory notes and the pleasantness of smell (Kermen et al., 2011). The one property that did not show tuning was the number of different 3D conformers of the molecule—this is not surprising as this particular property informs about the possible variations in the molecule but does not tell about the shape of any specific structure, unlike the other 12 properties. We also observed that the molecular properties of the odorants could be used to train a neural network model for the predicting the OR responses to new odorants. The accuracy of these predictions is likely to improve as more data becomes available for training the model.

We found systematic differences in the OR responses recorded using the empty-neuron system and the *Xenopus* oocyte expression system. Our results generalize the experimental observation made by Wang et al. using one pheromone receptor (Or13) in *Helicoverpa assulta* (Wang et al., 2016). The sensitivity of the two techniques might differ due to differences in the levels of receptor expression in the different kinds of cells, and differences between the odor delivery through the liquid medium in the oocyte recording technique and the volatile odor delivery in the empty-neuron technique. These results highlight the need for caution when interpreting negative results from the oocyte expression system.

Our dataset revealed no correlation between the oviposition indices and the preference indices of odors in mosquitoes or *Drosophila*. This result is consistent with previous work showing that sensory processing and the choice of behavior are expected to be state-dependent (Barrozo et al., 2011; Cohn et al., 2015; Gadonne et al., 2016; Sayin et al., 2018; Vogt et al., 2021). We also found that higher odor concentrations were in general more aversive than lower concentrations of the same odorants (Figure 6). This effect may also be related to our observations that landing assays resulted in lower or more negative preference indices than the non-landing assays for mosquitoes and that T-maze assays resulted in more negative preference indices than non-T-maze assays for flies (Figure 4): we speculate that these differences could be because the landing assays bring mosquitoes closer to the odor source and expose them to the odorants

with less air dilution, and perhaps the size and shape of the T-maze expose flies to higher concentrations than Y-maze assays. In *Drosophila*, a low concentration of apple cider vinegar triggers attraction through a small set of activated glomeruli, but a higher concentration triggers aversion through activation of an additional glomerulus (Semmelhack and Wang, 2009). Our results show that this concentration-dependent aversion is a more general pattern extending across odors and species.

Limitations of study

Because previous studies have focused mostly on a few species of mosquitoes, such as *A. gambiae*, *A. aegypti*, and *C. quinquefasciatus*, the MORE database has relatively fewer points for other species of mosquitoes. We have focused on mono-molecular odorants currently; future work may explore how to incorporate odor blend (Mozuraitis et al., 2020). We have captured some of the parameters from the experiments; however, some potentially relevant parameters, such as the flow rate of the odorized air, could not be included because of inconsistent or incomplete reporting in the literature.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
 - Lead contact
 - Materials availability
 - Data and code availability
- METHOD DETAILS
 - Data extraction
 - Physicochemical properties
 - Neural network model for OR response prediction
 - Analyses of behavioral assays
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.isci.2022.103938>.

ACKNOWLEDGMENTS

We would like to thank Elissa A Hallem (University of California, Los Angeles), John Carlson (Yale School of Medicine), Laurence J. Zwiebel (Vanderbilt University), and Sharon Hill (Swedish University of Agricultural Sciences) for sharing raw data from their papers (Hallem and Carlson, 2006; Hill et al., 2009; Wang et al., 2010). We thank members of the N.G. lab for helpful discussions. We thank Aalap Shah (Indian Institute of Technology Kanpur) for helping with the graphics. This work was supported by the DBT/Wellcome Trust India Alliance Fellowship [grant number IA/I/15/2/502091] and the SERB Core Research Grant [CRG/2020/004719] awarded to N.G.

