





Characterization of the chemoreceptor repertoire of a highly specialized fly with comparisons to other *Drosophila* species

Pedro Mesquita Fonseca¹, Lizandra Jaqueline Robe² , Tuane Letícia Carvalho² and Elgion Lucio Silva Loreto^{1,3} 

¹Universidade Federal do Rio Grande do Sul, Instituto de Biociências, Programa de Pós-Graduação em Genética e Biologia Molecular, Porto Alegre, RS, Brazil.

²Universidade Federal de Santa Maria, Centro de Ciências Naturais e Exatas, Programa de Pós-Graduação em Biodiversidade Animal, Santa Maria, RS, Brazil.

³Universidade Federal de Santa Maria, Centro de Ciências Naturais e Exatas, Departamento de Bioquímica e Biologia Molecular, Santa Maria, RS, Brazil.

Abstract

To explore the diversity of scenarios in nature, animals have evolved tools to interact with different environmental conditions. Chemoreceptors are an important interface component and among them, olfactory receptors (ORs) and gustatory receptors (GRs) can be used to find food and detect healthy resources. *Drosophila* is a model organism in many scientific fields, in part due to the diversity of species and niches they occupy. The contrast between generalists and specialists *Drosophila* species provides an important model for studying the evolution of chemoreception. Here, we compare the repertoire of chemoreceptors of different species of *Drosophila* with that of *D. incompta*, a highly specialized species whose ecology is restricted to *Cestrum* flowers, after reporting the preferences of *D. incompta* to the odor of *Cestrum* flowers in olfactory tests. We found evidence that the chemoreceptor repertoire in *D. incompta* is smaller than that presented by species in the *Sophophora* subgenus. Similar patterns were found in other non-*Sophophora* species, suggesting the presence of underlying phylogenetic trends. Nevertheless, we also found autapomorphic gene losses and detected some genes that appear to be under positive selection in *D. incompta*, suggesting that the specific lifestyle of these flies may have shaped the evolution of individual genes in each of these gene families.

Keywords: Olfaction, taste, odorant receptor, gustatory receptor.

Received: December 28, 2022; Accepted: May 07, 2024.

Introduction

Nature consists of many environments with different conditions and particularities. Animals living in each environment are in constant contact with many chemical compounds. Some of these compounds come from other living forms and are received and interpreted for many different purposes, such as finding sexual partners (Wyatt, 2003; Davis, 2007; Kurtovic *et al.*, 2007), identifying predators or parasites (Ebrahim *et al.*, 2015; Joseph and Carlson, 2015), or even finding and tasting food (Dahanukar *et al.*, 2005; Hallem *et al.*, 2006). Regardless of the nature of the chemical signals, animals usually interpret them through chemoreceptors present in their cells. Because chemoreceptors are directly involved in this first contact with nature, they are subject to differential selection pressures during animal evolution (Whiteman and Pierce, 2008). In this sense, chemoreceptor genes generally evolve fast, leading to great diversification in genic repertoire (McBride, 2007; McBride and Arguello, 2007; Vieira and Rozas, 2011; Cande *et al.*, 2013). These patterns generally lead to an association between odor detection and individual ecological needs (Carey *et al.*, 2010).

Since the early 20th century, many *Drosophila* species have been used as model organisms in various scientific fields, among which we can highlight *Drosophila melanogaster* Meigen, 1830 (Markow, 2015). Nevertheless, the genus includes 1,644 described species with many different traits, niches, and distributions (O'Grady and DeSalle, 2018; TaxoDros). Many of these species have provided excellent models for evolutionary genomics, and there is a wealth of phylogenetic information spanning hundreds of species and at least thirty well-annotated genomes (*Drosophila* 12 Genomes Consortium *et al.*, 2007; Chen *et al.*, 2014; O'Grady and DeSalle, 2018). *Drosophila* also provides several advantages while studying adaptations associated with different ecological habits, as it includes species with contrasting behaviors and niche breaths. Among these species we can identify generalists such as *D. melanogaster*, *Drosophila ananassae* Doleschall (1858), and *Drosophila simulans* Sturtevant (1919) and other specialists such as *Drosophila grimshawi* Oldenberg, 1914, *Drosophila erecta* Tsacas and Lachaise (1974) and *Drosophila sechellia* Tsacas and Bächli, 1981. Generalist habits allow exploring a variety of resources for feeding and oviposition, making such species resilient under conditions of limited resource availability or competition (Anholt, 2020). Conversely, specialists benefit from low competition for food or oviposition sites once they can overcome toxic or extreme niches (Anholt, 2020).

Among the different families of chemoreceptor genes, those coding for gustatory receptors (GR) and olfactory receptors (OR) are responsible for the detection of flavors or odorants, respectively (Clyne *et al.*, 1999; Gao and Chess, 1999; Scott *et al.*, 2001; Vosshall and Stocker, 2007; Martin *et al.*, 2013). In *D. melanogaster*, each of these two gene families harbors 60 genes (Robertson *et al.*, 2003). These genes are commonly involved in adaptations to different ecological contexts (Depetris-Chauvin *et al.*, 2015; Diaz *et al.*, 2018). Such a pattern has been demonstrated, for example, for the pestiferous species *Drosophila suzukii* Matsumura that differs from other species of the *D. melanogaster* group due to its characteristic of exploring fresh fruit rather than rotting ones, which is reflected in the OR and GR repertoire (Hickner *et al.*, 2016). Moreover, specialist *Drosophila* species can present a fivefold higher dN/dS ratio for OR and GR genes (McBride, 2007) and may also present faster rates of gene loss for these two gene families (McBride and Arguello, 2007). Nevertheless, according to Gardiner *et al.* (2008), endemism rather than niche specialization may account for much of this straightforward chemosensory gene loss.

Within *Drosophila*, the *D. flavopilosa* species group draws attention for its close association with flowers of *Cestrum* (Solanaceae), using this substrate as a unique resource for oviposition, larval development, and adult feeding (Brncic, 1966; Robe *et al.*, 2013). This group consists of 18 species (TaxoDros) divided into two subgroups and is considered part of the *virilis-repleta* radiation (Throckmorton, 1975), which is currently considered a subgenus of the *Drosophila* genus (the subgenus *Siphlodora*) (Yassin, 2013). Nevertheless, the exact phylogenetic position of the group within the subgenus remains controversial, as it can be considered either a sister group of the *D. annulimana* species group (Robe *et al.*, 2010) or a sister group of the *D. virilis* or *D. repleta* species groups (de Ré *et al.*, 2017). The flies of the *flavopilosa* group share some strict morphological and physiological adaptations for the use of *Cestrum* flowers: the pale yellow color of the body, which is cryptic in many *Cestrum* flowers; the medium to small size of the flies; oviposition at advanced stages of embryonic development, which matches with the ephemeral nature of *Cestrum* flowers; and the robust and full sting characteristic of ovipositors used to scarify the flower surface (Wheeler *et al.*, 1962; Pipkin *et al.*, 1966; Brncic, 1983; Ludwig *et al.*, 2002; Robe *et al.*, 2013). Although the occurrence of this group covers a large portion of the Neotropical region, most species are endemic to small patches (Robe *et al.*, 2013). The most widespread species in the group is *D. incompta*, which occurs from Argentina to Mexico (TaxoDros).

The large number of chemoreceptors that can be found in *Drosophila* possibly enables the exploration of different resources, providing means to the establishment of different niches. To relate the preference of a species to its specific resource, choice tests are usually performed comparing different substrates. In these tests, *Drosophila* usually prefers citric substrates for oviposition (Dweck *et al.*, 2013), but this choice can vary depending on the niche of each species.

Mansourian *et al.* (2018) demonstrated that *D. melanogaster* specimens are seasonal specialist of marula fruit (*Sclerocarya birrea*), preferring volatile marula odors over orange scents, although they preferred orange when tested with other fruit scents such as banana. Host switching between populations can also be assessed by comparing preferences for different odors. For example, for *D. mojavensis*, a cactophilic species that feeds on and breeds on various cactus species, population-specific preferences were previously detected (Date *et al.*, 2013).

The aim of this study is to contribute to the knowledge of the patterns of molecular evolution associated with the chemoreceptor repertoire in different ecological contexts. To this end, we first examined the olfactory preferences of *D. incompta* to test if there are preferences for *Cestrum* flowers extracts over other odors. We then listed OR and GR genes that are part of the chemoreceptor repertoire in this species and compared them to homologous sequences in other *Drosophila* species to gain insight into the forces that might be shaping the evolution of chemoreceptor genes in specialized species.

Material and Methods

Sampling

Specimens of the *Drosophila flavopilosa* group were obtained from flowers of *Cestrum strigilatum* (Solanaceae) collected in the city of Santa Maria, southern Brazil (-29.710553 latitude and -53.717070 longitude). The sampled flowers were kept in the laboratory until the adult flies emerged. *Drosophila incompta* was distinguished from the other species of the group by its external morphology and male genitalia patterns, according to Wheeler *et al.* (1962).

Olfactory choice tests

To test the flies' preference for different odors, we performed an olfactory choice test (Table S1) with adult flies of *D. incompta* collected from the field, as described above. The tests were also performed with adults of *D. melanogaster* from the Oregon-R strain maintained in our laboratory, as a control to test if the olfactometer was working well. Flies were tested up to two days after emergence, until the total number of 10 flies was achieved per test. The elected odors were: 1) extract of *Cestrum strigilatum* L. flowers [1 ml of a solution of non-rotting flowers, prepared with five entire flowers (approximately 5 g) macerated in 5 ml of distilled water]; 2) extract of *Brunfelsia uniflora* (pohl) (Solanaceae) [1 ml of a solution of non-rotting flowers, prepared with two entire flowers (approximately 5 g) macerated in 5 ml of distilled water]; 3) extract of banana fruits (1 ml of a solution prepared with 5 g of a mature banana macerated in 5 ml of distilled water); and 4) extract of orange fruits (1 ml of a solution prepared with 5 g of a mature orange with peel macerated in 5 ml of distilled water). These odors were chosen (considering the putatively strict adaptation of *D. incompta* to *Cestrum* spp. (Hofmann and Napp, 1984; Santos and Vilela, 2005), against the general profile of exploring different fermenting fruits presented by frugivorous *Drosophila* species (Sevenster and Van Alphen, 1993; Markow, 2015). Conversely, *Brunfelsia*

flowers were elected because they blossom concomitantly with *Cestrum* and have records of interaction with other *Drosophila* species (Frota-Pessoa, 1952; Malogolowkin, 1953; Sepel *et al.*, 2000; Schmitz and Valente, 2019).

