UNIVERSITÀ DEGLI STUDI DI MILANO

Scuola di Dottorato in Scienze Biologiche e Molecolari

XXVIII ciclo

The evolution of chemoreception in the invasive pest *Drosophila suzukii* and other arthropods

Sukanya Ramasamy

PhD Thesis

Scientific tutor: Dr. Omar Rota-Stabelli

Academic year: 2014-2015

SSD: BIO/10, BIO/11

Thesis was performed at the Agricultural Entomology group, Department of Sustainable Agro-Ecosystems and Bio-resources (DASB), Fondazione Edmund Mach (FEM), San Michele all' Adige, TN, Italy.

And

The School of Earth Sciences,

University of Bristol, UK

Thesis Committee

Supervisors

Dr. Omar Rota-Stabelli, Dr. Gianfranco Anfora

Fondazione Edmund Mach

Doctoral Coordinator

Prof. David Stephen Horner,

Università Degli Studi Di Milano

Front Cover is an illustration for Charles Dickens's 1883 Cricket on the Hearth by Fred Barnard, here depicting the force of natural selection on the insect

Contents

Abstract…………………………………………………………………………………….....1

1. **Introduction**……………………………………………………………………………..…......2

1.1. **Chemoreception in insects: biology and applications**…………………..……....2

1.1.1. The physiological and morphological bases of chemoreception……………......3

1.1.2. The molecular bases of chemoreception………………………………........….4
1.1.2.1. Odorant Receptors……………………………………………..….......5
1.1.2.2. Gustatory Receptors………………………………………….…........7
1.1.2.3. New classes of chemoreceptors: Ionotropic Receptors……….........7
1.1.2.4. Binding proteins – An Overview……………………………………8

1.1.3. The importance of chemoreception for applied studies……………......……9

1.2. **Evolution, Genomics, and chemoperception………………………………..…………12**

1.2.1. On the origin and evolution of chemoreception in insects………..………..12
1.2.1.1. Birth-Death process of gene evolution……………………............12
1.2.1.2. The origin of chemosensation…………………......…….……......12

1.2.2. Comparative genomics in the post-genomic era………………….……....13

1.3. ***Drosophila suzukii*- an invasive pest in Western countries………..…..….....15**

2. Main Objectives………………………………………………………………….........19

3. Methods…………………………………………………………………………….....20

4. Summary of Results…………………………………………………………….............25

4.1. **Genome and transcriptome of *D. suzukii*………………………………….…....25**

4.1.1. Sequencing and Assembly of genome and transcriptome…………………......25

4.1.2. Comparative genomics of repetitive elements in *Drosophila*..……..…..….…26

4.1.3. Incongruence between mitochondrial and nuclear phylogeny……………...27

4.2. **Investigating the evolution of sexual pheromone perception in *D. suzukii*…..28**

4.3. **The role of olfaction in the making of a pest…………………………………..….34**

.

4.3.1. Evolution of chemosensory genes in *D. suzukii*: increased turnover rate of ORs, GRs and OBPs………………………………………………………..….34

4.3.2. Making sense of the ecological significance of the duplicated/lost genes….40

4.3.3. Altered responses of sensory physiology driven by loss of function of key receptors……………………………………………………………….…42

4.4. Evolution of chemoreceptors in Arthropoda…………………………….….44

4.4.1. Odorant receptors and new findings in Palaeoptera……………………….…..44

4.4.2. Lineage based, rapid evolution of GRs………………………………………..45

5. Conclusion and Future Perspectives……………………………………………...…..48

References……………………………………………………………………………..55

Appendix A - Supporting Materials…………………………………………………65

Appendix B – Articles………………………………………………………………...70

Acknowledgements…………………………………………………………………71

Manuscripts

- ✓ Linking genomics and ecology to investigate the complex evolution of an invasive Drosophila pest
- ✓ Olfactory responses of *Drosophila suzukii* female to host plant volatiles
- ✓ Loss of Drosophila pheromone reverses its role in sexual communication in *Drosophila suzukii*
- ✓ The evolution of olfactory gene families in *Drosophila* and the genomic basis of chemical-ecological adaptation in *Drosophila suzukii*

Abstract

Chemoreception is the process that allows animals to respond to the chemical stimuli in their environment. In insects, this is mediated by specialized neurons expressing a variety of dedicated receptors and carrier proteins: olfactory (OR), gustatory (GR) and Ionotropic (IR) receptors, odorant binding (OBP) and chemosensory (CSP) proteins. How the evolution of these genes correlates with adaption to new ecological niches is still a debated topic; when these genes arose during arthropods evolution is also an unresolved question. To tackle the first of these issues I have studied these gene families in *Drosophila suzukii* Matsumura (Diptera: Drosophilidae), an invasive pest that, unlike other *Drosophila*, oviposits in fresh fruits. I have initially contributed in curating *D. suzukii*'s genome and transcriptome, and then annotated its entire repertoire of chemosensory genes. Analysis of these genes on a 14 *Drosophila* phylogenetic framework revealed that ORs, OBPs, and GRs are characterized by high turnover rates and uneven distribution of duplication and gene loss events on the phylogeny: these peculiar evolutionary patterns are consistent with a change in selective pressures. *D. suzukii* is characterised by loss of function of key ORs that bind volatiles typically released by fermenting substrates, providing with a rare example of ecological adaptation due to genes loss. I further present my work in annotating and studying genes involved in the metabolism and perception of an aggregating pheromone that may play an active role in *D. suzukii*'s peculiar biology. To inquire the origin of chemosensory genes in insects and other arthropods, I have screened various recently released genomes: I identified multiple lineage-specific expansions of GRs in each of the arthropod clades, an interesting loss of GRs in most crustaceans, and the first genomic evidence of ORs in the Palaeoptera. My postgraduate work has contributed in identifying receptors and ligands that will be used in downstream *D. suzukii* control applications, and has shed some light in the evolution of chemosensory genes in arthropods.

1. Introduction

1.1. Chemoreception in insects: biology and applications

Chemosensation, the ability to respond to chemical stimuli, is likely the first of the senses to have evolved in the history of life. While bacteria possess a simple, although efficient form of chemical communication using quorum-sensing (Miller and Bassler 2001), animals have evolved a complex system which involves membrane receptors, accessory proteins, as well as dedicated neurons and organs where these proteins are expressed. Not surprisingly, chemical communication is still widespread in the animal kingdom (Wyatt 2003; Steiger et al. 2011). Like many mammals such as dogs and elephants (Niimura et al. 2014), insects rely mostly on their chemosenses for their biology; for example, chemosensation is essential for food location, mate finding, oviposition site choice and predator avoidance (Firestein 2001). In fact, as a means of survival, insects have developed a series of morphological and molecular tools, to discriminate among the variety of chemical cues present in the environment, in most cases independently to mammals and other animals.

From a molecular point of view, chemosensation in insects is mediated by at least 5 different protein families. Since a single insect may possess hundreds of different types of chemosensory proteins, expressed in thousands of dedicated different sensory neurons, chemosensation is clearly a highly intricate and dynamic system. In my thesis, I explore this complexity in insects, particularly in the emerging model, *D. suzukii*. In this section (1.1). I describe the morphological and molecular basis of chemosensation in insects with a special emphasis on *Drosophila*.

1.1.1. The physiological and morphological bases of chemoreception

In insects, this chemosensory signal transduction starts at the peripheral sensory system that involves hair like appendages called sensilla, which cover some of the body surface, particularly legs and antennae. These appendages are single (in case of gustatory sensilla) or multi-porous (in odorant sensilla), housing the dendrites of gustatory or olfactory receptor neurons (GRNs/ORNs) within them (de Bruyne et al. 2001; Hallem and Carlson 2004). The olfactory system is not

homologous between insects and vertebrates but convergent on a functional level; similarly, their structural organization and molecular mechanism is quite distinct (Park et al. 2000; Hallem and Carlson 2004; Su et al. 2009).

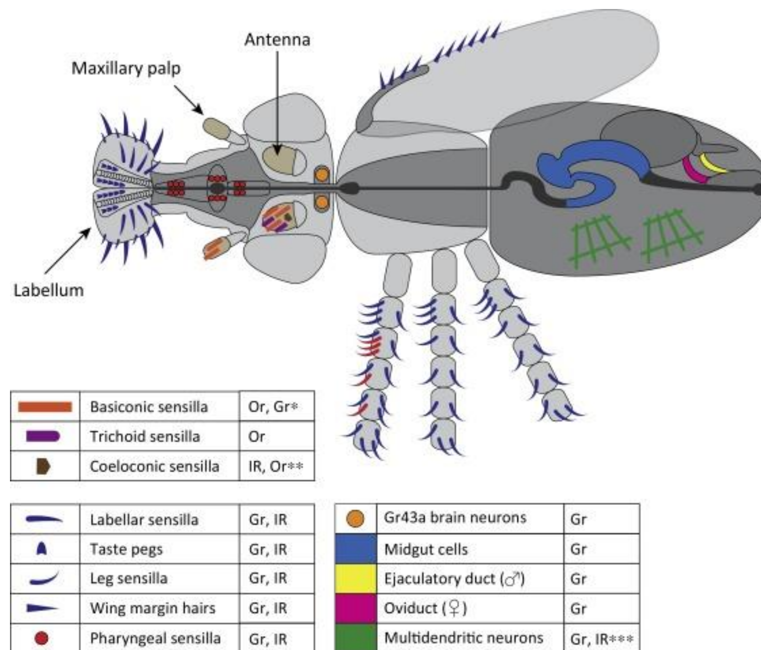


Fig 1.1. Distribution of chemosensory sensilla on the periphery of a fly. Light grey represents the exterior, dark grey, the interior. Tan represents the ORNs located on the antenna and maxillary palp. The red sensilla on the legs are male specific. Figure from (Joseph and Carlson 2015)

Morphologically, insect sensilla are classified into 3 types that differ in size, shape and cuticular structure (fig.1.1): club shaped basiconic sensilla, spine shaped trichoid sensilla and small cone shaped coeloconic sensilla (Shanbhag et al. 1999). The neurobiology of the chemosensory system is well studied in the *Drosophila genus*, because of the biological model *Drosophila melanogaster*. While the *Drosophila* antenna houses both basiconic and trichoid sensilla, maxillary palps bear only the basiconic. The distribution of sensilla types varies in male and female in most insects studied so far. At least in the antennae, males possess 30% more trichoid sensilla but 20% fewer basiconic sensilla than females (Stocker 1994), indicating a functional importance. Both basiconic and trichoid are housed in single walled sensilla. Each sensillum type has a different role. Basiconic is required for finding food (de Bruyne et al. 2010). Trichoid sensilla respond to odours involved in mate recognition (Hallem et al. 2006), such as cVA (cis vaccenyl acetate), an aphrodisiac compound released by male fruit flies that acts as a male-male aggressive pheromone and prevents

mated females from mating again. Coeloconic sensilla are unique,- they are shorter than the other two, and have two walls instead of one (Shanbhag et al. 1999): this type of sensilla is conserved and can be found in all of the insect orders (Croset et al. 2010). While the single walled basiconic and trichoid are predominantly housed only by ORNs, double walled coeloconic are housed both in the ORNs of the antennae and the GRNs of the taste organs.