AUTHOR CONTRIBUTIONS

AG and NG designed research; AG, SSS, AMM, PS, SG, KRK and AKG collected data; AG performed data analysis in consultation with NG; AG and NG wrote the manuscript with inputs from all authors.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: April 15, 2021

Revised: December 21, 2021

Accepted: February 14, 2022

Published: March 18, 2022

REFERENCES

- Afify, A., and Potter, C.J. (2020). Insect repellents mediate species-specific olfactory behaviours in mosquitoes. *Malar. J.* 19, 1–10.
- Afify, A., Betz, J.F., Riabinina, O., Lahondère, C., and Potter, C.J. (2019). Commonly used insect repellents hide human odors from *Anopheles* mosquitoes. *Curr. Biol.* 29, 1–12.
- Ali, A., Cantrell, C.L., and Khan, I.A. (2017). A new in vitro bioassay system for the discovery and quantitative evaluation of mosquito repellents. *J. Med. Entomol.* 54, 1328–1336.
- Anton, S., Van Loon, J.J.A., Meijerink, J., Smid, H.M., Takken, W., and Rospars, J.P. (2003). Central projections of olfactory receptor neurons from single antennal and palp sensilla in mosquitoes. *Arthropod. Struct. Dev.* 32, 319–327.
- Barrozo, R.B., Jarrault, D., Deisig, N., Gemenó, C., Monsempes, C., Lucas, P., Gadenne, C., and Anton, S. (2011). Mating-induced differential coding of plant odour and sex pheromone in a male moth. *Eur. J. Neurosci.* 33, 1841–1850.
- Blackwell, A., and Johnson, S.N. (2000). Electrophysiological investigation of larval water and potential oviposition chemo-attractants for *Anopheles gambiae*s. *Ann. Trop. Med. Parasitol.* 94, 389–398.
- Braks, M.A.H., Meijerink, J., and Takken, W. (2001). The response of the malaria mosquito, *Anopheles gambiae*, to two components of human sweat, ammonia and L-lactic acid, in an olfactometer. *Physiol. Entomol.* 26, 142–148.
- Caballero-Vidal, G., Bouysset, C., Grunig, H., Fiorucci, S., Montagné, N., Golebiowski, J., and Jacquín-Joly, E. (2020). Machine learning decodes chemical features to identify novel agonists of a moth odorant receptor. *Sci. Rep.* 10, 1–9.
- Carey, A.F., Wang, G., Su, C.-Y., Zwiebel, L.J., and Carlson, J.R. (2010). Odorant reception in the malaria mosquito *Anopheles gambiae*. *Nature* 464, 66–71.
- Charro, M.J., and Alcorta, E. (1994). Quantifying relative importance of maxillary palp information on the olfactory behavior of *Drosophila melanogaster*. *J. Comp. Physiol.* 175, 761–766.
- Chepurwar, S., Gupta, A., Haddad, R., and Gupta, N. (2019). Sequence-based prediction of olfactory receptor responses. *Chem. Senses* 44, 693–703.
- Cohn, R., Morante, I., and Ruta, V. (2015). Coordinated and compartmentalized neuromodulation shapes sensory processing in *Drosophila*. *Cell* 163, 1742–1755.
- Constantini, C., Birkett, M.A., Gibson, G., Ziesmann, J., Sagnon, N.F., Mohammed, H.A., Coluzzi, M., and Pickett, J.A. (2001). Electroantennogram and behavioural responses of the malaria vector *Anopheles gambiae* to human-specific sweat components. *Med. Vet. Entomol.* 15, 259–266.
- Cork, A., and Park, K.C. (1996). Identification of electrophysiologically-active compounds for the malaria mosquito, *Anopheles gambiae*, in human sweat extracts. *Med. Vet. Entomol.* 10, 269–276.
- Crauto, C., Marengo, L., Miller, P., and Shepherd, G. (2002). Olfactory Receptor Database: a metadata-driven automated population from sources of gene and protein sequences. *Nucleic Acids Res.* 30, 354–360.
- Dasgupta, S., Sheehan, T.C., Stevens, C.F., and Navlakha, S. (2018). A neural data structure for novelty detection. *Proc. Natl. Acad. Sci. U S A* 115, 13093–13098.
- de Bruyne, M., Foster, K., and Carlson, J.R. (2001). Odor coding in the *Drosophila* antenna. *Neuron* 30, 537–552.
- de Fouchier, A., Walker, W.B., Montagné, N., Steiner, C., Binyameen, M., Schlyter, F., Chertemps, T., Maria, A., François, M.C., Monsempes, C., et al. (2017). Functional evolution of Lepidoptera olfactory receptors revealed by deorphanization of a moth repertoire. *Nat. Commun.* 8, 1–11.
- Degennaro, M., McBride, C.S., Seeholzer, L., Nakagawa, T., Dennis, E.J., Goldman, C., Jasinskiene, N., James, A.A., and Vosshall, L.B. (2013). Orco mutant mosquitoes lose strong preference for humans and are not repelled by volatile DEET. *Nature* 498, 487–491.
- Dobritsa, A.A., Naters, W.V.D.G., Warr, C.G., Steinbrecht, R.A., and Carlson, J.R. (2003). Integrating the molecular and cellular basis of odor coding in the *Drosophila* antenna. *Neuron* 37, 827–841.