In each case, the extracts were combined in pairs and dropped onto a new piece of cotton (Figure 1). Ten flies of the same sex were placed in the center of the olfactometer (following Fuyama, 1976) with the gates closed for two minutes for acclimation. After, gates were opened, and flies were allowed to choose a side. After five minutes, flies located on each side of the artifact were counted. All tests were conducted in a dark room with only red light available, in the morning. Each test was repeated at least six times per treatment, considering replicates performed either for males or with females, as both sexes showed similar patterns. Analysis of variance (ANOVA) was performed to assess the significance of the preferences.

Molecular biology procedures

Total DNA was isolated from 20 males of *D. incompta* using a NucleoSpin Tissue XS kit (Macherey-Nagel, Düren, Germany) according to the manufacturer's protocol. Only males were employed at this stage due to the cryptic nature of *D. incompta* regarding other closed related species, which hampers females' identification (Wheeler *et al.*, 1962; Robe *et al.*, 2013). Two approaches were used to obtain shotgun sequences of the whole genome. First, 100-bp single-end reads were obtained using a Solexa-Illumina HiSeq 2000 Next Generation Sequencing (NGS) instrument (Fasteris DNA Sequencing Service). Furthermore, we also generated 100-bp paired end reads in a Solexa-Illumina HiSeq 2000 instrument (Illumina Inc., San Diego, USA) at Macrogen Sequencing Service (Korea) using the provider's protocols.

Other 20 adult males of *D. incompta* were directly collected from the medium with sampled flowers up to two hours after eclosion and further employed to isolate total

adult RNA using TRIzol reagent (ThermoFisher Scientific), according to the protocol of Rio *et al.* (2010). The mRNA was isolated using the mRNA Dynabeads C kit (Life Technologies) and libraries were prepared using the Seq RNA Total Ion V2 kit (ThermoFisher Scientific). Sequencing was performed using a Ion S5 sequencer (ThermoFisher Scientific).

The genome of *D. incompta* was assembled using a hybrid assembly approach for both paired-end and single-end libraries in hybridSPAdes (Antipov *et al.*, 2016). The performance of the de novo assembly was evaluated using Quast (Gurevich *et al.*, 2013) and Busco 5.4.6 software (Simão *et al.*, 2015). In the last case, the Diptera OrthoDBv.10 (odb10) dataset was employed, with default parameters (Table S2). Transcriptome de novo assembly was performed using Trinity (Grabherr *et al.*, 2011) under default parameters in the Galaxy platform (Afgan *et al.*, 2018).

Predicted coding sequences (cds) and protein genes for both genome and transcriptome drafts were identified using the Augustus 3.4.0 software (Stanke *et al.*, 2006). Both cds datasets were merged, and redundant sequences were removed using the cd-hit-2d program (Fu *et al.*, 2012). This resulted in a non-redundant hybrid annotation of predicted cds, including genome and transcriptome sequences.

Recovery of OR and GR gene sequences in *D. incompta*

To retrieve sequences related to OR and GR chemoreceptors in the genome of *D. incompta*, at first, the repertoire of olfactory and gustatory receptor genes of *D. melanogaster* (Robertson *et al.*, 2003) and *D. virilis* (*Drosophila* 12 Genomes Consortium *et al.*, 2007) were downloaded from FlyBase (Thurmond *et al.*, 2019), totaling 121 and 103 different genes sequences, respectively. *Drosophila melanogaster* was chosen because it is a model organism that has a well-annotated and well-studied repertoire of chemoreceptors. In turn, *D. virilis* entered as a query in

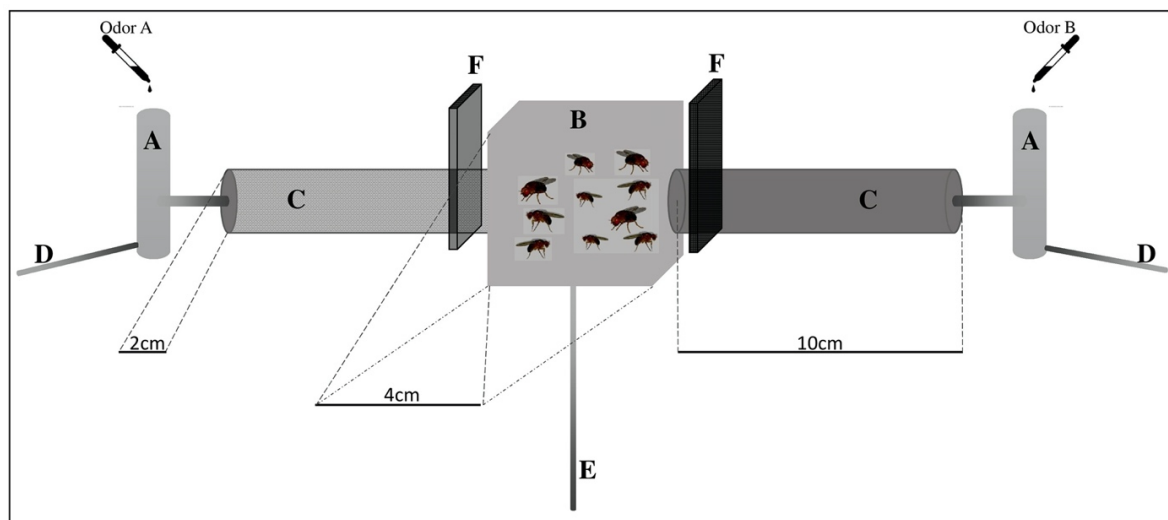


Figure 1 – Olfactometer model used to perform the olfactory preference tests. Legend: A, Extracts scent vessels; B, Flies waiting box; C, Preference tubes; D, Air inlet pipes; E, Air suction pipe; and F, Gates separating the flies from scents.

our searches because it is closely related to *D. incompta* (Robe *et al.*, 2010; De Ré *et al.*, 2017). These sequences were arranged into two separate matrices subdivided by chemoreceptor gene family.

The pipeline employed in the InsectOR tool (Karpe *et al.*, 2021) was then used to predict and validate OR and GR sequences in *D. incompta* genome. For this task, Exonerate was first used to align the protein sequences recovered for *D. melanogaster* and *D. virilis* against the draft genome of *D. incompta* assembled by HybridSpades. These alignment files, the draft genome, and the protein queries were then submitted as input to the InsectOR web server. As a result, detailed information on each gene prediction and cds and protein fasta sequences were obtained. Finally, for an exhaustive approach, three additional strategies were employed: 1) OR and GR gene sequences that were not found with InsectOR were searched against the Augustus annotation using protein sequences of *D. melanogaster* and *D. virilis* as queries in BLASTp and tBLASTn searches; 2) these sequences were also used as seed against the raw reads of *D. incompta* genome and transcriptome, through the use of aTRAM 2.0 software (Allen *et al.*, 2018), which simultaneously searches and assembles orthologous gene sequences; 3) eggNOG-mapper (Huerta-Cepas *et al.*, 2017) was employed to functionally annotate the entire genome of *D. incompta*, using the paired-end Solexa-Illumina raw reads with default parameters.

Repertoire of OR and GR gene sequences in other *Drosophila* species

To recover OR and GR gene sequences from other species, we relied on the repertoires found by Gardiner *et al.* (2008) for the first 12 *Drosophila* sequenced genomes (*Drosophila* 12 Genomes Consortium *et al.*, 2007) (*D. ananassae*, *D. erecta*, *D. grimshawi*, *D. mojavensis*, *D. melanogaster*, *D. persimilis*, *D. pseudoobscura*, *D. sechellia*, *D. simulans*, *D. virilis*, *D. willistoni*, and *D. yakuba*). To complement the matrices and provide a broader glimpse about species with different levels of niche specialization, we also retrieved OR and GR genes from the genomes of other 19 *Drosophila* species (*D. arizonae*, *D. biarmipes*, *D. bipunctinata*, *D. busckii*, *D. buzzatii*, *D. elegans*, *D. eugracilis*, *D. ficusphila*, *D. hydei*, *D. kikkawai*, *D. mauritiana*, *D. miranda*, *D. navojoa*, *D. novamexicana*, *D. obscura*, *D. rhopaloa*, *D. serrate*, *D. suzukii*, and *D. takahashii*) and outgroups (see below). These searches were performed through BLASTn in the NCBI database using OR and GR genes of *D. virilis* and *D. melanogaster* as queries. In each case, only the main copy of each gene was considered, and duplicated copies were not counted. Conversely, incomplete copies were eventually included in further analyses if they were the only available copies of the respective genes. We considered this approach a strategy to uncover the general evolutionary pattern of the two different gene families since it is not strongly affected by recent duplications or gene losses in other *Drosophila* species. Moreover, this approach is more robust to problems related to sequencing or assembling strategies.

Orthology assignments

To correctly assign orthologous relationships, the whole set of amino acid sequences of OR and GR gene families isolated for *D. melanogaster* and *D. virilis* were aligned against the putative orthologous copies retrieved for *D. incompta*. This alignment was performed under a PAM matrix, using the Clustal W algorithm (Thompson *et al.*, 1994), as available in Mega 7.0.26 software (Kumar *et al.*, 2018). OR and GR matrixes were then employed to reconstruct a phylogenetic tree, under maximum likelihood (ML) in the IQTREE v.1.6.12 software (Nguyen *et al.*, 2015), using 10,000 ultra-fast bootstrap replicates. The ML tree was finally visualized in FigTree v1.4 (2018) and rooted at the midpoint.

Evolutionary tests

For each individual gene, matrices were constructed containing all homologous sequences found in the different *Drosophila* species and outgroups (Table S3). *Scaptodrosophila lebanonensis* was generally used as an outgroup except in cases where no orthologous sequences were found for this species. In these cases, another species closely related to *Drosophilidae* for which orthologous sequences were available in NCBI was used as the outgroup: *Calliphora stygia*, *Ctenopseustis obliquana*, or *Rhopalosiphum maidis* for ORs; and *Bactrocera latifrons*, *Musca domestica*, or *Ceratitis capitata* for GRs. Codon-based alignments were then performed individually for each gene of the GR or OR gene families for which orthologous sequences of at least 441 bp were found for *D. incompta* using MACSE (Ranwez *et al.*, 2011).