Unlike the ORNs of the antennae and maxillary palp, the gustatory receptor neurons (GRNs) have a different distribution and architecture. They are mostly concentrated at proboscis, legs and wing margins (Stocker 1994). Strangely, a few GRNs are also reported in the ejaculatory duct and oviduct (Rice 1977; Stocker 1994). The external taste sensilla can be categorized on the basis of size, distribution and number of GRNs into three classes: short (s), intermediate (i) and long (l) types; where the short and long types have four neurons each, while the intermediate type has only 2 GRNs (Hiroi et al. 2002).

1.1.2. The molecular bases of chemosensation in insects

Chemosensation in insects is mediated by at least 5 different protein families, mostly expressed in ORN or GRN: Olfactory Receptors (OR), Gustatory Receptors (GR), Ionotropic Receptors (IR), Odorant Binding Proteins (OBP) and Chemosensory Proteins (CSP). In this section (1.2) I will provide a general introduction to the structure and function of these proteins.

Odour molecules in the air pass through pores in the external cuticle of the sensilla and enter the aqueous lymph where they bind to various types of proteins which are secreted by the supporting cells surrounding the ORN/GRNs (Olfactory Receptor Neuron/ Gustatory Receptor Neuron) (Shanbhag et al. 2001); two of them have been shown to play a key role in chemosensation: the odorant binding proteins (OBP) and the chemosensory proteins (CSP). The odour is transported to the dendrite surface of the receptor neurons by the binding protein OBP, after which the OR-Or83b complex triggers electrical stimuli, that are further processed in the higher centres of the brain. These stimuli are carried in a series of networks: the olfactory/gustatory receptor neurons are bipolar, spreading their dendrites in the sensillary lymph (where they bind odours), and project their axons (the odorant neurons) to spherical shaped, functional processing neuronal units called glomeruli in the antennal lobe (Hallem and Carlson 2004) of the brain, which is the functional equivalent (but not homolog) of the olfactory bulb of vertebrates (Hildebrand and Shepherd 1997).

In order to prevent the continuous stimulation, ODEs (Odorant Degrading Enzymes) help in degrading odorants bound to ORs (Leal 2013).

1.1.2.1. Odorant Receptors

First identified in mammals by (Buck and Axel 1991), odorant receptors have been found and annotated in *Drosophila* only after 7 years (Clyne et al. 1999; Gao and Chess 1999; Vosshall et al. 1999). The difficulty of identifying insect ORs from non-homologous mammalian ones (Benton et al. 2006; Wistrand et al. 2006), for example using similarity searches such as BLAST, explains this lag. In both insects and vertebrates, ORs are transmembrane receptors characterised by seven transmembrane (7TM) structures. But while mammalian ORs are a subfamily of GPCR proteins, insect ORs are characterised by an inverted topology, with an intracellular N-terminal and an extracellular C-terminal. The insect ORs constitute a very divergent set of genes, rapidly evolving through lineage specific expansions. Accordingly, the number of genes in each of the insect orders ranges from ~10 in human body louse (Kirkness et al. 2010) to ~ 300 in the wasp, *Nasonia vitripennis* (Robertson et al. 2010). As a comparison, mammalian ORs are generally numerous in number with 396 functional ORs in human and 1948 in elephant (Jiang and Matsunami 2015).

There are 62 odorant receptor proteins in *D. melanogaster*, resulting from the alternative splicing of transcripts from 60 OR loci. Among those 60 ORs, *Or83b/Orco* is quite conserved across insects (Jones et al. 2005) and reacts as a co-receptor, forming a heteromeric complex with all the other specific ORs; *Orco* does not act as an odorant receptor itself but rather as an ion channel (See below). Previous attempts to elucidate the molecular mechanism of the insect ORs revealed different scenarios (Sato et al. 2008; Wicher et al. 2008) (fig.1.2). While the first contribution proposed that ORs form a functional heteromeric complex of ligand binding OR along with *Orco* (itself an OR), conforming a gated cation channel, the second contribution suggested a G-protein coupled receptor-like mechanism coupled with a ligand gated cation channelling. In recent times, works on the functional mechanism of *Orco* revealed the presence of further signalling cascades such as PKC (Protein kinase C) (Sargsyan et al. 2011; Getahun et al. 2013). Currently, it is suggested that these 7TM proteins function as metabotropically acting cation channels. Because of their transmembrane nature, modelling a 3D structure of an insect OR has been impracticable until recently (Hopf et al. 2015).

Most ORs are broadly tuned to an extensive spectrum of odorants but few are narrowly tuned to specific volatiles. *Or19a* for example binds to terpenes found in citrus fruits, activating its oviposition circuit (Dweck et al. 2013); *Or56a* is invoked by Geosmin, sensing the toxic substrates, thereby modulating an avoidance behaviour (Stensmyr et al. 2012). Few other ORs such as *Or47b, Or67d, Or65a* and *Or88a* detect pheromones that trigger sexual and aggression behaviours (Kurtovic et al. 2007; Wang and Anderson 2010; Lebreton et al. 2014; Dweck et al. 2015). It is intriguing to note that ORs often work, along with OBPs, in a combinatorial fashion to distinguish the tiny fractions of odours from the natural environment to drive the olfactory circuit in modulating adaptive behaviour (Joseph and Carlson 2015).

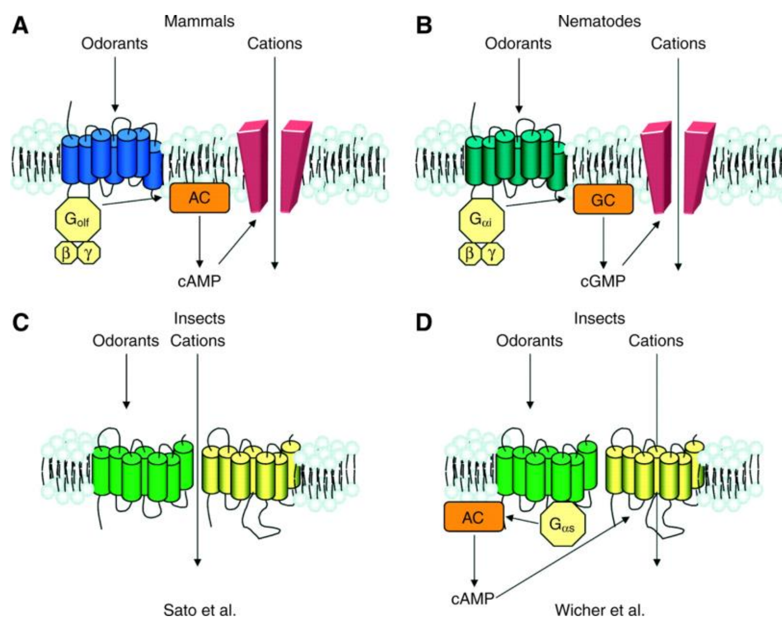


Fig 1.2. Comparison of the models of molecular mechanism in the Odorant receptors of A) Mammals B) Nematodes. They function as G-protein coupled receptors, activated by cAMP and cGMP pathways, respectively. C) and D) Insects – Two different models of functioning of odorant receptors. (Sato et al. 2008) suggests OR forms a ligand gated ion channel with Orco while according to (Wicher et al. 2008), OR forms a heterodimeric complex with Orco, creating a gated ion channel as well as activating a cyclic AMP pathway. Figure from (Pellegrino and Nakagawa 2009).

1.1.2.2. Gustatory Receptors

ORs and GRs share as less as 8% identity among them. ORs are insect-specific while GRs are found throughout Athropoda, making them relatively ancient; gustatory receptors are similar to ORs, both in their structure and molecular mechanism (they are inverted seven transmembrane proteins), but are non-homologous to other proteins. It is still unclear whether they use G-protein pathway or direct ligand gated channels, but, at least in Lepidoptera, they are believed to work like ORs in a combinatorial fashion (Zhang et al. 2011). Their three dimensional structure of these proteins could not be established so far.

There are 60 GRs loci/genes in *D. melanogaster* predicted to encode atleast 68 GR proteins, through alternative splicing (Clyne et al. 1999). The number of GRs in insects varies from 10 in bees to 200 in the beetle, *Tribolium castaenum.* Unlike the ORs, only half of the GRs have been functionally characterised so far in *D. melanogaster*: *Gr5a* works by activating the sugar-sensing neurons, eliciting a feeding behaviour, while *Gr66a* and *Gr93a* are involved in caffeine avoidance (Lee et al. 2009);*Gr43a, Gr64 (a-f)* family are required in detecting various sugars (Dahanukar et al. 2007; Fujii et al. 2015). Based on the spatial distribution and expression analysis, 33 GRs are localized and characterized into 4 different categories of bitter sensing neurons (Weiss et al. 2011), of which *Gr32a, Gr33a, Gr66a, Gr39a1* and *Gr89a* form the core of bitter sensing receptors (Lee et al. 2009; Moon et al. 2009; Weiss et al. 2011), that function analogous to Orco (Benton et al. 2006). Not all GRs are restricted to gustation. Four of them – *Gr32a, Gr33a, Gr39a* and *Gr68a* are involved with sexual behaviour through pheromone detection (Moon et al. 2009; Watanabe et al. 2011; Wang et al. 2011; Bray and Amrein 2003). *Gr28b* in *D. melanogaster* is involved in thermosensation (Thorne and Amrein 2008). Presence of atleast four other GRs in the antennae indicates their role in olfaction (Dunipace et al. 2001; Scott et al. 2001): among them, *Gr21a* and *Gr63a* are involved in perception of CO_2 (Jones et al. 2007; Kwon et al. 2007).

1.1.2.3. New classes of chemoreceptors: Ionotropic Receptors

The role of coeloconic sensilla in chemoperception has remained vague for years until recent studies found that these sensilla express a new type of receptors, called as ionotropic receptors (IRs) (Benton et al. 2009). Belonging to ionotropic glutamate receptor (iGluR) related gene family, they are categorized into two: Antennal IRs (aIRs) and divergent IRs (dIR). With the identification of *Ir25a* in various protostomes, IRs are now considered the most ancient class of chemoreceptors

(Croset et al. 2010). By using expression analysis, IRs have been linked not only to olfaction (aIR), as they are expressed in the olfactory neurons of the antennae (Benton et al. 2009), but also to taste for example in acid (Ai et al. 2010), ammonia, and amines sensing (Min et al. 2013). The IRs expressed in taste organs such as labellum, legs, pharynx and wing margins are very divergent, and hence called divergent IRs (dIRs) or gustatory IRs. While aIR are extremely conserved, dIR are fairly lineage-specific and share less than 8.5% sequence identity with the IRs of the same species or different species, suggesting a dynamic evolution related to the selectional pressure from occupying different geographical and ecological niches (Croset et al. 2010). Indeed, (Koh et al. 2014) show that a part of dIR gene family (*Ir20* clade; ~35 members) is expressed in GRNs, functioning as taste receptors, while *IR52c* and *Ir52d*, being involved in pheromone detection.

The origin of IRs can be traced back to the animal ionotropic glutamate receptors (iGluR), since both IR and iGluR share a common protein domain organisation. The IR gene family consists of as many as 66 members in flies (*D. melanogaster*) to 10 in nematodes (*Caenorhabditis elegans*) and 27 in molluscs (*Lottia gigantea*). Compared to dIRs, aIRs are very few, ranging from 17 in *D. melanogaster* to 2 in *L. gigantea.* There is no three dimensional structure of the IRs, but given their homology to iGluR, the ion channelling domains are expected to be identical in folding, to the predicted C-terminal of iGluR (Croset et al. 2010).