- Dweck, H.K.M., Ebrahim, S.A.M., Kromann, S., Bown, D., Hillbur, Y., Sachse, S., Hansson, B.S., and Stensmyr, M.C. (2013). Olfactory preference for egg laying on citrus substrates in *Drosophila*. *Curr. Biol.* 23, 2472–2480.
- Gadenne, C., Barrozo, R.B., and Anton, S. (2016). Plasticity in insect olfaction: to smell or not to smell? *Annu. Rev. Entomol.* 61, 317–333.
- Galizia, C.G., Münch, D., Strauch, M., Nissler, A., and Ma, S. (2010). Integrating heterogeneous odor response data into a common response model: a DoOR to the complete olfactome. *Chem. Senses* 35, 551–563.
- Geier, M., and Boeckh, J. (1999). A new Y-tube olfactometer for mosquitoes to measure the attractiveness of host odours. *Entomol. Exp. Appl.* 92, 9–19.
- Geier, M., Sass, H., and Boeckh, J. (1996). A search for components in human body odour that attract female *Aedes aegypti*. *Ciba Found. Symp.* 200, 132–148.
- Guha, L., Seenivasagan, T., Iqbal, S.T., Agrawal, O.P., and Parashar, B.D. (2014). Behavioral and electrophysiological responses of *Aedes albopictus* to certain acids and alcohols present in human skin emanations. *Parasitol. Res.* 113, 3781–3787.
- Halle, E.A., and Carlson, J.R. (2006). Coding of odors by a receptor repertoire. *Cell* 125, 143–160.
- Healy, T.P., and Copland, M.J.W. (2000). Human sweat and 2-oxopentanoic acid elicit a landing response from *Anopheles gambiae*. *Med. Vet. Entomol.* 14, 195–200.
- Helfand, S.L., and Carlson, J.R. (1989). Isolation and characterization of an olfactory mutant in *Drosophila* with a chemically specific defect. *Proc. Natl. Acad. Sci. U S A* 86, 2908–2912.
- Hill, S.R., Hansson, B.S., and Ignell, R. (2009). Characterization of antennal trichoid sensilla from female Southern house mosquito, *Culex quinquefasciatus* say. *Chem. Senses* 34, 231–252.
- Islam, J., Dhiman, S., Tyagi, V., Duarah, S., Zaman, K., and Chattopadhyay, P. (2017a). Behavioural and electrophysiological responses of mosquito vectors *Aedes aegypti*, *Anopheles stephensi* and *Culex quinquefasciatus* to an ethyl ester: ethyl 2-aminobenzoate. *J. Insect Behav.* 30, 343–358.
- Islam, J., Zaman, K., Tyagi, V., Duarah, S., Dhiman, S., and Chattopadhyay, P. (2017b). Protection against mosquito vectors *Aedes aegypti*, *Anopheles stephensi* and *Culex quinquefasciatus* using a novel insect repellent, ethyl anthranilate. *Acta Trop.* 174, 56–63.
- Joseph, R.M., Devineni, A.V., King, I.F.G., and Heberlein, U. (2009). Oviposition preference for and positional avoidance of acetic acid provide a model for competing behavioral drives in *Drosophila*. *Proc. Natl. Acad. Sci. U S A* 106, 11352–11357.
- Kermen, F., Chakirian, A., Sezille, C., Jousain, P., Le Goff, G., Ziessel, A., Chastrette, M., Mandairon, N., Didier, A., Rouby, C., and Bensafi, M. (2011). Molecular complexity determines the number of olfactory notes and the pleasantness of smells. *Sci. Rep.* 1, 1–5.
- Kistler, K.E., Vosshall, L.B., and Matthews, B.J. (2015). Genome engineering with CRISPR-Cas9 in the mosquito *aedes aegypti*. *Cell Rep.* 11, 51–60.
- Kline, D.L., Takken, W., and Wood, J.R. (1990). Field studies on the potential of butanone, carbon dioxide, honey extract, 1-octen-3-ol, l-lactic acid and phenols as attractants for mosquitoes. *Med. Vet. Entomol.* 4, 383–391.
- Knaden, M., Strutz, A., Ahsan, J., Sachse, S., and Hansson, B.S. (2012). Spatial representation of odorant valence in an insect brain. *Cell Rep.* 1, 392–399.
- Liu, N., Xu, F., Marengo, L., Hyder, F., Miller, P., and Shepherd, G.M. (2004). Informatics approaches to functional MRI odor mapping of the rodent olfactory bulb: OdorMapBuilder and OdorMapDB. *Neuroinformatics* 2, 3–18.
- Liu, X., Su, X., Wang, F., Huang, Z., Wang, Q., Li, Z., Zhang, R., Wu, L., Pan, Y., Chen, Y., et al. (2011). ODOFactor: a web server for deciphering olfactory coding. *Bioinformatics* 27, 2302–2303.
- Logan, J.G., Stanczyk, N.M., Hassanali, A., Kemei, J., Santana, A.E.G., Ribeiro, K.A.L., Pickett, J.A., and Mordue (Luntz), A.J. (2010). Arm-in-cage testing of natural human-derived mosquito repellents. *Malar. J.* 9, 1–10.
- Macwilliam, D., Kowalewski, J., Kumar, A., Pontrello, C., and Ray, A. (2018). Signaling mode of the broad-spectrum conserved CO2 receptor is one of the important determinants of odor valence in *Drosophila*. *Neuron* 97, 1153–1167.