Subsequently, two different selection tests were performed to evaluate the presence of negative or positive selection in the OR and GR genes of *D. incompta*: A) the adaptive branch-site REL test for episodic diversification (aBSREL; Smith *et al.*, 2015) was performed to test whether the branch leading to *D. incompta* evolved under positive selection; and B) the fixed effects likelihood test (FEL; Pond and Frost, 2005) was used to derive the nonsynonymous substitution (α) and synonymous substitution (β) rates, testing whether alpha is significantly higher than beta at each site in the sequence of *D. incompta*. The latter test was applied only to matrices that showed a signal of positive selection for *D. incompta* genes in the aBSREL test. In both cases, individual gene codon alignments were submitted to the Datamonkey web platform (Weaver *et al.*, 2018), and analysis were performed with default parameters.

Results

Olfactory preference tests

Despite exploring a limited variety of odors, the olfactory preference tests showed a clear preference of *D. incompta* for *Cestrum* flowers compared to fermented banana extract ($p < 0.0001$). *Drosophila incompta* also preferred *Cestrum* flowers compared to fermented orange extract ($p < 0.0001$) and *Brunfelsia uniflora* (Pohl.) flowers ($p = 0.0005$). For *D. melanogaster*, the test of preference for *Cestrum* versus

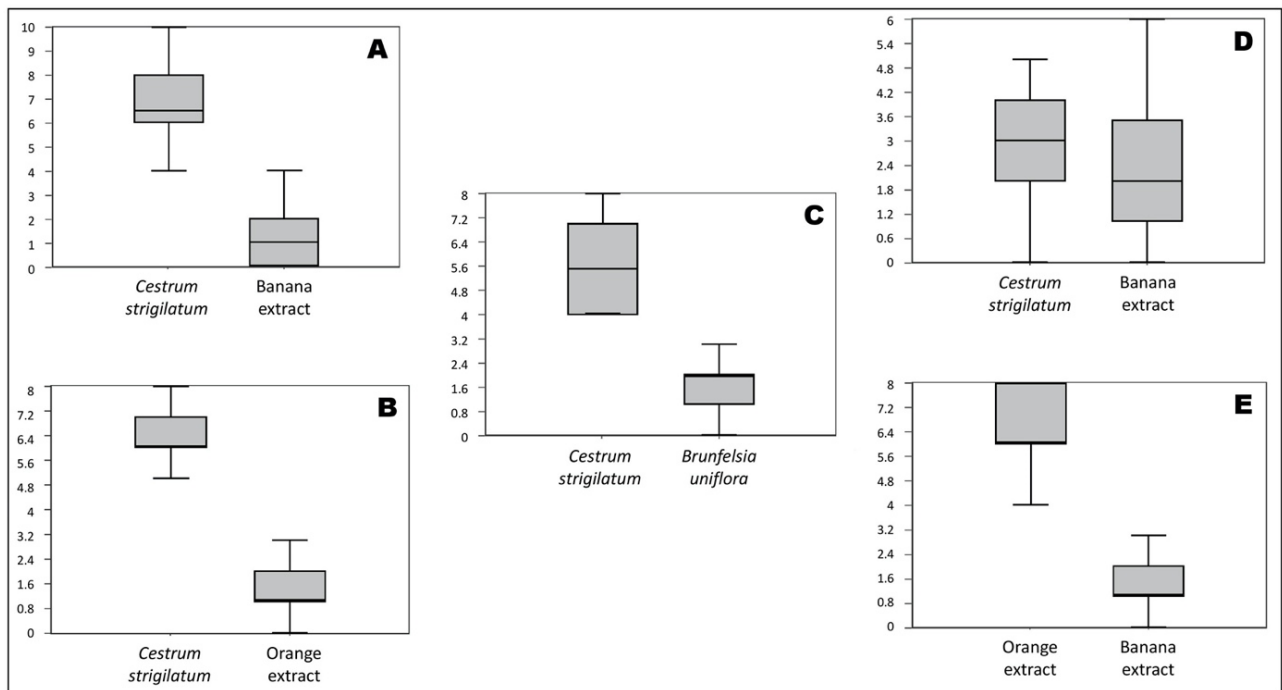


Figure 2 – Results of ANOVAs performed to test the preference of *D. incompta* (**A, B, and C**) and *D. melanogaster* (**D and E**) for each pair of extract scents. **A and D**: *Cestrum strigilatum* against banana. $F\text{-Stat} = 267.0286, p = 1.4772e-22$ and $F\text{-Stat} = 0.03866, p = 0.84555$, respectively; **B**: *C. strigilatum* against orange. $F\text{-Stat} = 80.2002, p = 1.246e-7$; **C**: *C. strigilatum* against *Brunfelsia uniflora*. $F\text{-Stat} = 25.71407, p = 0.0005$; **E**: Orange against banana. $F\text{-Stat} = 74.88571, p = 1.9737e-7$. Y axis represents the percentage of the preference of flies for both compared extracts at every paired comparison.

fermented banana extract showed no statistically significant difference. On the other hand, our results corroborated the results found by Mansourian *et al.* (2018) about the significant preference of *D. melanogaster* for orange over banana extracts (Figure 2, Table S1), showing that our olfactometer provides an accurate description of smell perception and preference for the tested extracts in the target species.

Genome-wide analyses

The *de novo* genome assembly on hybridSPAdes resulted in 187,091 contigs from 289,083,143 raw reads (including both, single and paired end sequences), which resulted in 2,437 matches in Busco (Table S2). This was contrasted by the 11,209 contigs recovered by Trinity for the transcriptome. The hybrid dataset including genome and transcriptome nonredundant cds recovered by Augustus resulted in 2,660 matches in Busco (Table S2).

Repertoire of genes related to OR and GR gene families

After confirming orthology in a ML tree reconstructed for each family (Figs. 1S and 2S), we recovered a total of 84 sequences related to different chemoreceptors genes in *D. incompta*, of which 40 and 44 were identified as GRs and ORs, respectively (Tables 1 and 2). For ORs, 42 were recovered by InsectOR, and 11 and 37 were found by similarity searches using *D. melanogaster* and *D. virilis* genes as queries, respectively (Table S4). For GRs, all 40 genes were recovered by InsectOR, whereas 6 and 37 were found by similarity searches using *D. melanogaster* and *D. virilis* genes as queries, respectively (Table S4). Among the genes not recovered by

InsectOR, EggNog mapper and aTRAM were able to recover incomplete copies of two ORs.

Notwithstanding, results for the other *Drosophila* species were based on alternative methodologies, the number of chemoreceptor genes associated with ORs genes varied from 38 (in *D. busckii*) to 60 (in *D. melanogaster*, *D. sechellia* and *D. simulans*) (Table 1, Table S5). GRs varied from minimum values of 34 (in *D. busckii*) to maximum values of 60 (in *D. melanogaster*, *D. sechellia*, and *D. simulans*).

Regarding the distribution of gene losses or gains, evidence of phylogenetic signal was found for both ORs and GRs (Figures 3 and 4). In this context, it was interesting to see that the number of sequences found for both GR and OR was generally lower in non-*Sophophora* species. Nevertheless, two specific gene losses that are either autapomorphic or homoplastic for *D. incompta* were found for the genes ORs *Or1a*, and *Or94b* (Fig. 3).

Selection tests

Of the 84 chemoreceptor genes found for *D. incompta*, 81 had a continuous coding sequence of at least 441 bp (147 aa) and were included in further evolutionary analysis. Three ORs were not included due to their fragmentary sequences.

Of the 81 chemoreceptors matrices used for selection tests, six presented signals of positive selection on the aBSREL approach (Table 2): OR19a, OR42a, OR63a, GR57a, GR64e and GR77a ($p < 0.05$). These six matrices were also tested on FEL analysis and presented signals of positive selection ranging from one site for GR64e to 17 sites for GR63a, according to the following distribution: four sites for GR77a, five sites for OR19a, six sites for GR57a and eight sites for OR42a.