1.1.2.4. Binding Proteins – An Overview

Binding proteins- OBPs and CSPs are secreted in large quantities by the surrounding cells of the receptor neurons ORNs/GRNs. Although they have specific binding properties, their exact function remains unclear except for the role of LUSH (*Obp56a*) which, together with *Or67a* is involved with pheromone communication (Xu et al. 2005; Vosshall and Stensmyr 2005) and few others in taste perception (see below). All genes of these families are likely involved in transporting the odour molecules to the receptor sites present in the dendrites of the ORNs.

Similar to their OR-GRs counterparts, the Arthropod specific OBPs and CSPs are not homologous to the mammalian binding proteins, which are lipocalins (Tegoni et al. 2000). This is because, insect OBPs and CSPs form alpha-helical domain structures folded in two different patterns: OBPs characterized by 6 conserved cysteines, while CSPs by 4 conserved cysteines residues respectively (Angeli et al. 1999; Leal et al. 1999), connected by disulphide bridges. In Lepidoptera, OBPs are found to be antennal specific (Vogt et al. 2002) and more related to ORs, while CSPs are more

affiliated to GRs. The occurrence of GRs and CSPs in *Daphnia pulex* (Crustacean, which does not express ORs), and the characteristic of insect-specific ORs and OBPs gave rise to the hypothesis of OR-OBP and GR-CSP functional dependencies. This direct correspondence does not hold true always: in some hymenopterans (Calvello et al. 2003; Calvello et al. 2005), CSPs are found to be antennal specific. Also, the presence of both OBP and CSP in non-chemosensory tissues indicates other physiological functions (See review, Pelosi et al. 2014). *Obp57d* and *Obp57e* are involved in taste perception and host plant interactions and their evolution may have helped in driving the ecological specialisation in *Drosophila sechellia* (Matsuo et al. 2007; Matsuo 2008; Harada et al. 2012a).

OBPs in insects are very divergent, with less than 10% sequence identity (protein level) between species and sometimes, within a species (Pelosi et al. 2014) and are classified further into 9 sub-families based on function and phylogenetic relationship (Vieira et al. 2007). The number of groups of orthologous genes ranges from 34 OBPs and 3 CSPs within the genus *Drosophila* to 2 OBPs and 2 CSPs across Hexapoda. On the whole, only *Obp73a* and *Obp59a* are found across the whole Hexapoda which is highly suggestive of the reminiscent of the OR co-receptor, *Orco* (*Or83b*) (Vieira and Rozas 2011). However, this is currently under doubt, given the absence of *Obp59a* in *Apis mellifera* together with low level expressions in *Cerapachys biroi* (McKenzie et al. 2014), which might also be an independent case of gene loss. To date, three dimensional structures of various binding proteins have been predicted in Mosquitoes (Leite et al. 2009; Murphy et al. 2013) and Moths (Zhou et al. 2009) using X-ray and NMR techniques. Also, the crystalline structure of OBP LUSH/*Obp56a* and the binding pockets in *D. melanogaster* is very well studied, given its importance in pheromone communication and alcohol binding (Kruse et al. 2003; Thode et al. 2008; Ader et al. 2010).

1.1.3. The importance of chemoreception for applied studies

Insects are found everywhere, from unperturbed forests to agricultural lands and metropolitan areas. The insect-plant and insect-human interactions bring about both positive and negative implications for humans. While honey from bees and silk from moths have been used since ancient times, many infectious diseases are transmitted by vectors such as *Anopheline* mosquitoes that cause more than one million human deaths per year (Korenromp 2005). Likewise, in agriculture, insects are responsible for ~ 35% of yield loss (Pimentel 1991). The highly successful radiation of insects can

be partly attributed to their highly complex and flexible chemosensory toolkit. Chemical communication plays a vital role in shaping the insect's response to the environment: insects utilize their sense of smell and taste equally in survival and reproduction, wherein pheromones govern mating behaviour.

One way of controlling insects is to understand and manipulate how insects communicate among each other and with the external environment. There are few classical strategies using chemosensory-based control: i) Repellence, for example using DEET, a compound that binds receptor *Ir40* and is used to repel and manage various insects particularly *Anopheles* malarial vectors worldwide (Klun et al. 2004); ii) Mass trappings, for example of Tsetse flies in Africa using attractive chemicals based on cattle volatiles and urine; the same strategy applied in attracting Mediterranean fruit flies with lures such as ammonium acetate, putrescine and trimethylamine (Jannin 1999), or *Drosophila suzukii* using Droskidrink (Ioriatti et al. 2014), or ; the Colorado beetle, *Leptinotarsa decemlineata* –a major pest of potato crops in the Unites States, using specific pheromones (Kuhar et al. 2006); iii) mating disruption using pheromones, which has been highly successful for the European grape-berry moths, *Lobesia botrana* and *Eupoecilia ambiguella* or the codling moth *Cydia pomonella* (Arn and Louis 1997). Currently, other applied measures in use or under evaluation are vibrational disruptions, biological controls (ie.,parasitoids), sterile insect technique (SIT), and introduction of BT-crops: most of these methods are used without a clear understanding of their biological and genetic bases and/or are still inefficient or environmentally-hazardous (van der Goes van Naters and Carlson 2006).

The progress in the molecular genetics of taste and smell both at molecular and cellular level along with the rapid accumulation of –omics data provide us with new opportunities for tackling insect disease vectors and pests. For example, the knowledge of neuronal responses in identifying specific volatiles, combined with structure and evolution of receptor genes, can ultimately help in modelling the binding pockets for ligand-receptor study. For example, the Ray group used chemo-informatics to unveil the mechanism behind DEET avoidance by identifying its binding activity to olfactory receptor, *Ir40* (Kain et al. 2013). This can further help in finding safe but efficient alternatives to DEET, which mostly remains unaffordable in under-developed countries and moreover, is hazardous to human and animal health. For the near future, targeting chemosensation is a promising tool in controlling insect vectors and pests.

One of the aims of my thesis is to exploit the chemosensory system of *Drosophila suzukii* to find the target genes and specific ligands which are likely to have a profound impact on the behaviour of the pest.

1.2. Evolution, genomics, and chemoperception

1.2.1. On the origin and evolution of chemoreception in insects

1.2.1.1. Birth-and-death process of gene evolution

All major chemosensory gene families are highly variable in size across the arthropods, and there is a high level of divergence even between closely related species (Sanchez-Gracia et al. 2009); this is reminiscent of an involvement of these genes in adapting to specific ecological niches. Indeed, the insect chemosensory mechanism is one of the most highly dynamic systems where continuous gain and loss process is thought to follow a stochastic model of birth-and-death (BD) (Guo and Kim 2007a; McBride and Arguello 2007; Vieira et al. 2007; Sanchez-Gracia et al. 2009; Croset et al. 2010). According to (Nei and Rooney 2005), the BD model of evolution describe a scenario in which new genes are randomly lost or gained by gene duplication: depending on the selective forces involved, few duplications may be maintained and eventually get fixed in the population, whereas others are deleted or become non-functional through deleterious mutations.

In accordance with BD, most of the gene gains in *Drosophila* chemoreceptors are the result of tandem duplications, and hence are clustered in the genome (Vieira et al. 2007). On average, the level of birth and death is higher for the receptor genes than in OBPs. Not only, BD rates are higher in chemoreceptor gene family (λ=0.006 for ORs, λ=0.011 for GRs and OBP, λ=0.005 (McBride and Arguello 2007; Vieira et al. 2007; Gardiner et al. 2008)) than in the overall genome (λ=0.0012,)(Hahn et al. 2007). Gain and loss events are unevenly distributed across the *Drosophila* lineage. For instance, specialists such as *D. sechellia* and *D. erecta*, have lost significant OR and GR receptors compared to generalist species, suggesting a non-neutral evolution (McBride and Arguello 2007), indicating a possible role of chemosensation in ecological adaption. This is also reflected in the dN/dS analysis where the median ω for chemosensory genes is relatively higher (0.05 to 0.22) (Sanchez-Gracia et al. 2009).

1.2.1.2. The origin of chemosensation

Arthropods are the most diverse and successful phylum of animals that live or have ever lived on earth. They have conquered virtually all ecosystems including extreme environments such as Antartica and deserts, and were the first animals to colonize land, ~130 Million Years Ago (mya) before vertebrates (Anderson et al. 2013). By coupling genome data and fossil records, it was found that arthropods colonized land at least four times independently: the first terrestrialisation event was that of the myriapods (centipedes and millepedes at 510 mya) followed by that of arachnids and hexapods (insects and their kin) nearly synchronous with or even earlier than the first land plants (Ordovician; 470 mya); crustacean colonization of land (eg: by isopods) occurred much later (Rota-Stabelli et al. 2013b). We know from earlier works (Pelosi et al. 2006; Penalva-Arana et al. 2009; Sanchez-Gracia et al. 2009) that odorant receptors should have originated during the transition of organisms from water to land, as an adaptation to sense air-borne mediated chemicals, likely during the split of Crustacea/Hexapoda (Sanchez-Gracia et al. 2009). However, results from (Missbach et al. 2014) contradicts this general view and suggest a non-causal link between OR emergence and terrestrialization. Instead, the OR coreceptor (*Orco*) has been found in Zygentoma (silverfish), but not in earlier hexapod lineages such Archeognatha. This scenario set the origin of ORs within the flying insects (Paleoptera plus Neoptera) concomitantly with the emergence of vascular plants, rather than within the hexapods (Entognatha plus Insecta) concomitantly with terrestrialization events. It is not clear, however if the origin of a mature OR system (involving both *Orco* and specialized OR) occurred in the Neoptera or earlier in the ancestor of Palaeoptera, a question that I will address in my thesis. In any case, the ability to perceive odours on land must have evolved independently various times in Hexapoda, Myriapoda and Chelicerata either through a system based on ORs or using other type of receptors.

While ORs seems to be restricted to insects, GRs and IRs are present in other Arthropoda such as Crustacea (*Daphnia pulex*) (Penalva-Arana et al. 2009), Chelicerata (*Ixodes scapularis, Tetranychus urticae*) (Gulia-Nuss et al. 2016) and Myriapoda (Chipman et al. 2014), suggesting that these arthropod lineages rely on GR and IR for odour sensing, untangling, but also complicating the evolutionary history of peripheral arthropod chemosensory system. It is still unclear whether the vertebrate GPCRs and the insect chemoreceptors (OR and GR) originated from a common ancestor. Recent evidence points to a scenario where arthropod chemoreceptors might have originated from GUR (Gustatory-related receptors) receptors of nematodes and/or GRL (GR-like receptors) of Cnidaria (Robertson et al. 2003). Unlike ORs and GRs, the evolutionary history of

IRs can be traced to protostomia (550-850 mya), with the presence of the conserved antennal receptor, *Ir25a* in the entire Ecdysozoa, Lophotrochozoa and Cnidaria. Similarly, another aIR, such as *Ir93a* is present throughout the Arthropoda. As previously discussed IRs are closely related to the ionotropic glutamate receptors (iGluR). A likely scenario is that IRs evolved from an iGluR (involved in neurotransmission). The drastic, lineage based evolution of dIRs are attributed to retro-transposition, through the process of random re-insertions of reverse transcribed copies of the parental genes (Croset et al. 2010). The evolution of GRs and IRs along the Arthropoda phylogeny is poorly understood: for example it is clear that branchiopods, ticks and mites possess GRs, but nothing is known in other crustacean or chelicerate classes. For this reason in the last part of my PhD studies (one year spent at the University of Bristol with Davide Pisani) I have annotated a large dataset of GRs and IRs from various arthropods with fully sequenced genomes, and performed phylogeny of the GRs rooting them with GR-Like genes as outlined in section 4.4.2.