- Marengo, L., Wang, R., McDougal, R., Olender, T., Twik, M., Bruford, E., Liu, X., Zhang, J., Lancet, D., Shepherd, G., and Crasto, C. (2016). ORDB, HORDE, ODORactor and other on-line knowledge resources of olfactory receptor-odorant interactions. Database 2016, baw132.
- McBride, C.S. (2016). Genes and odors underlying the recent evolution of mosquito preference for humans. *Curr. Biol.* 26, R41–R46.
- Mehr, Z.A., Rutledge, L.C., Buescher, M.D., Gupta, R.K., and Zakaria, M.M. (1990). Attraction of mosquitoes to diethyl methylbenzamide and ethyl hexanediol. *J. Am. Mosq. Control Assoc.* 6, 469–476.
- Morel, P. (2018). Gramm : grammar of graphics plotting in Matlab. *J. Open Source Softw.* 3, 1–4.
- Moriwaki, H., Tian, Y.S., Kawashita, N., and Takagi, T. (2018). Mordred: a molecular descriptor calculator. *J. Cheminform.* 10, 1–14.
- Mozūraitis, R., Hajkazemian, M., Zawada, J.W., Szymczak, J., Pålsson, K., Sekar, V., Biryukova, I., Friedländer, M.R., Koekemoer, L.L., Baird, J.K., et al. (2020). Male swarming aggregation pheromones increase female attraction and mating success among multiple African malaria vector mosquito species. *Nat. Ecol. Evol.* 4, 1395–1401.
- Münch, D., and Galizia, C.G. (2016). DoOR2 .0 - comprehensive mapping of *Drosophila melanogaster* odorant responses. *Sci. Rep.* 6, 1–9.
- Olender, T., Nativ, N., and Lancet, D. (2013). HORDE: comprehensive resource for olfactory receptor genomics. *Methods Mol. Biol.* 1003, 23–38.
- Pates, H.V., Takken, W., Stuke, K., and Curtis, C.F. (2001). Differential behaviour of *Anopheles gambiae sensu stricto* (Diptera: Culicidae) to human and cow odours in the laboratory. *Bull. Entomol. Res.* 91, 289–296.
- Price, G.D., Smith, N., and Carlson, D.A. (1979). The attraction of female mosquitoes (*Anopheles quadrimaculatus* SAY) to stored human levels of relative humidity, temperature, and carbon dioxide. *J. Chem. Ecol.* 5, 383–395.
- Raji, J.I., Melo, N., Castillo, J.S., Gonzalez, S., Saldana, V., Stensmyr, M.C., and DeGennaro, M. (2019). *Aedes aegypti* mosquitoes detect acidic volatiles found in human odor using the IR8a pathway. *Curr. Biol.* 29, 1253–1262.e7.
- Riabina, O., Task, D., Marr, E., Lin, C.C., Alford, R., O’Brochta, D.A., and Potter, C.J. (2016). Organization of olfactory centres in the malaria mosquito *Anopheles gambiae*. *Nat. Commun.* 7, 1–12.
- Saberi, M., and Seyed-Allaei, H. (2016). Odorant receptors of *Drosophila* are sensitive to the molecular volume of odorants. *Sci. Rep.* 6, 1–11.
- Sayin, S., Boehm, A.C., Kobler, J.M., de Backer, J.F., and GrunwaldKadow, I.C. (2018). Internal state dependent odor processing and perception—the role of neuromodulation in the fly olfactory system. *Front. Cell Neurosci.* 12, 1–17.