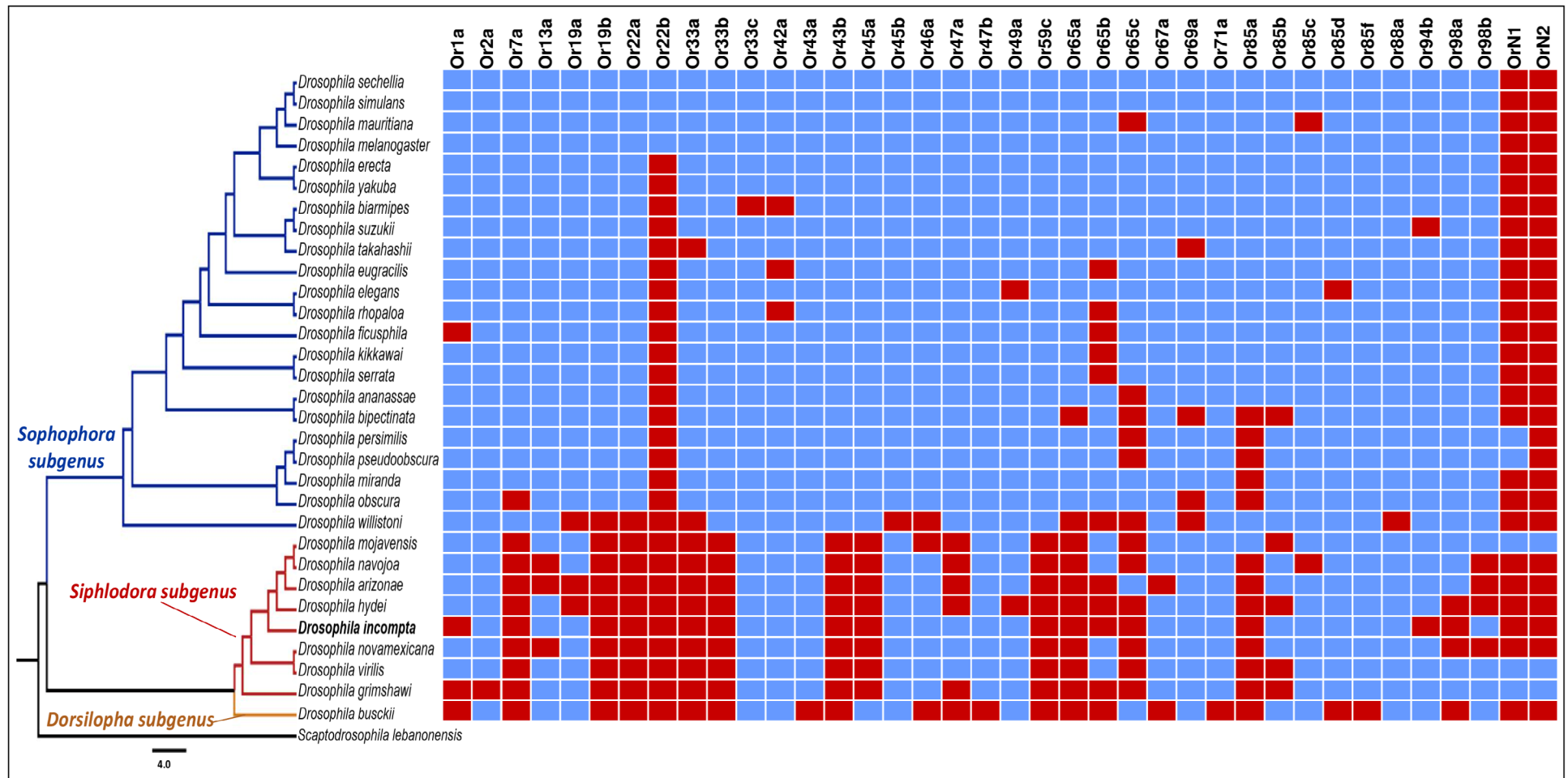


Figure 3 – Presence (blue boxes) and absence (red boxes) of 38 olfactory genes (*OR*) analyzed across the *Drosophila* phylogeny, modified from Robe *et al.* (2010), Yang *et al.* (2012), Yassin (2013), Hjelmen and Johnston (2017), and O’Grady and DeSalle (2018). Orthologues that were found in all analyzed species were not depicted on this graph.

Table 1 – Number of olfactory receptors (OR) and gustatory receptors (GR) related genes retrieved for *D. incompta* and each of the other 30 *Drosophila* species.

Species	Number of OR	Number of GR
<i>D. sechellia</i>	60	60
<i>D. simulans</i>	60	60
<i>D. mauritiana</i>	58	57
<i>D. melanogaster</i>	60	60
<i>D. erecta</i>	59	56
<i>D. yakuba</i>	59	59
<i>D. biarmipes</i>	57	55
<i>D. suzukii</i>	58	55
<i>D. takahashii</i>	57	56
<i>D. eugracilis</i>	57	56
<i>D. elegans</i>	57	52
<i>D. rhopaloa</i>	57	55
<i>D. ficusphila</i>	57	51
<i>D. kikkawai</i>	58	51
<i>D. serrata</i>	58	51
<i>D. ananassae</i>	58	54
<i>D. bipectinata</i>	54	47
<i>D. persimilis</i>	58	52
<i>D. pseudoobscura</i>	58	52
<i>D. miranda</i>	58	50
<i>D. obscura</i>	56	48
<i>D. willistoni</i>	48	47
<i>D. mojavenensis</i>	49	43
<i>D. navojoa</i>	45	40
<i>D. arizonae</i>	43	38
<i>D. hydei</i>	41	38
<i>D. incompta</i>	44	40
<i>D. novamexicana</i>	46	38
<i>D. virilis</i>	50	40
<i>D. grimshawi</i>	46	40
<i>D. busckii</i>	38	34

Note – For each species, the total number of orthologous sequences was considered, disregarding duplications. *Drosophila melanogaster*, *D. simulans*, *D. sechellia*, *D. yakuba*, *D. erecta*, *D. ananassae*, *D. pseudoobscura*, *D. persimilis*, *D. mojavenensis*, *D. virilis*, and *D. grimshawi* had their total numbers of OR and GR based on previous works (Guo and Kim, 2007; Nozawa and Nei, 2007; Gardiner *et al.*, 2008) in addition to homology searches.

Table 2 – Repertoire of olfactory receptors (OR) and gustatory receptors (GR) found in *D. incompta*, with their respective CDS fragment size and evidence of positive selection.

OR	CDS Fragment size recovered for <i>D. incompta</i> (bp)	Signals of positive selection for <i>D. incompta</i> according to aBSREL ($p \leq 0.05$)	Number of sites with evidence of positive selection according to FEL ($p < 0.05$)	GR	CDS Fragment size recovered for <i>D. incompta</i> (bp)	Signals of positive selection for <i>D. incompta</i> according to aBSREL ($p \leq 0.05$)	Number of sites with evidence of positive selection according to FEL ($p < 0.1$)
<i>Or2a</i>	876	–	–	<i>Gr2a</i>	1074 *	–	–
<i>Or9a</i>	1152 *	–	–	<i>Gr5a</i>	648	–	–
<i>Or10a</i>	1212 *	–	–	<i>Gr8a</i>	690	–	–
<i>Or13a</i>	678	–	–	<i>Gr9a</i>	1092 *	–	–
<i>Or19a</i>	1092	Yes ($p = 0.0005$)	5 ($p < 0.0450$)	<i>Gr10a</i>	1080	–	–
<i>Or22c</i>	1050	–	–	<i>Gr21a</i>	1113 *	–	–

Table 2 – Cont.

<i>OR</i>	CDS Fragment size recovered for <i>D. incompta</i> (bp)	Signals of positive selection for <i>D. incompta</i> according to aBSREL ($p \leq 0.05$)	Number of sites with evidence of positive selection according to FEL ($p < 0.05$)	<i>GR</i>	CDS Fragment size recovered for <i>D. incompta</i> (bp)	Signals of positive selection for <i>D. incompta</i> according to aBSREL ($p \leq 0.05$)	Number of sites with evidence of positive selection according to FEL ($p < 0.1$)
<i>Or23a</i>	834	–	–	<i>Gr22e</i>	1050	–	–
<i>Or24a</i>	1057	–	–	<i>Gr23a</i>	778	–	–
<i>Or30a</i>	600	–	–	<i>Gr28a</i>	753	–	–
<i>Or33c</i>	1146 *	–	–	<i>Gr28b</i>	564	–	–
<i>Or35a</i>	1227 *	–	–	<i>Gr32a</i>	637	–	–
<i>Or42a</i>	1158 *	Yes ($p = 0.0067$)	8 ($p < 0.0402$)	<i>Gr33a</i>	1442 *	–	–
<i>Or42b</i>	1194 *	–	–	<i>Gr39a</i>	609	–	–
<i>Or43a</i>	1119	–	–	<i>Gr39b</i>	933	–	–
<i>Or45b</i>	1179 *	–	–	<i>Gr43a</i>	441	–	–
<i>Or46a</i>	1158 *	–	–	<i>Gr47b</i>	1101	–	–
<i>Or47a</i>	1164	–	–	<i>Gr57a</i>	1245 *	Yes ($p = 0.0000$)	6 ($p < 0.0326$)
<i>Or47b</i>	945	–	–	<i>Gr58b</i>	696	–	–
<i>Or49a</i>	1212	–	–	<i>Gr58c</i>	1071	–	–
<i>Or49b</i>	1113 *	–	–	<i>Gr59b</i>	459	–	–
<i>Or56a</i>	1101 *	–	–	<i>Gr59d</i>	1023	–	–
<i>Or59a</i>	981	–	–	<i>Gr59e</i>	864	–	–
<i>Or59b</i>	1194 *	–	–	<i>Gr59f</i>	1152	–	–
<i>Or63a</i>	1266 *	Yes ($p = 0.0000$)	17 ($p < 0.0446$)	<i>Gr61a</i>	1122	–	–
<i>Or67a</i>	1212 *	–	–	<i>Gr63a</i>	1428 *	–	–
<i>Or67b</i>	1032 *	–	–	<i>Gr64a</i>	1191	–	–
<i>Or67c</i>	1212 *	–	–	<i>Gr64b</i>	831	–	–
<i>Or67d</i>	1170	–	–	<i>Gr64c</i>	999	–	–
<i>Or69a</i>	1143 *	–	–	<i>Gr64e</i>	1071	Yes ($p = 0.040$)	1 ($p = 0.0457$)
<i>Or71a</i>	1107 *	–	–	<i>Gr64f</i>	1059	–	–
<i>Or74a</i>	1215 *	–	–	<i>Gr66a</i>	753	–	–
<i>Or82a</i>	1101 *	–	–	<i>Gr68a</i>	703	–	–
<i>Or83a</i>	1320 *	–	–	<i>Gr77a</i>	1257 *	Yes ($p = 0.0000$)	4 ($p < 0.0415$)
<i>Or83c</i>	999	–	–	<i>Gr89a</i>	1095 *	–	–
<i>Or85b</i>	496	–	–	<i>Gr93a</i>	1299	–	–
<i>Or85c</i>	1110	–	–	<i>Gr93c</i>	1170 *	–	–
<i>Or85d</i>	564	–	–	<i>Gr94a</i>	1185 *	–	–
<i>Or85e</i>	1326	–	–	<i>Gr97a</i>	1207	–	–
<i>Or85f</i>	1203	–	–	<i>Gr98a</i>	855	–	–
<i>Or88a</i>	1203	–	–	<i>Gr98b</i>	747	–	–
<i>Or92a</i>	807	–	–	–	–	–	–
<i>Or94a</i>	1164	–	–	–	–	–	–
<i>Or98b</i>	927	–	–	–	–	–	–
<i>Orco</i>	1029	–	–	–	–	–	–

Note – The * symbol after the size number indicates a complete CDS recovered. Blank values on the selection tests columns correspond to genes for which we could not mount genes matrices due to the quality of the sequences. Genes with no significant signal of positive selection present the dash signal (-) on the selection tests columns.