1.2.2. Comparative genomics in the post-genomic era

With cost-effective sequencing technologies and advanced development of software tools, it is now possible to sequence genomes de novo and perform basic comparative genomics in-house. While earlier genome sequencing relied on Sanger-based sequencing technologies, NGS (next generation sequencing) currently dominate the sequencing market with technologies including Illumina, Pacific Biosciences, Roche/454, ABI/Solid and Ion Torrent: all these technologies guarantee high throughput at reasonable cost. Choosing a platform can be tricky, as each sequencing technology offers different outputs. Among all, Pacific Biosciences (Eid et al. 2009) gives the longest of the read lengths (14 kb), but with a high error rate (Ribeiro et al. 2012). While ABI/Solid has a low error rate, it produces shorter reads (75 bp) with a longer run time, making it useful in identifying the transcribed regions in the genome (Horner et al. 2010). Though Roche gives longer read lengths (700 bp) and Ion Torrent has a high throughput (80-100 Mb/h), both their outputs are prone to homopolymer-associated errors (Loman et al. 2012). On the other hand, Illumina gives the highest throughput with least number of consensus error: for example, Illumina HiSeq 2500 offers high throughput (900 Gb to 1 Tb/ 6 days), reasonably high quality reads (2 x 250 paired-end reads; accuracy > 99.5%) for a cheaper price and in a relatively, quick time. For these reasons, we have used Illumina HiSeq for the whole genome sequencing (WGS) and transcriptome projects described

in my thesis. Therefore, given the ease of generating NGS data and the many new dedicated bioinformatics tools, it is not surprising at all, that as of March, 2016 there are 6,136 genome projects (inclusive of organelles) of animals in the databases, of which 975 are of insects. In 2011, i5K project was launched, that aims to sequence 5000 genomes of insects and other arthropods, initiated mostly towards insecticide and pesticide resistance (Robinson et al. 2011).

One of the biggest bottlenecks of NGS genome sequencing is the assembly of raw data. The rapid development in the sequence assembling resulted in a dramatic increase in number of assembly tools that follow three main paradigms: Greedy, overlap-layout-consensus, and De Bruijn graph. Each of these tools are tailored to specific sequencing platforms and projects. For example, assemblers based on the De Bruijn Graph such as Velvet (Zerbino and Birney 2008) and ABySS (Simpson et al. 2009) are more popular for the whole genome assembly, especially to suit the outcome of Illumina and SOLiD. Their functionality is inclusive of, but not limited to accurate shorter read fragments, produced from high coverage data sets. On the other hand, Trinity (Haas et al. 2013) and Oases (M.H. Schulz 2012) (based on Velvet) are widely used for transcriptomic data. In any case, one of the most common hindrance in sequence assembly and mapping are repetitive regions: when a read length is shorter than the repeat length (and maps to more than one region), the assembler is confused by this artefact. Hence, the accuracy of the data highly depends on how much an assembler is willing to tolerate the errors, given, the similarities of distinct regions (Nagarajan and Pop 2013). This can partly be overcome by in –depth sequencing to produce high coverage data. Moreover, to increase the accuracy and size of the assembly, sequencing paired-end or mate-paired data can help in producing overlapping pairs/mates that can be stitched together, which are in average, twice the length of the reads produced by the single-end sequencing. Although there are no means of predicting a correct genome assembly, most of the genomic studies assess the quality of post-assembly using N50 values, to approximately measure the contiguity of an assembly. N50 is the statistical measure of the reads that includes contigs that make up greater than or equal to 50% of the assembly size. Hence, lower the N50 value, less contiguous is the assembly. However, recent works suggest a non-correlation of N50 value to the actual quality of an assembly (Earl et al. 2011). Upstream of assembly, there are also popular pre-assembly editing tools such as fastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc) and fastx (http://hannonlab.cshl.edu/fastx_toolkit) toolkit to perform quality control that can help in improving and evaluating the quality of the data to be assembled. Once the assembly is retrieved, there are plenty of options to choose from various annotation tools that one can use, often depending on the source and size of the data (WGS/RNA-seq/single-celled). Irrespective of the

methods, a useful first analysis in genome assembly is masking the repeats, which refers to both low complexity regions and mobile genetic elements (Smit et al. 1996-2010).

The processed/assembled genome data can be annotated in two ways. An empirical approach predict genes using pre-existing gene models, which are used to train the intron-exon boundaries prediction (Yandell and Ence 2012). In a second approach annotation is *ab initio* and the software use mathematical models to predict the genes and their alternative splice-forms. However, even in the absence of reference gene models, aligning EST, RNA-seq and protein sequences to a genome can assist in training the algorithm as in case of MAKER (Holt and Yandell 2011) that combines AUGUSTUS (Stanke et al. 2004) and SNAP (Korf 2004). The next phase of gene prediction is annotation of specific gene families/orthology. This is also done in two ways: i) manually- which is labour intensive and slow, ii) automatically- using automated pipelines that can be less reliable but most commonly used, especially for smaller genome projects. Often, the manually curated datasets result in relatively high quality annotations (Misra et al. 2002; Yandell et al. 2005), as in the case of the human genome project. In general, orthologous gene identification works well only with reliable information from closely related reference species. In the absence of a reference the risk is to find misorthologs, false positives, and false negatives. Extremely gapped/incomplete genome assemblies clearly worsen the correct identification of orthologs. In general, there is always a delicate trade-off between specificity and coverage/depth/completeness of the genome sequencing: A clear growing problem in comparative genomics is that, for economic reasons, low coverage/poorly assembled genomes are preferred, causing errors in gene identification (Yandell and Ence 2012).

To make sense of genes annotated in genomes, it is essential to study them on an evolutionary framework. The most important issue is llikely the comparison with closely related species, commonly called as ‘sister species’. This allows identification of genes that are shared between the species of interest and the sister species: such genes are not unique and are likely less important in describing the peculiar biology of the species of interest. Conversely, genes that are unique are good candidate to play a role in determing the phenotype of the species of interest. Divergence times may also play an important role in making sense of comparative genomics because they allow inferring the rate of gene gain and loss in the various lineages, as I have done when studying genes on a Birth and Death framework (see Results; section. 4.3.1). Furthermore, proper outgroup choice plays a significant role, as it can influence the structure of the phylogenetic tree based on its rooting position, and further help in polarising genomic event on the ingroup species gene tree.

1.3. *Drosophila suzukii*- an invasive pest in Western countries

With increased global trade, western markets import huge quantities of goods from tropical regions such as South America and South Eastern Asia. Occasionally, such goods are infested with invasive species that easily gets disperse: one of those is *Drosophila suzukii* Matsumura (Diptera: Drosophilidae, commonly known as SWD - Spotted Wing Drosophila) , a fruit fly endemic to South East Asia, first reported in Japan in 1930s (Kanzawa 1939; Mitsui et al. 2010). In 2008, arrival of the species in the US and Europe was reported synchronously in California, Spain and Italy, mainly near the ports (Walsh et al. 2011; Cini et al. 2012; Rota-Stabelli et al. 2013a). After a decade, *D. suzukii* has now spread to UK, and Scandinavia in North East Europe to Serbia in the west, thus showing a high dispersal potential. Also, it is well established in North America, Canada and in some parts of South America (Asplen et al. 2015). Further reports are expected to see a trend in the spreading of the pest in other unconfirmed territories. Like *D. melanogaster*, most species of *Drosophila* are attracted by, feed and oviposit only on rotten substrate. *D. suzukii* similarly feeds on over ripened or decaying fruits at adult stage, but it oviposits on fresh, soft and stone fruits causing a menace to fruit crop production. *D. suzukii* pierces the fruits using the serrated ovipositor (fig 1.3) (Kanzawa 1939; Walsh et al. 2011), an evolutionary trait shared with two other species, *Drosophila subpuchrella and Drosophila pulchrella*. However, these two species are not reported to produce any economic damage to the crops in their native geographical location (Mitsui et al. 2010). The infestation of *D. suzukii* leads to the primary problem of larval feeding inside the fruit pulp. However, what makes the fruit inedible is the infection caused by fungi and bacteria from the pierced and damaged skin of the infested fruit. This led to unmarketable fruits and huge agricultural losses in the dispersed regions. For example, severe economic damage was registered in Trentino province of Italy in Europe, losing about 3 million € in 2011 alone (Ioriatti et al. 2015a). In US, pacific coastal States suffered atleast 500 million $ annually, which formed a whopping 76% of total soft fruit production (Walsh et al. 2011). Although cherries and berries are preferred substrates, the host range of *D. suzukii* is very wide with infestations reported on pears, apricots, figs, tomatoes, peaches and, more threatening for Italian economy, to some variety of grapes (Rota-Stabelli et al. 2013a; Ioriatti et al. 2015b).

Fig 1.3. Serrated ovipositor of *D. suzukii.* Figure from Department of Entomology, University of Florida.

Monitoring and management of the pest have relied on emergency measures to effectively control the population, mostly resorting to pesticides. These systems however can leave hazardous chemicals on fruits when treatments are close to harvest (Rota-Stabelli et al. 2013a). Hence, other alternative and eco-friendly strategies such as mass trappings, mating disruptions, biological controls using parasitoids are being applied and/or studied to tackle the spread (Cattel et al. 2016; Rossi-Stacconi et al. 2016). Studying the chemosensory repertoire in *D. suzukii* has the potential of providing genes and ligands that can be used for downstream control strategies based on repellents and attractants.

Given the significance of the current status of pest management, two draft genomes of *D. suzukii* have been sequenced and assembled, one from Italy (discussed in my thesis) and another from North America. (Ometto et al. 2013) (That I have co-authored) confirmed that *D. suzukii* is well nested within the *Drosophila* clade, and further in the melanogaster subgroup. *D. melanogaster* is one of the most studied model organisms and a carefully mapped genome, refined gene annotations coupled with experimental data can significantly help in understanding the biology of *D. suzukii*. Also, thanks to the availability of the *Drosophila biarmipes* genome, comparing the genome of *D. biarmipes,* a non-pest, sister species, can be a good strategy in exploiting the uniqueness of *D.*

suzukii. In fact, doing comparative genomics, particularly of chemosensory genes, could be a fundamental step for downstream applied research.

The shift in preference for ripe fruits in *D. suzukii* also offers a unique possibility for comparative evolutionary studies on the adaptive origin of new ecological and behavioural traits. Throughout the last decades, *Drosophila* proved to be an excellent model organism for a wide range of studies, from sexual selection to shifts in food preference (Dekker et al. 2006; McBride 2007; Stensmyr et al. 2012). Moreover, the availability of rich literary resource based on physiological findings of *D. melanogaster* chemoreceptors, can be exploited (Muench and Galizia 2015) to interpret *D. suzukii's* behaviour. The first part of the thesis that I present here is one of the very few works to combine bioinformatic and experimental data to demonstrate the ecological adaption in a *Drosophila* species. Studying the evolution of the chemoreceptor genes in *Drosophila suzukii* as an insect pest and evolutionary model (where we have sequenced both its genome and transcriptome), using comparative genomics can help to unveil its ecological adaption. Further, the result of this work has provided with a list of candidate genes (chemoreception and pheromone synthesis), involved in the ecological shift in the behaviour of *D. suzukii*. Most of these receptors/genes, and the respective neurons where they are expressed will become the targets of detailed neurophysiological and behavioural experiments for further downstream applications.