- Scott, J.W., Sherrill, L., Jiang, J., and Zhao, K. (2014). Tuning to odor solubility and sorption pattern in olfactory epithelial responses. *J. Neurosci.* 34, 2025–2036.
- Semmelhack, J.L., and Wang, J.W. (2009). Select *Drosophila glomeruli* mediate innate olfactory attraction and aversion. *Nature* 459, 218–223.
- Simonnet, M.M., Berthelot-Grosjean, M., and Grosjean, Y. (2014). Testing *Drosophila* Olfaction with a Y-maze Assay. *J. Vis. Exp.* <https://doi.org/10.3791/51241>.
- Spitzen, J., Spoor, C.W., Grieco, F., ter Braak, C., Beeuwkes, J., van Brugge, S.P., Kranenborg, S., Noldus, L.P.J.J., van Leeuwen, J.L., and Takken, W. (2013). A 3D analysis of flight behavior of *Anopheles gambiae sensu stricto* malaria mosquitoes in response to human odor and heat. *PLoS One* 8, 1–12.
- Su, C.Y., Menz, K., and Carlson, J.R. (2009). Olfactory perception: receptors, cells, and circuits. *Cell* 139, 45–59.
- Vinauger, C., Van Breugel, F., Locke, L.T., Tobin, K.K.S., Dickinson, M.H., Fairhall, A.L., Akbari, O.S., and Riffell, J.A. (2019). Visual-olfactory integration in the human disease vector mosquito *Aedes aegypti*. *Curr. Biol.* 29, 2509–2516.e5.
- Vogt, K., Zimmerman, D.M., Schlichting, M., Hernandez-Nunez, L., Qin, S., Malacon, K., Rosbash, M., Pehlevan, C., Cardona, A., and Samuel, A.D.T. (2021). Internal state configures olfactory behavior and early sensory processing in *Drosophila* larvae. *Sci. Adv.* 7, 1–11.
- Wang, G., Carey, A.F., Carlson, J.R., and Zwiebel, L.J. (2010). Molecular basis of odor coding in the malaria vector mosquito *Anopheles gambiae*. *Proc. Natl. Acad. Sci. U S A* 107, 4418–4423.
- Wang, B., Liu, Y., He, K., and Wang, G. (2016). Comparison of research methods for functional characterization of insect olfactory receptors. *Sci. Rep.* 6, 1–10.
- Yu, B.T., Ding, Y.M., and Mo, J.C. (2015). Behavioural response of female *Culex pipiens pallens* to common host plant volatiles and synthetic blends. *Parasit. Vectors* 8, 1–8.
- Zwicker, D., Murugan, A., and Brenner, M.P. (2016). Receptor arrays optimized for natural odor statistics. *Proc. Natl. Acad. Sci. U S A* 113, 5570–5575.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Data website	Github	https://neuralsystems.github.io/MORE/
Software and algorithms		
Code developed	Github	https://github.com/neuralsystems/MORE
MATLAB	MathWorks	https://www.mathworks.com/products/matlab.html
Python	Python Software Foundation	https://www.python.org
Gramm plotting toolbox	Morel (2018)	https://github.com/piermorel/gramm
WebPlotDigitizer	AnkitRohatgi	https://automeris.io/WebPlotDigitizer/
Physiochemical properties	Moriwaki et al. (2018)	https://github.com/mordred-descriptor/mordred
PubChemPy	PubChem	https://pubchempy.readthedocs.io/en/latest/guide/introduction.html