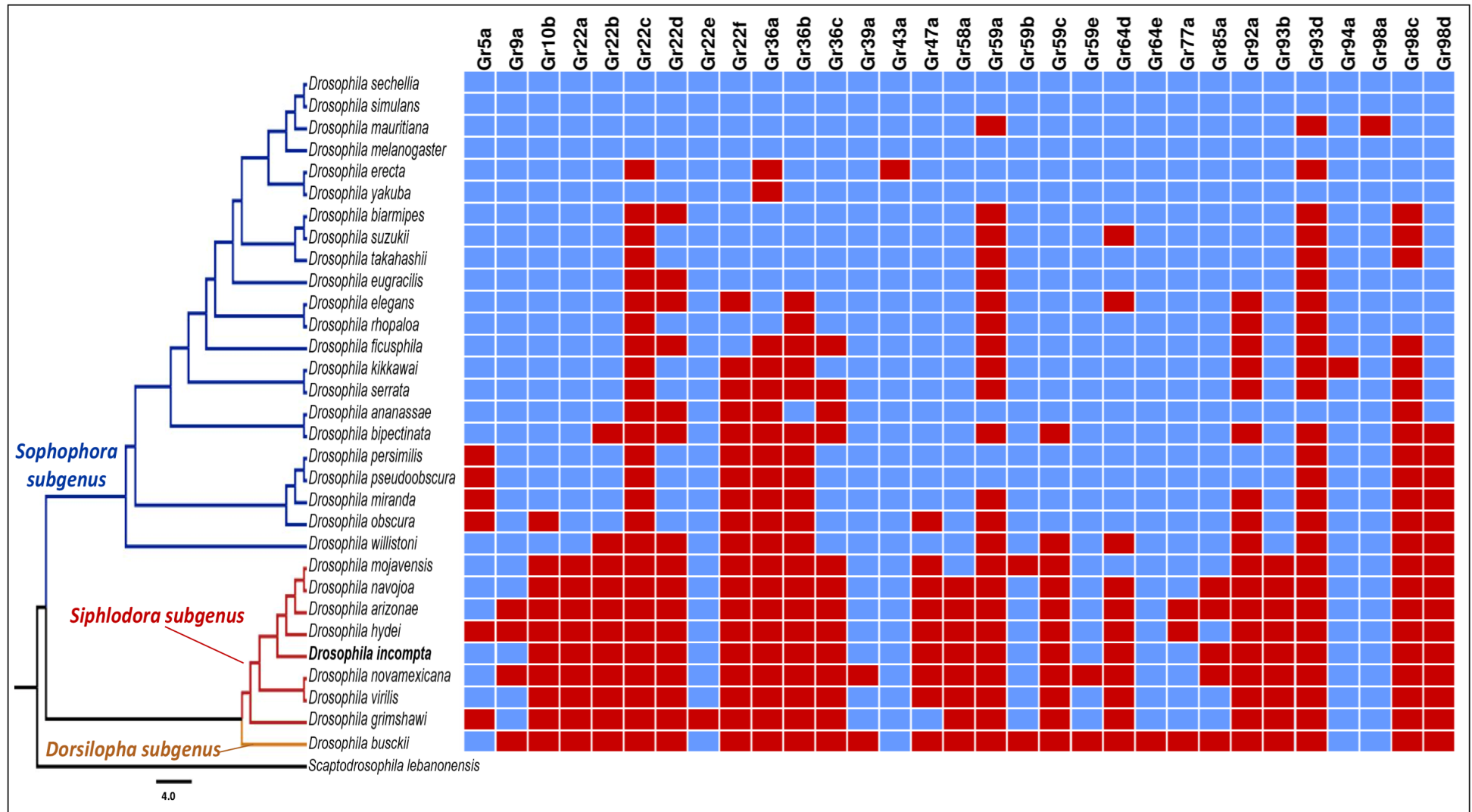


Figure 4 – Presence (blue boxes) and absence (red boxes) of 31 gustatory genes (**GR**) analyzed across the *Drosophila* phylogeny, modified from Robe *et al.* (2010), Yang *et al.* (2012), Yassin (2013), Hjelmen and Johnston (2017), and O’Grady and DeSalle (2018). Orthologs that were found in all analyzed species were not depicted on this graph.

Discussion

Drosophila has been widely used as a model to study the forces shaping the evolution of chemoreceptor gene families in different ecological contexts. Nevertheless, the role that selection and genetic drift have played in the evolution of these genes in specialized species is still controversial (McBride, 2007; McBride and Arguello, 2007; Gardiner *et al.*, 2008). This study contributes to the understanding of the patterns of molecular evolution exhibited by the OR and GR gene families in a species with a strict niche of the *D. flavopilosa* group, *D. incompta*, and provides several insights into understanding the general patterns of the entire *Drosophila* genus.

Molecular evolution of OR and GR gene families in *D. incompta*

Although it has been previously demonstrated that *D. incompta* can only explore *Cestrum* flowers as feeding and breeding resources (Sepel *et al.*, 2000; Santos and Vilela, 2005; Robe *et al.*, 2013), the ethological or physiological characteristics associated with this restricted ecology have not yet been recognized. Our results regarding the olfactory preferences of *D. incompta* suggest a chemosensory basis for this condition, as indicated by the clear preference for extracts of *Cestrum* flowers compared to other scents (*Brunfelsia*, banana, or orange extract scents). It is generally expected that at least part of this specialization is related to chemosensory changes, ultimately linked to genetic alterations. This hypothesis is supported by some of the results presented here for the OR and GR gene families in *D. incompta*, such as some autapomorphic or homoplastic gene losses suggested for some genes, in contrast to the signals of positive selection detected for others.

It is well known that both OR and GR gene families frequently evolve through gene duplication, gene loss, and pseudogenization (Gardiner *et al.*, 2008; Benton, 2015). In particular, it has been suggested that during host specialization there may be a general trend toward loss of some chemosensory genes with concomitant positive selection on others (McBride, 2007; McBride and Arguello, 2007). For *D. incompta*, orthologous sequences were found for 84 of the 120 chemoreceptor genes examined in this study. Different strategies were employed for this task, among which InsectOR revealed the most sensitive, being able to recover 42 of the 44 ORs, and all the GRs, through the use of a mixed query matrix containing *D. melanogaster* and *D. virilis* chemoreceptor genes. Consistency of these results were further revealed by the ML tree reconstructed with amino acid sequences for each family, where most identifications performed with InsectOR recovered monophyletic groups. In fact, the only exceptions to this pattern were related to polytomous or poorly supported relationships, in which putatively orthologous sequences encompassed short branches that were located as adjacent in the three. Such results are probably an outcome related to the insufficient phylogenetic signal presented by some sequences, or to the positive selection detected for others (see below), since none *D. incompta* sequence appeared as nested within clades compounded by alternative orthogroups.

Although the small number of genes recovered for *D. incompta* seems to be consistent with the hypothesis

suggesting a general trend of gene loss in specialized species (McBride and Arguello, 2007; Croset *et al.*, 2010), this pattern should be interpreted with caution. At first, it is important to consider that at least some of these putative losses could be an outcome of methodological sources of error related to sequencing coverage, genome assembling or annotation. Moreover, a closer inspection to the general phylogenetic profile reveals that most of the missing chemoreceptor genes detected in *D. incompta* appear to have been lost in the ancestors (see below), reflecting a phylogenetic signal or a historical contingency. Indeed, only two of the 36 missing genes of *D. incompta* provided evidence of autapomorphic or homoplastic gene loss in the target species. This is true, for example, for the putative loss of *OR1a* and *OR94b* in the *D. incompta* genome, which were previously shown to respond to specific odorants, like the aldehyde (E)-2-hexenal (Fishilevich *et al.*, 2005) and the phenol 4-methylphenol (Kreher *et al.*, 2005), respectively. Interestingly, neither of these compounds seem to have been isolated in *Cestrum* yet (Alrabayah *et al.*, 2024).

The effects of differential selection pressure in the repertoire of chemoreceptor genes of *D. incompta* could be directly tested in the positive selection tests, when tree genes of both the GR and the OR gene families showed signals of positive selection. Interestingly, two of the six genes exhibiting such signals (OR63a and GR57a) were found in all other species examined, suggesting that their products serve some important functions that are somehow conserved across the *Drosophila* phylogeny. Otherwise, GR64e seems to be missed only in *D. busckii*, and was previously shown to confer responsiveness to glycerol, a byproduct of yeast fermentation, affecting feeding preferences in *Drosophila* (Wisotsky *et al.*, 2011). Since yeasts are common inhabitants of many angiosperm nectars (Pozo *et al.*, 2009), this result raises the possibility that yeasts are the effective feeding resources used by *D. incompta* larvae and/or adults in *Cestrum* flowers, suggesting that the *D. flavopilosa* group may not have diverged so much from the ancestral Drosophilidae feeding habits (Throckmorton, 1975).

Evolution of OR and GR genes in *Drosophila*

By increasing the number of species studied, we improved our understanding of the evolutionary patterns of the OR and GR gene families across the phylogeny of *Drosophila*. In this sense, we found a large variation in the number of chemoreceptor genes among species, with the higher and lower numbers of OR and GR genes generally found for species of the subgenus *Sophophora*, especially for species of the *melanogaster* subgroup, and for *D. busckii*, respectively. Because *D. melanogaster* was one of the species used as a query or seed in our searches, the higher number of chemoreceptor genes found for this species and other closely related taxa may be a methodological artifact. The lower number of orthologous sequences found for *D. busckii* could also be a biased result, as this species is not closely related to either of the two species used as a query or seed in our searches (Yassin, 2013; O'Grady and DeSalle, 2018). Moreover, chemoreceptor annotation for this species has not yet received special attention. Nevertheless, such a pattern

does not explain the results found for species closely related to *D. virilis*, since the last species not only presents a well-annotated genome but was also used as a query or seed in our searches. Even so, the number of orthologous sequences found in this species is close to the lower range for the OR and GR genes (50 and 40, respectively). Another point to consider is that we are only looking at one copy of each gene, which means that the scenario presented focuses on more ancient events of gene duplication. In this sense, the number of genes may be somewhat underestimated for some species. This is the case for *D. grimshawi*, which has already been shown to have multiple copies of some loci associated with recent gene duplications (Guo and Kim, 2007; Nozawa and Nei, 2007; Gardiner *et al.*, 2008). Even so, it is important to mention that our OR family annotation generally agrees with Ramasamy *et al.* (2016) concerning the number of genes for each species. The few divergences may be explained by the use of different annotation tools, albeit the main approach of both studies was the BLAST analyses.