2. Main objectives

2.1. To unveil the evolutionary dynamics of chemoreception in a pest species, using *D. suzukii* and its genome as a model.

I shall contribute in assembling and curating *D. suzukii* genome using up to date protocols and methods. I shall then focus on the annotation of genes involved in chemosensation (ORs, GRs, aIRs, OBPs, CSPs) and in their analyses under an evolutionary framework. The goal is to unveil uniqueness in the repertoire of chemosensory genes in *D. suzukii*, and link this data with phenotype in search of likely adaptive genomic changes that would explain some of *D. suzukii* peculiar biology. Genes of interest, as well as the chemical ligands associated with them, may ultimately become target of downstream applied research.

2.1. To increase our understanding of the origin and evolution of chemosensory receptors in Arthropod.

For doing so, I shall search various recently released genomes of hexapods, crustaceans, myriapods and chelicerates in search for ORs, GRs, and IRs. Careful annotations of these receptors may clarify their distribution on the species phylogeny and shed light on some long lasting questions such as the origin of smell in insects.

3. Methods summary

I outline here a brief overview of methods I have used during my studies. I have summarised them in the bioinformatic pipeline of figure 3.1 which depicts a flow of experiment from genome/transcriptome assembly and gene annotation, to phylogenetic and birth and death (BD) analysis.

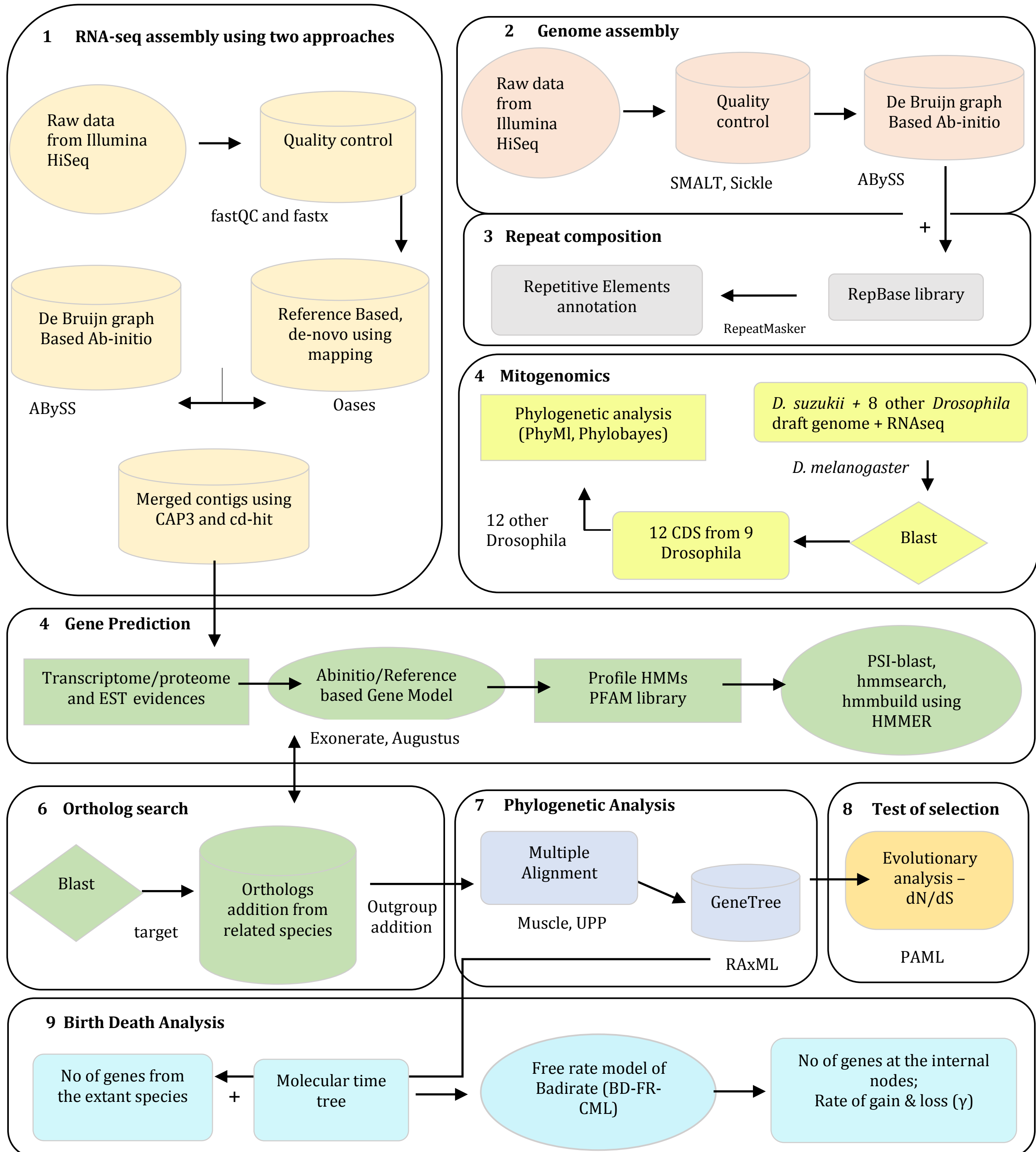


Figure 3.1.The pipeline of bioinformatics experiments I have conducted in my thesis.

For each of the projects of my thesis (See Results), I followed one or more sections of this pipeline. Here I provide a summary of each of section while a complete description is provided in the methods section of each of the attached manuscripts.

RNA-seq assembly

Data quality was evaluated with fastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/, last accessed April 3, 2013) and Tallymer (Kurtz et al. 2008). Low-quality positions were trimmed using fastx (http://hannonlab.cshl.edu/fastx_toolkit/, last accessed April 3, 2013) with a threshold of 0.3. I assembled the resulting 30,951,598 read pairs using two distinct approaches. First, we used Oases (M.H. Schulz 2012) with k-mers ranging from 25 to 53, obtaining 24,358 contigs (length 100–15,000 bp). In the second approach, I used ABySS (Simpson et al. 2009) with k-mer 45 and obtained 140,736 contigs. The two sets were merged using cd-hit (Li and Godzik 2006) with an identity threshold of 100% and eventually super-assembled using CAP3 (Huang 1999) using default settings.

Genome assembly

Unlike RNA-seq data, the genome was cleaned for Wolbachia and Mitochondria using Smalt (http://www.sanger.ac.uk/resources/software/smalt) by mapping the reads using genomes of five *Wolbachia* strains (*W. ananassae, W. melanogaster, W. simulans, W. willinstoni*, and *wRi*) and the *D. melanogaster* mitochondrial DNA (mtDNA) respectively. The resulting reads were further cleaned using Sickle (https://github.com/najoshi/sickle), after which both the 180bp and 300bp libraries scored, on average, a qvalue of 35. We used ABySS (Simpson et al. 2009) to assemble the resulting 20Gb raw data with k-mer size ranging from 48 to 64 (Table 3.1).

K-mer size	N contigs	n:200	n:N50	N80	N50	N20	max	sum (Mbp)
48	1,399,155	93,256	7,826	1,369	4,756	18,300	208,969	185.3
54	1,200,237	105,190	9,204	1,283	4,445	15,559	169,965	195.7
64	**961,286**	**131,597**	**12,820**	**1,089**	**3,565**	**11,309**	**169,947**	**209.6**

Table 3.1. Genome assembly statistics. Abyss trials with different k-mer size. n:200 is the number of contigs shorter than 200 bp, n:N50 is the number of contigs longer than the median, N80 is the size of the 80 percentile, N50 is the median contig size, N20 is the size of the 20 percentile, sum is the overall contigs size in millions of base pairs.

After quality assessment of the assemblies, we retained as best assembly, the one obtained using a k-mer of 64. Choosing a k-mer is often tricky, as it is a trade-off between the accuracy and repeat-identified problems. K-mer is influenced by factors such as assembly size, assembly errors and N50 value. We chose K-mer of 64 as it gave the maximum size of assembly, with an acceptable N50.

Repeat composition

I studied the composition of repeat elements in *D. suzukii* and 12 other *Drosophila* using RepeatMasker (Smit et al. 1996-2010). I analysed the entire genomes without distinguishing between euchromatin and heterochromatin partitions, as these information are either incomplete or unknown for most of the *Drosophila* species used in this study. I used all fragments irrespective of their length, because the *D. suzukii* genome assembly and some of the other draft genomes contained many contigs shorter than the 200 kb limit recommended (Consortium 2007). I quantified the presence and size of repeats as the percentage of repeated sequences over the draft genome size. This approach has the advantage of reducing biases due to the uncertain draft genome size of the different species, which may vary due to the different assembly strategies and/or genome quality levels, and may not reflect the actual genome size. To account for this inaccuracy, we further calculated the percentage of total repeats using two contrasting and conservative estimates of the putative average *Drosophila* genome size (a minimum at 130 Mb and a maximum at 180 Mb).

Mitogenomics and Phylogenomics

I used *D. melanogaster* mitochondrial DNA as a reference to extract the CDS in *D. suzukii* and 8 other *Drosophila* (*D. biarmipes*, *D. bipectinata*, *D. elegans*, *D. eugracilis*, *D. ficusphila*, *D. kikkawai*, *D. rhopaloa*, and *D. takahashi*). Comparison with both assembled genome and transcriptomes yielded partial sequences, and revealed several putative NUMTs ('nuclear mitochondrial DNA sequences', portion of mitochondrial DNA that have been transferred into nuclear genome). The nearly complete mitogenome of *D. suzukii* was assembled creating a consensus between genome and transcriptome data. The sequence of the 13 mtDNA protein coding genes from *D. suzukii* and from 8 *Drosophila* above were then aligned using MUSCLE with the orthologues from 13 other *Drosophila* with an annotated/published mitogenome. For phylogenetic analysis, we further assembled a concatenated alignment of 91 nuclear coded CDS extracted from the transcriptome of *D. suzukii* and other 21 *Drosophila* (details are in Ometto et al. 2013). Both datasets were processed into three different ways to make three types of datasets: the first one contained all 3 nucleotide positions; the second had the 3rd codon position removed; the third contained the corresponding amino acid data. Further, we inferred phylogeny from each of these

datasets using different phylogenetic frameworks (Bayesian and ML), both homogenous and more sophisticated heterogeneous models such as CAT + GTR on a Dayhoff recoded data set as in (Rota-Stabelli et al. 2013c).

Gene prediction and Ortholog Search

To annotate genes I have used ab-initio algorithms such as Augustus in combination with PSI-blast (Stanke and Waack 2003) (Altschul et al. 1997). Moreover, I performed individual manual blast searches to identify and annotate genes both in *D. suzukii* and *D. biarmipes* (more details are in the manuscript, Ramasamy et al, 2016). For the GR and IR genes of the Arthropoda, I used tblastn extensively along with evidence-based gene model software Exonerate (Slater and Birney 2005). The resulting annotations were used to search the genomes using hmmsearch and hmmbuild with HMMer (Mistry et al. 2013).