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Nitin Gupta (guptan@iitk.ac.in).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- All the data reported in this paper will be shared by the lead contact upon request.
- The code developed in this study can be accessed from the GitHub repository (<https://github.com/neuralsystems/MORE>).
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

METHOD DETAILS

Data extraction

Tabulated data from some papers were obtained by requesting the original authors ([Hallem and Carlson, 2006](#); [Hill et al., 2009](#); [Wang et al., 2010](#)). From some papers, when the data were not available directly, WebPlotDigitizer tool (provided by AnkitRohatgi) was used to manually extract the data from plots.

Different papers have reported odorant concentrations using many different formats such as vol/vol (V/V), wt/vol (W/V), molarity, parts per million (ppm), or weight/area (mg/cm²), which makes the comparisons difficult. Wherever possible, we converted the concentrations to common notations and units, as either fractions (V/V or W/V) or to g/ml (W/V); in case of dry odorant applied on a filter paper, we mentioned the amount of odorant use after setting concentration type as "Dry". To take a few examples of the conversions used: 1.11×10^{-5} M of lactic acid was converted to 0.000001 g/mL ([Braks et al., 2001](#)); 1 ppm was converted to 1 mg/L ([Islam et al., 2017a](#)); in one study, 0.025 mL of 0.01 mg/cm² odor was added on 6.6 cm² cloth, which was converted to equivalent W/V concentration in g/mL given by (0.000001 g/cm²) × (6.6 cm²)/(0.025 mL) ([Mehar et al., 1990](#)).

The EAG responses reported in some studies were not normalized ([Cork and Park, 1996](#); [Guha et al., 2014](#)), but in other studies were normalized with respect to a reference odor, such as 1-octen-3-ol ([Blackwell and Johnson, 2000](#); [Constantini et al., 2001](#)). To maintain uniformity, if the responses were not normalized, we

normalized them with 1-octen-3-ol, if present, or with another suitable reference odor depending on the dataset; the reference odor used is noted for each EAG data-point.

In OR electrophysiological measurements, some studies report the responses after subtracting the spontaneous firing rate (Hallem and Carlson, 2006), while some report the raw response without the background subtraction (Dobritsa et al., 2003). In the latter cases, we subtracted the response to the solvent (e.g., paraffin oil) from the odor responses of each OR.