Despite some putative sources of bias, the general picture suggests that the distribution of OR and GR genes across the *Drosophila* phylogeny may reflect a phylogenetic signal related to gene gains or losses in ancestral populations that have been stably inherited over millions of years. These conclusions are further supported when the presence or absence of each of the 16 OR and 20 GR genes of *D. melanogaster* not found in *D. incompta* are evaluated in a phylogenetic context. Of the ORs not found in *D. incompta*, six were also not found in any non-*Sophophora* species (OR7a, OR33a, OR33b, OR43b, OR59c, and OR65a), and three were found in only one species of the subgenus *Siphlodora* (OR65c in *D. arizonae*, and OR85a in *D. mojavensis*) or the subgenus *Dorsilopha* (OR45a in *D. busckii*). This pattern also holds for the gene family GR, where most genes not found in *D. incompta* appear to be largely restricted to the *Sophophora* subgenus. Indeed, 16 GR genes (GR10b, GR22a, GB22b, GR22c, GR22d,

GR22f, GR36a, GR36b, GR36c, GR59a, GR59c, GR92a, GR93b, GR93d, GR98c, and GR98d) were not found in any non-*Sophophora* species, and two of them (GR22c and GR93d) were present in only five *Sophophora* species. These absences may indicate ancestral gene losses in non-*Sophophora* species, gene gains in *Sophophora* species, or even high divergence compared to query sequences reaching the limit of sensitivity of BLAST. In fact, comparisons involving the mean number of chemoreceptor genes between *Sophophora* with 57.36 ($\sigma = 2.44$) for OR and 53.82 ($\sigma = 3.96$) for GR genes, and non-*Sophophora* species with 44.67 ($\sigma = 3.53$) for OR and 39 ($\sigma = 2.31$) for GR genes, further agrees with these putative phylogenetic trends (Figure 5). Nevertheless, there are interesting exceptions to this rule, like *D. willistoni* and *D. busckii*, which stands out as a *Sophophora* and a non-*Sophophora* species with a lower number of OR and GR genes respectively.

Because species in the *melanogaster* group often have the greater number of OR and GR genes (60 for both in *D. melanogaster*), it is tempting to speculate that, in at least some cases, the number of chemoreceptor genes also defines evolutionary potentials related not only to niche breadth but also to the distribution range of individual species within each of the major lineages of *Drosophila*. Indeed, several *Sophophora* species are recognized as generalists and cosmopolitans, such as *D. ananassae* (Singh, 1996), *D. kikkawai* (Ramniwas and Kajla, 2012), *D. mauritiana* (Tsacas and David, 1974), *D. melanogaster* (David and Capy, 1988; Lachaise and Silvain, 2004; Markow, 2015), *D. obscura* (Begon, 1975; Markow and O'Grady, 2008), *D. persimilis* (Wogaman and Seiger, 1983), *D. pseudoobscura* (Carson, 1971), *D. serrata* (Kelemen and Moritz, 1999; Schiffer and McEvey, 2006), *D. simulans* (Lemeunier *et al.*, 1986), *D. sukukii* (Asplen *et al.*, 2015; Biondi *et al.*, 2016), and *D. willistoni* (Spassky *et al.*, 1971, Cordeiro and Winge, 1995; Powell, 1997), a pattern that appears to be less common in

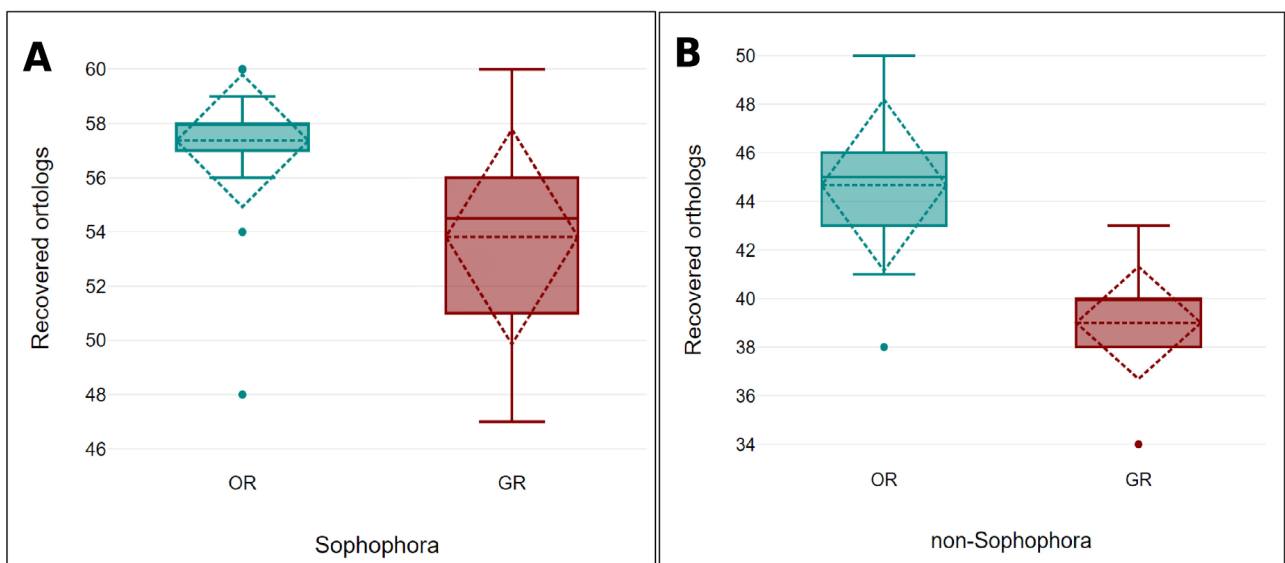


Figure 5 – Box plot graph comparing the total number of orthologs found for each of the two gene families by different *Drosophila* lineages (A: *Sophophora* species and B: non-*Sophophora* species). The bold horizontal line in the boxes represents the median, while the dashed lines represent the mean with standard deviation. Dots represent the outliers.

non-*Sophophora* species. Nevertheless, this hypothesis also has some important exceptions, as in the case of *D. sechellia*, a *Sophophora* species with 120 chemoreceptor genes which is endemic to the Seychelles islands in the Indian Ocean and is specialized to a single resource, the fruit of *Morinda citrifolia* (Tsacas and Bächli, 1981; Louis and David, 1986; Gerlach, 2009); and *D. erecta*, another species of this subgenus, which has 59 and 56 OR and GR genes, respectively, and is endemic to west-central Africa, and specialized on ripe fruit of *Pandanus* spp. (Lachaise and Tsacas, 1974; Rio *et al.*, 2010; Linz *et al.*, 2013). This reinforces that, although several trends and potentials seem to have been inherited from ancestral populations, the patterns and processes behind the evolution of OR and GR gene families are certainly quite complex and need to be evaluated under different perspectives.

Conclusion

The evolution of specialization in the use of specific resources is an interesting topic that is not yet fully understood (Etges, 2019; Markow, 2019; Auer *et al.*, 2020). Here, we demonstrated the clear preference of *D. incompta* for the scent of macerated *Cestrum* flowers compared to other resources and characterized some genomic changes on the OR and GR genes that are possibly related to this behavior. In characterizing the chemoreceptor repertoire of *D. incompta*, we not only detected some autapomorphic or homoplastic gene losses, but also found several indications of positive selection on different components of the two major gene families that encompass some of the basic tools for interacting with nature (Benton, 2015). However, most gene losses in this species appear to have first occurred in its ancestors located at different nodes of the *Drosophila* phylogenetic tree. So, generally speaking, our results do not seem to corroborate the idea that specialized species have a more limited number of ORs and GRs. Nevertheless, the precise relationship between these changes and the ecological potential of each species remains largely unclear, and it is not currently possible to decipher the putative causes and consequences of ecological specialization.

Acknowledgements

This study was supported by research grants and fellowships from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). PMF received a research grant from CNPq – GD. ELSL and LJR are research fellows of CNPq – PQ (305613/2020-0 and 314206/2021-3, respectively).

Conflict of Interest

The authors declare that there is no conflict of interest.

Author Contributions

PMF conducted the samples collections, the olfactory choice tests, the chemoreceptors repertoire searches and analyzes and wrote the manuscript; LJR analyzed the evolutionary data and wrote the manuscript; TNC performed the chemoreceptors repertoire searches and analyzes and wrote the manuscript; ELSL conceived the study and wrote the manuscript. All authors read and approved the final version.

References

- Afgan E, Baker D, Batut B, van den Beek M, Bouvier D, Cech M, Chilton J, Clements D, Coraor N, Grüning BA, Guerler A *et al.* (2018) The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Res* 46:W537-W544.
- Allen JM, LaFrance R, Folk RA, Johnson KP and Guralnick RP (2018) aTRAM 2.0: An improved, flexible locus assembler for NGS data. *Evol Bioinform Online* 14:1176934318774546.
- Alrabayah IN, El Hawary SS, El-Kadder EM, Essa A and Elraey M (2024) Genus *Cestrum* therapeutic potential: An updated review of its phytochemical, pharmacological, and morphological features. *Egyptian J Chem* 67:375-393.
- Anholt RR (2020) Chemosensation and evolution of *Drosophila* host plant selection. *Science* 23:100799.
- Antipov D, Korobeynikov A, McLean JS and Pevzner PA (2016) hybridSPAdes: An algorithm for hybrid assembly of short and long reads. *Bioinformatics* 32:1009-1015.
- Asplen MK, Anfora G, Biondi A, Choi DS, Chu D, Daane KM, Gibert P, Gutierrez AP, Hoelmer KA, Hutchison WD *et al.* (2015) Invasion biology of spotted wing *Drosophila* (*Drosophila suzukii*): A global perspective and future priorities. *J Pest Sci* 88:469-494.
- Auer TO, Khallaf MA, Silbering AF, Zappia G, Ellis K, Álvarez-Ocaña R, Arguello JR, Hansson BS, Jefferis GS, Caron SJ *et al.* (2020) Olfactory receptor and circuit evolution promote host specialization. *Nature* 579:402-408.
- Begon M (1975) The relationships of *Drosophila obscura* Fallen and *D. subobscura* Collin to naturally-occurring fruits. *Oecologia* 20:255-277.
- Benton R (2015) Multigene family evolution: Perspectives from insect chemoreceptors. *Trends Ecol Evol* 30:590-600.
- Biondi A, Traugott M and Desneux N (2016) Special issue on *Drosophila suzukii*: From global invasion to sustainable control. *J Pest Sci* 89:603-604.
- Brcnc D (1966) Ecological and cytogenetic studies of *Drosophila flavopilosa*, a neotropical species living in *Cestrum* flowers. *Evolution* 1:16-29.
- Brcnc D (1983) The *flavopilosa* group of species as an example of flower-breeding species. In: Ashburner M, Carson HL and Thompson JN (eds) *The genetics and biology of Drosophila*. 3rd edition. Academic Press, New York, pp 360-377.
- Cande J, Prud'homme B and Gompel N (2013) Smells like evolution: The role of chemoreceptor evolution in behavioral change. *Curr Opin Neurobiol* 23:152-158.
- Carey AF, Wang G, Su CY, Zwiebel LJ and Carlson JR (2010) Odorant reception in the malaria mosquito *Anopheles gambiae*. *Nature* 464:66-71.
- Carson HL (1971) Speciation and the founder principle. *Stadler Genetics Symposium* 3:51-70.
- Chen ZX, Sturgill D, Qu J, Jiang H, Park S, Boley N, Suzuki AM, Fletcher AR, Plachetzki DC, FitzGerald PC *et al.* (2014) Comparative validation of the *D. melanogaster* modENCODE transcriptome annotation. *Genome Res* 24:1209-1223.
- Clyne PJ, Warr CG, Freeman MR, Lessing D, Kim J and Carlson JR (1999) A novel family of divergent seven-transmembrane proteins: Candidate odorant receptors in *Drosophila*. *Neuron* 22:327-338.
- Cordeiro AR and Winge H (1995) Levels of evolutionary divergence of *Drosophila willistoni* sibling species. In: Levine L (ed), *Genetics of Natural Populations of Theodosius Dobzhansky*. 1st edition. Columbia University Press, New York, pp. 241-261
- Croset V, Rytz R, Cummins SF, Budd A, Brawand D, Kaessmann H, Gibson TJ and Benton R (2010) Ancient protostome origin of chemosensory ionotropic glutamate receptors and the evolution of insect taste and olfaction. *PLoS Genet* 6:e1001064.