(Gene) Phylogenetic analysis

To help in the annotation of the *Drosophila* gene trees (OR, GR, OBP, CSP, aIR), I constructed six-species gene phylogenies. The gene and protein sequences for *D. erecta, D. ananassae and D. pseudoobscura* were downloaded from FlyBase (Drysdale et al. 2005) and their orthologous relationships predicted by OrthoDB (Waterhouse et al. 2013). I then added the orthologues of *D. melanogaster*, *D. suzukii* and *D. biarmipes*, and built multiple sequence alignments at both nucleotide and protein level for each of the 3 families with MUSCLE (Edgar 2004b) using TranslatorX (Abascal et al. 2010). The genome sequences for various arthropods were downloaded either from i5k genome project ((Consortium 2013); https://www.hgsc.bcm.edu/i5k-pilot-project-summary) or European nucleotide archives at EBI (http://www.ebi.ac.uk). However, for the GR and IR identification in Arthropoda, I used UPP (Ultra-large alignments using Phylogeny-aware Profiles) (Nguyen et al. 2015), which is quite useful for highly diverged proteins like chemoreceptors. Also, the dataset was huge and filled with fragmented sequences, as in most cases the gene prediction cannot find the initial N-terminal regions. The profile-based alignment from UPP was aligned further with Muscle (Edgar 2004a) and manually cleaned for gap-filled sites. For all the alignments, phylogenetic analysis was done using maximum likelihood framework, with 100 bootstrap replicates, configured in RAxML v.7.2.8 (Stamatakis 2014). Since the input sequences were amino-acids, protein models of replacement- GTR (Waddell and Steel 1997) or one of its empirical versions, LG (Le and Gascuel 2008), were used to compute tree search, together with a

Gamma distribution with four discrete categories and an empirical estimation of amino acid frequencies (F).

Signatures of selection

We used PAML to calculate ω (dN/dS) for each of the coding proteins involved in chemoreception. While the ratio of synonymous substitution to non-synonymous substitution (dN/dS) of zero indicate purifying or negative selection, dN/dS above one is an evidence of positive selection. Under free ratio models implemented in codeml, we calculated branch specific rates (model=0, NSsites=0; model=2, NSsites=0). Also, to identify *D. suzukii* specific positive selections, we used branch-site specific model (model=M2a, NSsites=2) to test the occurrence of affecting sites. Prior to the test, alignments were manually curated for frame shifts (See Methods in Ramasamy et al 2016). I did not check for signatures of selection in the fragmented, partial dataset of Arthropod chemoreceptors.

Birth Death Analysis

I mapped the evolution of each of the gene families (OR, GR, aIR, OBP) on a 14 *Drosophila* species trees using the tree topology proposed by (Ometto et al. 2013). I estimated the gene family size at each internal node and family turn-over rates for each branch by using stochastic models implemented in BadiRate version 1.35 (Librado et al. 2011). The program uses the information of the divergence time and the number of genes at the extant species over a phylogenetic framework. I used the BDI-FR-CML model where the probability of genes at each internal node is modelled by maximum likelihood assuming that each lineage evolves independently. I have also predicted the rate of expansion or contraction on a time tree using the gene gain and loss information at each branch as: Rate of Expansion/Contraction = No of gains + No of losses / Divergence time in mya.

4. Summary of results

The work from this thesis has contributed to 4 publications in international peers review journals (included in Appendix B). Here I present a synthetic summary of these contributions. These results include mainly my own work. However, because of the strong interdisciplinary nature of my thesis project, some sections describe work that I have conducted together with some colleagues of mine: in such cases I have used the plural form (we) in presenting results. In few circumstances, I have further presented work done entirely by colleagues of mine; this is because these results were an essential complement to my own results: in these cases I explicitly referred to "my colleagues".

4.1. Genome and Transcriptome of *D. suzukii*

To aid gene annotation and ease understanding the biology and the evolution of *D. suzukii*, we sequenced both its genome and transcriptome. I contributed in cleaning and assembling both genome and transcriptome, I compared the repeat composition, annotated mitochondrial genomes, and helped in using genome data to clarify the phylogenetic affinities of *D. suzukii*.

4.1.1. Sequencing and Assembly of genome and transcriptome

We sequenced and assembled a draft genome and transcriptome of *D. suzukii* from an Italian Alpine population. Assembly of the nuclear genome was performed using both 180 bp and 300 bp libraries. The 180 bp library generated 67,153,264 100 base read pairs totaling 14.3 gigabases (Misof et al.) and the 300 bp library 51,792,255 100 base read pairs covering 10.4 Gb. Contigs that are longer than 1 kb have been submitted to the European Nucleotide Archive at EBI web site (http://www.ebi.ac.uk/, last accessed April 3, 2013) with the accession numbers: CAKG01000001–CAKG01061569. The draft genome was sequenced to high depth (an average of 80× coverage) and comprises 49,558 contigs spanning a total of 160 Mb. The RNAseq sequencing generated a total of 35.7 million 100 base paired reads. The final assembled data set consisted of 25,810 putative transcripts with lengths varying from 50 to 16,500 bp. The size of the genome and transcriptome are comparable with that of *D. melanogaster* and other sequenced *Drosophila.*

4.1.2. Comparative genomics of repetitive elements in *D. suzukii*

Results show that nearly 9% of the *D. suzukii* genome is composed of repeats: this is comparable to amount of repeats found in *D. biarmipes* and *D. takahashii*. However, the association between the phylogeny of the *Drosophila* species and the composition of transposable elements is not congruent for few closely related species. For example, *D. simulans* has only a quarter of the *D. sechellia*'s repeat composition (See fig.4.1). This might be due to the lineage-specific effects of horizontal gene transfers (Biémont and Cizeron 1999). The two classes of repeats, retrotransposons and DNA transposons, differ in their functional mechanism of transposition. With the presence or absence of long terminal repeats, retrotransposons can be further distinguished into LTR and non-LTR families which differ in their abundance of distribution in *Drosophila*. LTR which are more conserved than the non-LTRs, are found higher (fig. 4.2) in *D. suzukii*.

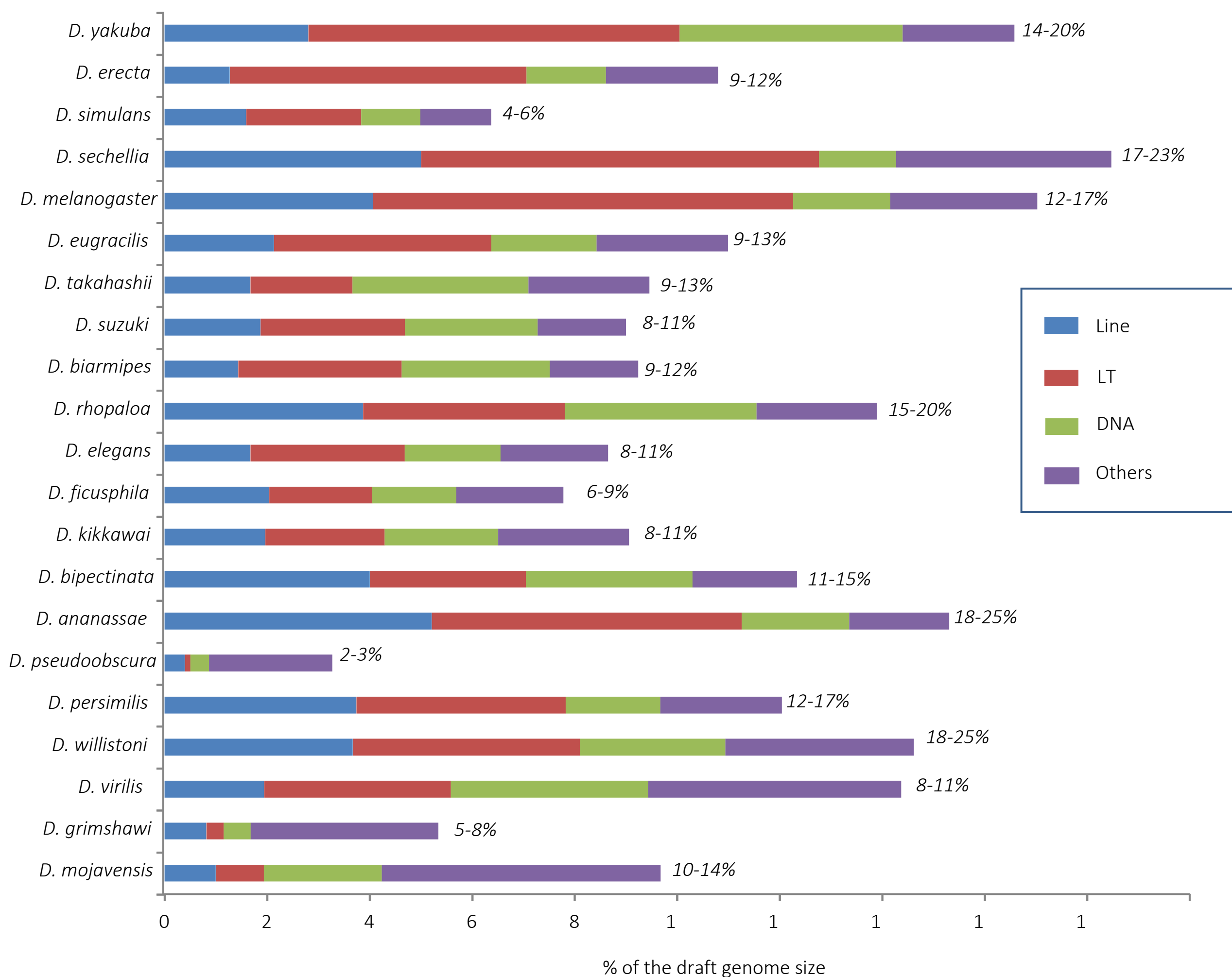


Figure 4.1. Repeat elements in *D. suzukii* genome. The distribution and number of repeats in *D. suzukii* is similar to that of sister species *D. biarmipes* and *D. takahashii*, consistent with their phylogeny.

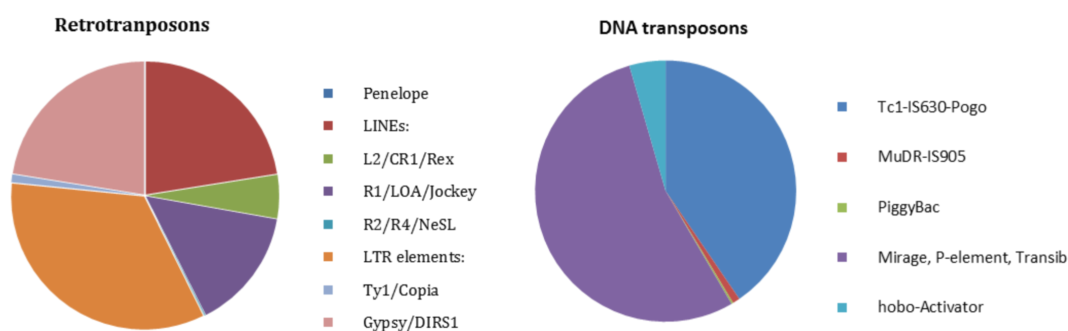


Figure 4.2. Composition of Retro-transposon and DNA transposon in *D. suzukii.*

High occurrences of gypsy and copia elements and absence of Bel/Pao and SINEs family is in line with the characteristic of most *Drosophila* species. The total and family-wise composition of repeat elements in *D. suzukii* is similarly shared by the closely related species characterized by a different biology: henceforth, it is possible to exclude repeat enrichment as a reason behind the peculiar biology of *D. suzukii.*

4.1.3. Incongruence between mitochondrial and nuclear phylogeny

Comparative genomics relies on a robust phylogenetic tree, which should be used to polarize evolution of genes. For this reason, we used data from the *D. suzukii* mitogenome and transcriptome to conduct a comprehensive multi-locus genome scaled phylogenetic and dating analysis using 20 additional *Drosophila* species. We conducted two separate analyses using two distinct datasets, a mitogenomic one and a transcriptomic one. For the mitogenomic dataset, I have assembled *D. suzukii* mtDNA in 15 contigs, spanning 14,736 bp; from this mitogenome, I extracted coding sequences and generated an alignment of 21 *Drosophila* species. Colleagues of mine generated a 91 nuclear coded protein-coding genes extracted from the transcriptomes of the same 21 species. All analyses converged on a tree that confirmed a sister relationship between the *suzukii* and *takahashii* subgroups, but failed to converge on some internal relationships of *Drosophila*. This is most likely because of a lack of phylogenetic signal in the mtDNA. For example, the apparently robust bootstrap support (97%) in mtDNA phylogeny against the sister relationship *eugracilis-melanogaster* subgroup (supported instead by nuclear dataset) vanishes when highly saturated third codon positions are excluded, or when an amino acid data set was employed; this indicate that signal contradicting the nuclear phylogeny carried by mitochondrial genomes is concentrated in unreliable (likely saturated) third codon positions whose substitutions for the most part are synonymous (Ometto et al. 2013).