Physicochemical properties

The properties of the odorants were obtained from the PubChem database using a MATLAB script (created by Vincent Scalfani). We could obtain a set of 28 properties by this automated approach. Some of the properties were missing for some of the odorants. To ensure that the curve-fitting in the subsequent steps can be done reliably, we analyzed only those properties which took at least ten different numerical values among our set of odorants, leaving us with the following 13 properties: Molecular Weight; Molecular Volume; Octanol-water Partition Coefficient (computationally predicted); Complexity (an indicator of the complexity of the molecular structure, calculated using the Bertz/Hendrickson/Ihlenfeldt formula); Conformer Count 3D (the number of conformers); Effective Rotor Count 3D (number of effective rotors); Feature Count 3D (total number of 3D features); Heavy Atom Count (number of non-hydrogen atoms); Rotatable Bond Count (number of rotatable bonds); Topological Polar Surface Area (estimate of the area that is polar); XStericQuadrupole3D (x component of the quadrupole moment); YStericQuadrupole3D (y component of the quadrupole moment); ZStericQuadrupole3D (z component of the quadrupole moment).

The response data were taken from experiments that used the concentration of 10^{-2} V/V or W/V, which was the most frequently used concentration in the dataset. To calculate the p values, we first shuffled the mapping between the odorant property and response for 1000 times and then calculated the p value as the fraction of times $\sigma_{\text{shuffle}} \leq \sigma_{\text{actual}}$ (where, σ_{shuffle} and σ_{actual} are the standard deviations of fitted Gaussian on the shuffled and actual data, respectively). For fitting the gaussian, we used the MATLAB 'fit' function with 'gauss1' argument.

Neural network model for OR response prediction

We obtained 1842 physicochemical descriptors from PubChem and Mordred (Moriwaki et al., 2018). From these descriptors, we first removed the descriptors with missing values or variance < 0.005 , leaving 645 descriptors. Next, we removed the descriptors with mutual correlation greater than 0.95 (Caballero-Vidal et al., 2020). This finally gave us 295 descriptors, which were used to represent odorants in the model. For predicting the OR responses to the odorants, we used the Deep Learning Toolbox in MATLAB with default parameters to implement a feedforward artificial neural network with three hidden layers, each having 50 neurons. The dataset including 112 odorants for each OR, which were randomly partitioned into a training set (70% of the odorants), a model validation set (15% of the odorants), and a test set (15% of the odorants). The model performance is reported using only the predictions on the test odorants. For comparison, we randomly shuffled the OR-odorant response matrix and used the shuffled responses as the control predictions.

Analyses of behavioral assays

The exact behavioral assay used in each study is mentioned with the corresponding data-points in the database. In all analyses (Figures 4, 5, and 6), only those data-points were used for which the odor concentration was known. The concentrations were rounded to the nearest power of 10. In the comparison of preference index and oviposition index, the two indices were compared only when both the indices were measured at the same concentration (after rounding). After applying this criterion, we had 11 data-points in *A. aegypti* (shown in Figure 5); species with less than 5 data-points each were not analyzed. In Figure 4, if multiple data-points from different studies were available for an odor for the same assay category (example: T-maze or non-T-maze) and the same concentration, they were averaged into a single data-point. In Figure S4, an odor was included only if it was tested in the compared species at the same concentration (if the same concentration was not available, the closest available concentration within ± 10 folds was taken). The correlations were analyzed for only those pairs of species for which at least 10 common odors were available. In Figure 6A and 6B, we considered an odor-concentration if at least two data-points were



available for that concentration with the same assay in the same species (they were averaged into a single data-point).

QUANTIFICATION AND STATISTICAL ANALYSIS

All the analyses were conducted in MATLAB. The Gramm plotting toolbox ([Morel, 2018](#)) was used to draw the plots. The non-parametric Wilcoxon signed rank test was used to calculate the p values of paired sample comparisons.