- Dahanukar A, Hallem EA and Carlson JR (2005) Insect chemoreception. *Curr Opin Neurobiol* 15:423-430.
- Date P, Dweck HK, Stensmyr MC, Shann J, Hansson BS and Rollmann SM (2013) Divergence in olfactory host plant preference in *D. mojavensis* in response to cactus host use. *PLoS One* 8:e70027.
- David JR and Capy P (1988) Genetic variation of *Drosophila melanogaster* natural populations. *Trends Genet* 4:106-111.
- Davis RL (2007) The scent of *Drosophila* sex. *Neuron* 54:14-16.
- De Ré FC, Robe LJ, Wallau GL and Loreto EL (2017) Inferring the phylogenetic position of the *Drosophila flavopilosa* group: Incongruence within and between mitochondrial and nuclear multilocus datasets. *J Zool Syst Evol Res* 55:208-221.
- Depetris-Chauvin A, Galagovsky D and Grosjean Y (2015) Chemicals and chemoreceptors: Ecologically relevant signals driving behavior in *Drosophila*. *Front Ecol Evol* 3:41.
- Diaz F, Allan CW and Matzkin LM (2018) Positive selection at sites of chemosensory genes is associated with the recent divergence and local ecological adaptation in cactophilic *Drosophila*. *BMC Evol Biol* 18:144.
- Doleschall CL (1858) Derde leijdage tot de kennis der dipteran fauna van Nederlandsh Indie. *Natr. Tijdschr. Nederland Indie* 17:73-128.
- Drosophila 12 Genomes Consortium, Clark AG, Eisen MB, Smith DR, Bergman CM, Oliver B, Markow TA, Kaufman TC, Kellis M, Gelbart W, Iyer VN *et al.* (2007) Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* 450:203-218.
- Dweck HK, Ebrahim SA, Kromann S, Bown D, Hillbur Y, Sachse S, Hansson BS and Stensmyr MC (2013) Olfactory preference for egg laying on citrus substrates in *Drosophila*. *Curr Biol* 23:2472-2480.
- Ebrahim SA, Dweck HK, Stökl J, Hofferberth JE, Trona F, Weniger K, Rybak J, Seki Y, Stensmyr MC, Sachse S *et al.* (2015) *Drosophila* avoids parasitoids by sensing their semiochemicals via a dedicated olfactory circuit. *PLoS Biol* 13:e1002318.
- Etges WJ (2019) Evolutionary genomics of host plant adaptation: Insights from *Drosophila*. *Curr Opin Insect Sci* 36:96-102.
- Fishilevich E, Domingos AI, Asahina K, Naef F, Vosshall LB and Louis M (2005) Chemotaxis behavior mediated by single larval olfactory neurons in *Drosophila*. *Curr Biol* 15:2086-2096.
- Frota-Pessoa O (1952) Flower-feeding *Drosophilidae*. *Drosoph Inf Ser* 26:101-102.
- Fu L, Niu B, Zhu Z, Wu S and Li W (2012) CD-HIT: Accelerated for clustering the next-generation sequencing data. *Bioinformatics* 28:3150-3152.
- Fuyama Y (1976) Behavior genetics of olfactory responses in *Drosophila*. I. Olfactometry and strain differences in *Drosophila melanogaster*. *Behav Genet* 6:407-420.
- Gao Q and Chess A (1999) Identification of candidate *Drosophila* olfactory receptors from genomic DNA sequence. *Genomics* 60:31-39.
- Gardiner A, Barker D, Butlin RK, Jordan WC and Ritchie MG (2008) *Drosophila* chemoreceptor gene evolution: Selection, specialization and genome size. *Mol Ecol* 17:1648-1657.
- Gerlach J (2009) Conservation of the Seychelles sheath-tailed bat *Coleura seychellensis* on Silhouette Island, Seychelles. *Endanger Species Res* 8:5-13.
- Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q *et al.* (2011) Trinity: Reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nat Biotechnol* 29:644.
- Guo S and Kim J (2007) Molecular evolution of *Drosophila* odorant receptor genes. *Mol Biol Evol* 24:1198-1207.
- Gurevich A, Saveliev V, Vyahhi N and Tesler G (2013) QUAST: Quality assessment tool for genome assemblies. *Bioinformatics* 29:1072-1075.
- Hallem EA, Dahanukar A and Carlson JR (2006) Insect odor and taste receptors. *Annu Rev Entomol* 51:113-135.
- Hickner PV, Rivaldi CL, Johnson CM, Siddappaji M, Raster GJ and Syed Z (2016) The making of a pest: Insights from the evolution of chemosensory receptor families in a pestiferous and invasive fly, *Drosophila suzukii*. *BMC Genom* 17:648.
- Hjelmén CE and Johnston JS (2017) The mode and tempo of genome size evolution in the subgenus *Sophophora*. *PLoS One* 12:e0173505.
- Hofmann PR and Napp M (1984) Genetic-environmental relationships in *Drosophila incompta*, a species of restricted ecology. *Rev Bras Genet* 7:21-39.
- Huerta-Cepas J, Forslund K, Coelho LP, Szklarczyk D, Jensen LJ, Von Mering C and Bork P (2017) Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. *Mol Biol Evol* 34:2115-2122.
- Joseph RM and Carlson JR (2015) *Drosophila* chemoreceptors: A molecular interface between the chemical world and the brain. *Trends Genet* 31:683-695.
- Karpe SD, Tiwari V and Ramanathan S (2021) InsectOR—Webserver for sensitive identification of insect olfactory receptor genes from non-model genomes. *PLoS One* 16:e0245324.
- Kelemen L and Moritz C (1999) Comparative phylogeography of a sibling pair of rainforest *Drosophila* species (*Drosophila serrata* and *D. birchii*). *Evolution* 53:1306-1311.
- Kreher SA, Kwon JY and Carlson JR (2005) The molecular basis of odor coding in the *Drosophila* larvae. *Neuron* 46:445-456.
- Kumar S, Stecher G, Li M, Knyaz C and Tamura K (2018) MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol* 35:1547-1549.
- Kurtovic A, Widmer A and Dickson BJ (2007) A single class of olfactory neurons mediates behavioural responses to a *Drosophila* sex pheromone. *Nature* 446:542-546.
- Lachaise D and Tsacas L (1974) Les drosophilidae des savanes préforestières de la région tropicale de Lamto (Cote-d'Ivoire). II. Le peuplement des fruits de *Pandanus candelabrum* (pandanaceae). *Ann Univ Abidjan* 7:153-192.
- Lachaise D and Silvain JF (2004) How two Afrotropical endemics made two cosmopolitan human commensals: The *Drosophila melanogaster*-*D. simulans* palaeogeographic riddle. *Genetica* 120:17-39.
- Lemeunier F, Tsacas L, David J and Ashburner M (1986) The *melanogaster* species group. In: Ashburner M, Thompson JR Jr and Carson HL (eds) *The genetics and biology of Drosophila*, vol. 3, Academic Press, New York, pp 147-256.
- Linz J, Baschwitz A, Strutz A, Dweck HK, Sachse S, Hansson BS and Stensmyr MC (2013) Host plant-driven sensory specialization in *Drosophila erecta*. *Proc Royal Soc B* 280:20130626.
- Louis J and David JR (1986) Ecological specialization in the *Drosophila melanogaster* species subgroup: A case study of *D. sechellia*. *Acta Ecol Gener* 7:215-229.
- Ludwig A, Vidal NM, Loreto EL and Sepel LM (2002) *Drosophila incompta* development without flowers. *Drosoph Inf Ser* 85:40-41.
- Malogolowkin C (1953) Sobre a genitália dos drosophilídeos. IV. A genitália masculina no subgênero *Drosophila* (Diptera, Drosophilidae). *Rev Bras Biol* 13:245-264.
- Mansourian S, Enjin A, Jirle EV, Ramesh V, Rehermann G, Becher PG, Pool JE and Stensmyr MC (2018) Wild African *Drosophila melanogaster* are seasonal specialists on marula fruit. *Curr Biol* 28:3960-3968.
- Markow TA and O'Grady P (2008) Reproductive ecology of *Drosophila*. *Funct Ecol* 747-759.
- Markow TA (2015) The natural history of model organisms: The secret lives of *Drosophila* flies. *Elife* 4:e06793.