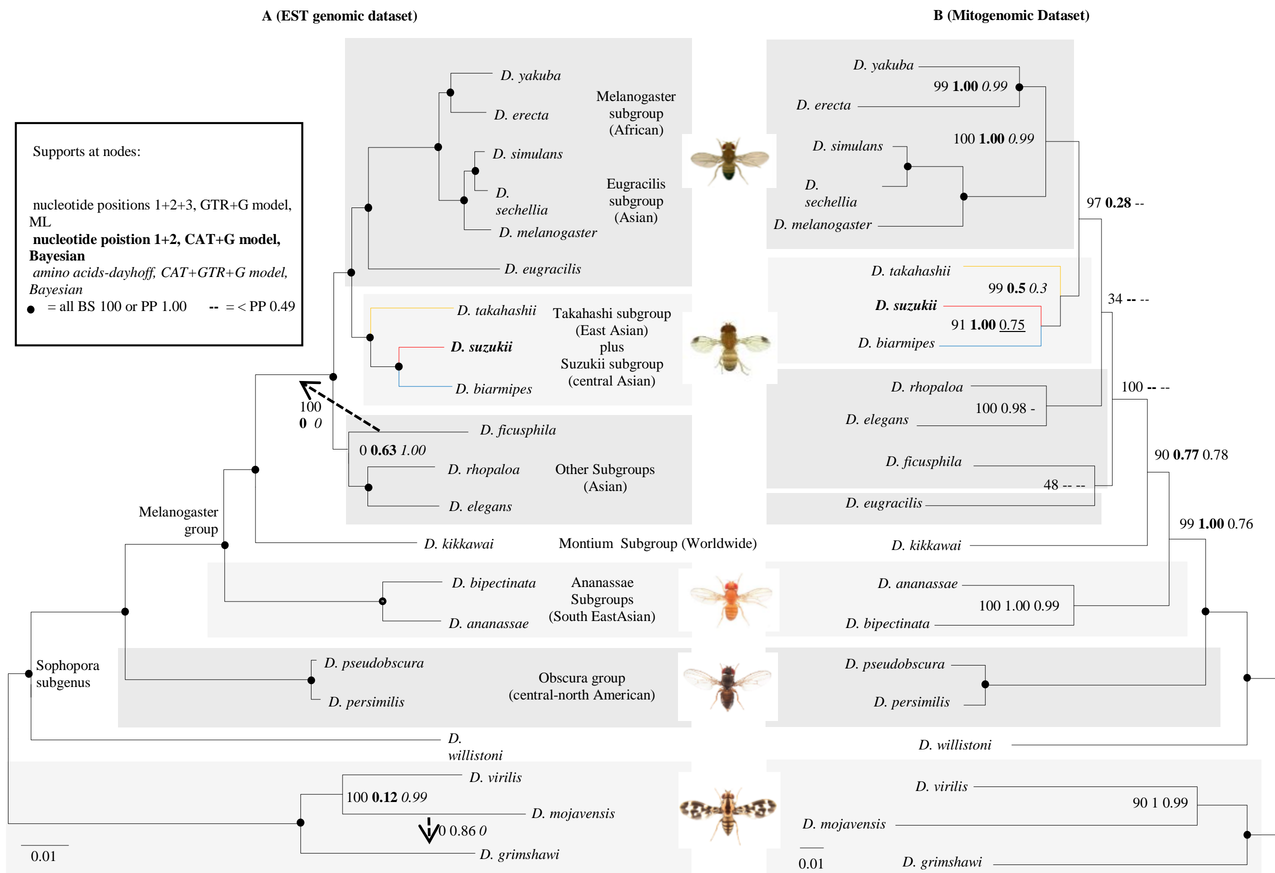


Figure 4.3. The evolutionary affinities of *D. suzukii* and the other *Drosophila* species inferred from phylogenomic and mitogenomic data. (*A*) Phylogenetic analyses of 91 orthologous nuclear genes. (*B*) Phylogenetic analyses of 12 mitochondrial genes. Both data sets support an Asian affinity of *D. suzukii*, and a sister relationship with *D. biarmipes*.

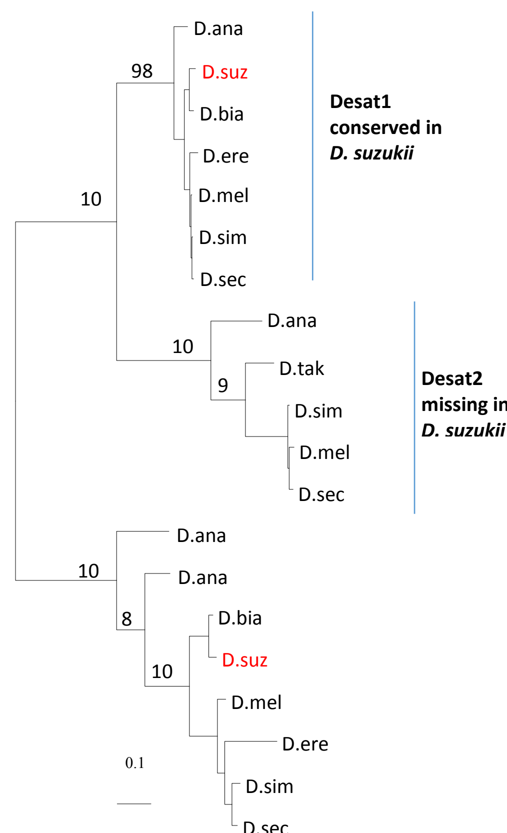
4.2. Investigating the evolution of sexual pheromone perception in *D. suzukii*

The power of *D. suzukii* genome data resides in being a useful repository for quickly obtaining orthologues for comparative studies aimed at understanding its ecology, biology and behavior. Apart from olfaction and gustation, pheromone detection is a key aspect of sexual communication and hence, reproductive behavior. Long chain hydrocarbons, present in the cuticle and known as cuticular hydrocarbons (CH), play a major role in insect's reproduction. By combining gas chromatography and mass spectrometer, some of my colleagues measured the CH profile of *D.*

suzukii for male and female, separately. Apart from minor differences, both sexes exhibited isomorphic CH profile. However, the CH profile of *D. suzukii* surprisingly shows lack of cVA: Among the CH, *cis*-11-octadecenyl acetate (cVA) is the best studied pheromone communication system (Symonds and Wertheim 2005). cVA increases mate acceptance in females, reduces attractiveness of newly mated females, reduces male–male courtship, while increasing aggression between males (Wang and Anderson 2010). cVA is found throughout the *Drosophila* lineage, even in basal ones like *obscura* and *immigrans*; hence, its complete absence in *D. suzukii* males is striking given its prominent role in both sexual and social behavior. In accordance to that my colleagues found that the ejaculatory bulb, where the cVA is produced has significantly shrunk in volume, compared to *D. melanogaster.*

In an attempt to find the molecular bases of cVA absence in *D. suzukii*, I have searched the *D. suzukii* genome for various Desaturase and Elongase gene families involved with the biosynthesis of species-specific cuticular hydrocarbons in *D. melanogaster*. Desaturases are essential for lipid metabolism and to maintain the structure and function of biological membranes, and in insects are further involved with the biosynthesis of cuticular hydrocarbons and pheromones. Although there are 8 Desaturases in the gene family of *D. melanogaster*, so far only three have been implied in biosynthesis of hydrocarbons and mate recognition: *desat1*, *desat2* and *desatF* (Roelofs and Rooney 2003) (Chertemps et al. 2006).

A



B

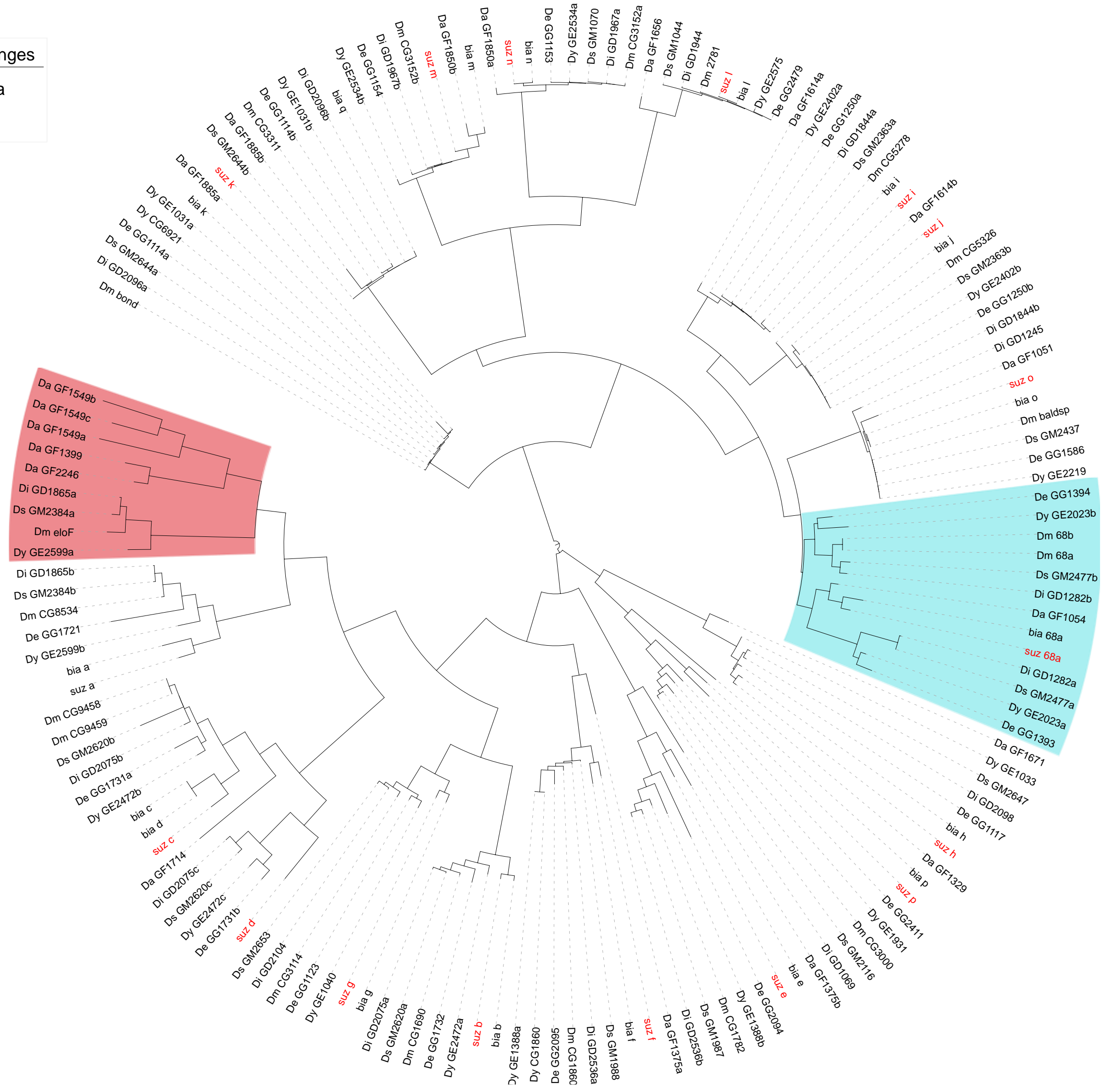


Figure 4.4. Phylogenetic analysis of (A) Desaturase family, where *D. suzukii* misses *Desat2* gene and (B) Elongase family, where two key genes *EloF* (red shade) and *Elo68a* (blue shade) are highlighted. EloF is missing in *D. suzukii*, while Elo68a being conserved. The tree is rooted using midpoint.