- Markow TA (2019) Host use and host shifts in *Drosophila*. *Curr Opin Insect Sci* 31:139-145.
- Martin F, Boto T, Gomez-Diaz C and Alcorta E (2013) Elements of olfactory reception in adult *Drosophila melanogaster*. *Anat Rec* 296:1477-1488.
- McBride CS (2007) Rapid evolution of smell and taste receptor genes during host specialization in *Drosophila sechellia*. *Proc Natl Acad Sci U S A* 104:4996-5001.
- McBride CS and Arguello JR (2007) Five *Drosophila* genomes reveal nonneutral evolution and the signature of host specialization in the chemoreceptor superfamily. *Genetics* 177:1395-1416.
- Nguyen LT, Schmidt HA, Von Haeseler A and Minh BQ (2015) IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* 32:268-274.
- Nozawa M and Nei M (2007) Evolutionary dynamics of olfactory receptor genes in *Drosophila* species. *Proc Natl Acad Sci U S A* 104:7122-7127.
- O'Grady PM and DeSalle R (2018) Phylogeny of the genus *Drosophila*. *Genetics* 209:25.
- Pipkin SB, Rodriguez RL and Leon J (1966) Plant host specificity among flower-feeding neotropical *Drosophila* (Diptera: Drosophilidae). *Am Nat* 100:135-156.
- Pond SLK and Frost SD (2005) Not so different after all: A comparison of methods for detecting amino acid sites under selection. *Mol Biol Evol* 22:1208-1222.
- Powell JR (1997) Progress and prospects in evolutionary biology: The *Drosophila* model. Oxford University Press, Oxford, 576 p.
- Pozo MJ, de Vega C, Canto A and Herrera CM (2009) Presence of yeasts in floral nectar is consistent with the hypothesis of microbial-mediated signaling in plant-pollinator interactions. *Plant Signal Behav* 4:1102-1104.
- Ramasamy S, Ometto L, Crava CM, Revadi S, Kaur R, Horner DS, Pisani D, Dekker T, Anfora G and Rota-Stabelli O (2016) The evolution of olfactory gene families in *Drosophila* and the genomic basis of chemical-ecological adaptation in *Drosophila suzukii*. *Genome Biol Evol* 8:2297-2311.
- Ramiwas S and Kajla B (2012) Divergent strategy for adaptation to drought stress in two sibling species of *montium* species subgroup: *Drosophila kikkawai* and *Drosophila leontia*. *J Insect Physiol* 58:1525-1533.
- Ranwez V, Harispe S, Delsuc F and Douzery EJ (2011) MACSE: Multiple Alignment of Coding SEquences accounting for frameshifts and stop codons. *PLoS One* 6:e22594.
- Rio DC, Ares M, Hannon GJ and Nilsen TW (2010) Purification of RNA using TRIzol (TRI reagent). *Cold Spring Harb Protoc* 2010:pdb-rot5439.
- Robe LJ, De Ré FC, Ludwig A and Loreto EL (2013) The *Drosophila flavopilosa* species group (Diptera, *Drosophilidae*) An array of exciting questions. *Fly* 7:59-69.
- Robe LJ, Loreto EL and Valente VL (2010) Radiation of the "*Drosophila*" subgenus (*Drosophilidae*, Diptera) in the Neotropics. *J Zool Syst Evol Res* 48:310-321.
- Robertson HM, Warr CG and Carlson JR (2003) Molecular evolution of the insect chemoreceptor gene superfamily in *Drosophila melanogaster*. *Proc Natl Acad Sci U S A* 100:14537-14542.
- Santos RD and Vilela CR (2005) Breeding sites of Neotropical Drosophilidae (Diptera): IV. Living and fallen flowers of *Sessea brasiliensis* and *Cestrum* spp. (Solanaceae). *Rev Bras Entom* 49:544-551.
- Schiffer M and McEvey SF (2006) *Drosophila bunnanda*: A new species from northern Australia with notes on other Australian members of the *montium* subgroup (Diptera: Drosophilidae). *Zootaxa* 1333:1-23.
- Schmitz HJ and Valente VL (2019) The flower flies and the unknown diversity of Drosophilidae (Diptera): A biodiversity inventory in the Brazilian fauna. *Pap Avulsos Zool* 59:e20195945.
- Scott K, Brady Jr R, Cravchik A, Morozov P, Rzhetsky A, Zuker C and Axel R (2001) A chemosensory gene family encoding candidate gustatory and olfactory receptors in *Drosophila*. *Cell* 104:661-673.
- Sepel LM, Golombieski RM, Napp M and Loreto EL (2000) Seasonal fluctuations of *D. cestri* and *D. Incompta*, two species of the *flavopilosa* group. *Drosoph Inf Serv* 83:122-126.
- Sevenster JG and Van Alphen JJ (1993) A life history trade-off in *Drosophila* species and community structure in variable environments. *J Anim Ecol* 62:720-736.
- Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV and Zdobnov EM (2015) BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31:3210-3212.
- Singh BN (1996) Population and behaviour genetics of *Drosophila ananassae*. *Genetica* 97:321-329.
- Smith MD, Wertheim JO, Weaver S, Murrell B, Scheffler K and Kosakovsky Pond SL (2015) Less is more: An adaptive branch-site random effects model for efficient detection of episodic diversifying selection. *Mol Biol Evol* 32:1342-1353.
- Spassky B, Richmond RC, Perez-Salas S, Pavlovsky O, Mourao CA, Hunter AS, Hoenigsberg H, Dobzhansky T and Ayala FJ (1971) Geography of the sibling species related to *Drosophila willistoni*, and of the semispecies of the *Drosophila paulistorum* complex. *Evolution* 25:129-143.
- Stanke M, Keller O, Gunduz I, Hayes A, Waack S and Morgenstern B (2006) AUGUSTUS: *Ab initio* prediction of alternative transcripts. *Nucleic Acids Res* 34:W435-W439.
- Sturtevant AH (1919) A new species closely resembling *Drosophila melanogaster*. *Psyche* 26:153-155.
- Thompson JD, Higgins DG and Gibson TJ (1994) CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22:4673-4680.
- Throckmorton LH (1975) The phylogeny, ecology and geography of *Drosophila*. In: King RC (ed). *Handbook of genetics*, vol 3, Plenum Press, New York, pp 421-469.
- Thurmond J, Goodman JL, Strelets VB, Attrill H, Gramates LS, Marygold SJ, Matthews BB, Millburn G, Antonazzo G, Trovisco V *et al.* (2019) FlyBase 2.0: The next generation. *Nucleic Acids Res* 47:59-65.
- Tsacas L and David J (1974) *Drosophila mauritiana* n. sp. du groupe *melanogaster* de l'île Maurice [Dipt. Drosophilidae]. *Bull Soc Entomol* 79:42-46.
- Tsacas L and Lachaise D (1974) Quatre nouvelles especes de la Cote-d'Ivoire du genre *Drosophila*, groupe *melanogaster* et discussion de l'origine du sous-groupe *melanogaster* (Diptera: Drosophilidae). *Ann Univ Abidjan* 7:193-211.
- Tsacas L and Bächli G (1981) *Drosophila sechellia*, n. sp., huitieme espece du sous-groupe *melanogaster* des îles Sechelles (Diptera, Drosophilidae). *Rev Fr Entomol* 3:146-150.
- Vieira FG and Rozas J (2011) Comparative genomics of the odorant-binding and chemosensory protein gene families across the Arthropoda: Origin and evolutionary history of the chemosensory system. *Genome Biol Evol* 3:476-490.
- Vosshall LB and Stocker RF (2007) Molecular architecture of smell and taste in *Drosophila*. *Annu Rev Neurosci* 30:505-533.
- Weaver S, Shank SD, Spielman SJ, Li M, Muse SV and Kosakovsky Pond SL (2018) Datamonkey 2.0: A modern web application for characterizing selective and other evolutionary processes. *Mol Biol Evol* 35:773-777.

- Wheeler MR, Takada H and Brncic D (1962) The *flavopilosa* species group of *Drosophila*. Univ Texas Publ Genetics 6205:395-414.
- Whiteman NK and Pierce NE (2008) Delicious poison: Genetics of *Drosophila* host plant preference. Trends Ecol Evol 23:473-478.
- Wisotsky Z, Medina A, Freeman E and Dahanukar A (2011) Evolutionary differences in food preference rely on Gr64e, a receptor for glycerol. Nat Neurosci 14:1534-1541.
- Wogaman DJ and Seiger MB (1983) Light intensity as a factor in the choice of an oviposition site by *Drosophila pseudoobscura* and *Drosophila persimilis*. Can J Genet Cytol 25:370-377.
- Wyatt TD (2003) Pheromones and animal behaviour: Communication by smell and taste. Cambridge University Press, Cambridge.
- Yang Y, Hou ZC, Qian YH, Kang H and Zeng QT (2012) Increasing the data size to accurately reconstruct the phylogenetic relationships between nine subgroups of the *Drosophila melanogaster* species group (*Drosophilidae*, Diptera). Mol Phylogenet Evol 62:214-223.
- Yassin A (2013) Phylogenetic classification of the *Drosophilidae* Rondani (Diptera): The role of morphology in the postgenomic era. Syst Entomol 38:349-364.

Internet Resources

- FigTree v1.4 (2018), Software, <http://tree.bio.ed.ac.uk/software/figtree/> (accessed 4 March 2024).
- The database on taxonomy of Drosophilidae (TaxoDros), <https://www.taxodros.uzh.ch/> (accessed 24 September 2023).

Supplementary Material

The following online material is available for this article:

Figure S1 – Phylogenetic tree reconstructed with amino acid sequences related to the odorant receptor (OR) gene family in *D. incompta*, *D. melanogaster* and *D. virilis*.

Figure S2 – Phylogenetic tree reconstructed with amino acid sequences related to the gustatory receptor (GR) gene family in *D. incompta*, *D. melanogaster* and *D. virilis*.

Table S1 – Sample design of the five olfactory choice tests performed with *Drosophila incompta* (A, B, and C) and *D. melanogaster* (D and E), comparing odors of *Cestrum* × Banana (replicates A and D), *Cestrum* × Orange (replicates B), *Cestrum* × *Brunfelsia uniflora* (replicates C), and Banana × Orange (replicates E).

Table S2 – Metrics and statistics of *Drosophila incompta* genome and transcriptome draft assemblies.

Table S3 – Accession numbers (GenBank – NCBI) of the sequences used in evolutionary analysis matrices.

Table S4 – Olfactory (OR, sheet A) and gustatory receptor (GR, sheet B) orthologues recovered for *Drosophila incompta* by each tool.

Table S5 – List of presence (X on the table) or absence (- on the table) of olfactory (OR, sheet A) and gustatory receptor (GR, sheet B) orthologues retrieved for each of the analyzed *Drosophila* species.

Associate Editor: Fabricio Rodrigues dos Santos

License information: This is an open-access article distributed under the terms of the Creative Commons Attribution License (type CC-BY), which permits unrestricted use, distribution and reproduction in any medium, provided the original article is properly cited.