I have extracted, and annotated these genes in *D. suzukii* and *D. biarmipes* and conducted phylogenetic analyses to understand their evolution in *D. suzukii*. Results show that while *desat1* and *desatF* are conserved throughout the *Drosophila* phylogeny, but that *D. suzukii* and *D. biarmipes* have lost *desat2* (See fig.4.4*A*). I further conducted a comprehensive analysis of Elongase orthologs in various *Drosophila* species and I found a complex scenario in which *EloF* and *Elo68a* are part of a large family of Elongases. We have named these putative Elongases using alphabetic letters from A to Q, leaving the F to *EloF*. Most of these putative Elongases are

conserved. *Elo68a*, which is directly involved in cVA production, is normally present in *D. suzukii*, but *EloQ* and *EloF* are missing in *D. suzukii* and *EloC* is likely duplicated. (see fig.4.4*A*, red shade in fig. 4.4*B*). *EloF* is mainly involved in female pheromone synthesis and courtship, and in *D. simulans*, non-expression of *EloF*, along with *DesatF* is thought to have played a role in species isolation and in turn, speciation (Chertemps et al. 2007). *EloF* is a candidate gene involved in the loss of cVA production in *D. suzukii.*

In *Drosophila,* cVA is recognized by *Or67d*, a receptor expressed on T1 neurons of trichoid sensilla, which innervate the glomerulus DA1 present in the antennal lobe (Kurtovic et al. 2007). My colleagues found that *D. suzukii* has less number of T1 sensilla compared to other *Drosophila*, in accordance with a reduced sensitivity for cVA. On the other hand, the T4 sensilla that houses *Or65a* neurons are more abundant in *D. suzukii*. In *D. melanogaster*, it is shown that *Or65a* is involved in suppressing *c*VA-mediated male-male aggression and decreasing receptivity towards recently mated females (Liu et al. 2011). Also DL3, the section of glomeruli that innervates T4 sensilla in *D. suzukii* is significantly enlarged. To understand if this difference was accompanied by molecular variations, I have extracted and analyzed *Or67d* and *Or65a* and the results show that these genes are well conserved in *D. suzukii* (fig. 4.5).

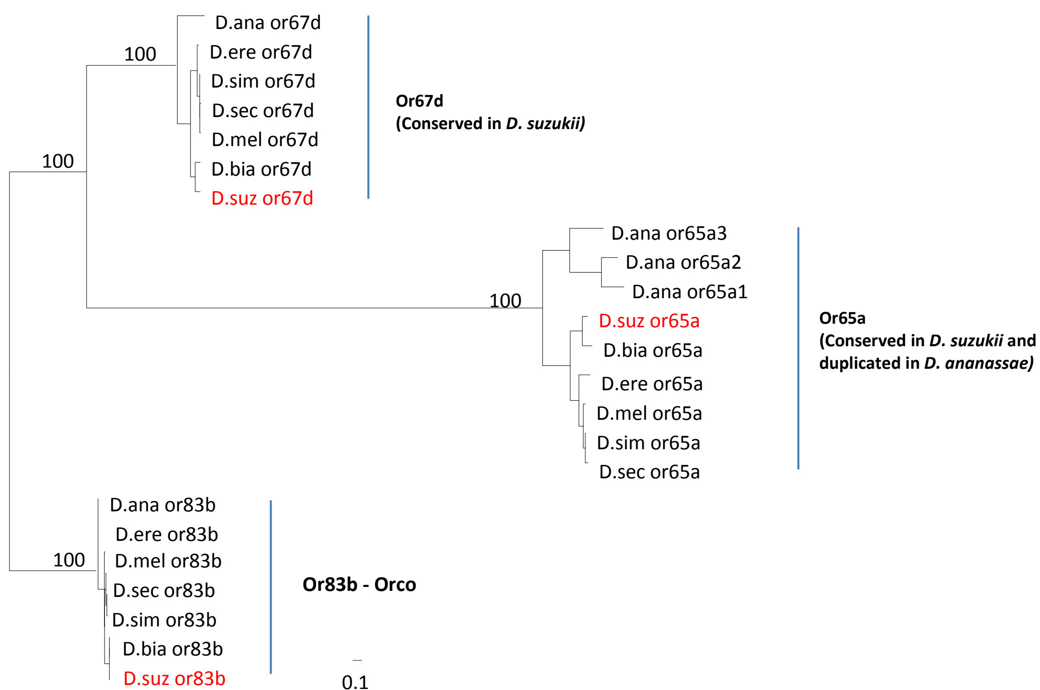


Figure 4.5. cVA odorant receptors are conserved in *D. suzukii*. Phylogenetic tree of *Or67d* and *Or65a* in *D. suzukii* and other *Drosophila*. These genes are extremely conserved in *D. suzukii* and the other species sampled. This suggests that these genes are indeed expressed in neurons of T1 and T4 sensilla and are structurally constrained to recognize *c*VA, as indicated by the electrophysiological responses. The tree is rooted with the OR co-receptor, *Orco*.

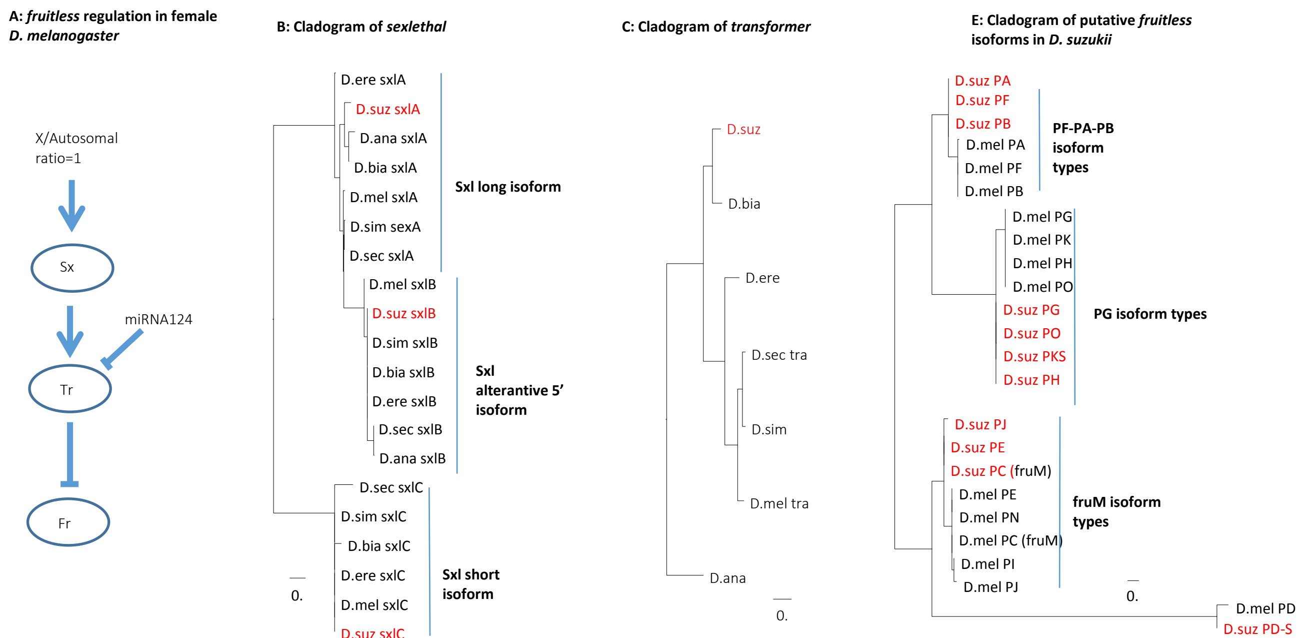


Figure 4.6. Fruitless/*c*VA signalling pathway is conserved in *D. suzukii.* A) Simplified scheme of the signalling cascade controlling fruitless mediated sexual characterization in *Drosophila*. B) *sexlethal* is present with both its two main isoforms in *D. suzukii*, which however miss an extremely conserved region at aa 332 C) *Transformer* is also conserved in both female and male isoforms in *D. suzukii*. One of the binding site of miRNA124 is variable in *D. suzukii*, but still capable of binding miRNA124 (see panel D). D) *Transformer* gene of *D. suzukii* contains both miRNA-124 binding sites. Left shows the binding site at 3'UTR of *transformer* in *D. melanogaster* (upper) and *D. suzukii* (lower). Right panel shows the binding site at the last *transformer* exon. E) A cladogram of putative *Fruitless* isoforms in *D. suzukii* compared with known isoforms in *D. melanogaster*, with each possible exons aligned. All the alignment from which trees have been constructed are available for downloading at http://dx.doi.org/10.6084/m9.figshare.865654.

It can be interpreted that both these genes are indeed expressed in T1 and T4 sensilla respectively, structurally constrained to recognize cVA. Given such a switch in the size between the two glomeruli of T1 and T4 sensilla of *D. suzukii*, it's possible that the cVA played an antagonistic role in mating behavior.

I have further analyzed a range of transcription factors involved with sexual behaviors and response to cVA in *D. melanogaster* (*fruitless (fru), sexlethal (sxl), transformer (tra) and doublesex (dsx)*). I found *Fruitless* (Hall 2002) to be conserved in *D. suzukii,* as well as miRNA-124 binding sites upstream of *transformer*, a factor directly involved in *c*VA production in male *D. melanogaster.* A male-specific splicing variant of *fruitless*, *Fru^M*, causes sex-specific neuronal growth, targeting and corresponding behaviour (Demir and Dickson 2005). My gene annotations show that the *fru* region, spanning 100 kb of genomic DNA, contains all putative exons to build the various isoforms found in *D. melanogaster*, including Fru^M fig. 4.6*E*; (Stockinger et al. 2005). I also found that *sexlethal* (Billeter et al. 2006)), *transformer* (Fernández et al. 2010)) and *doublesex* (Rideout et al. 2010)), are well conserved in *D. suzukii* (fig. 4.6). However, we noted that the volume of DA1, a section of glomeruli is sexually isomorphic in *D. suzukii*. This contrasts with *D. melanogaster*, where *Fru* causes a substantial dimorphism in volume and behaviour between males and females (Stockinger et al. 2005). Similarly, the two other glomeruli receiving input from sensory neurons in the *fru* circuitry, VA1v and VL2a, were also sexually isomorphic. However, DL3 was 19% enlarged in male compared with female *D. suzukii*, whereas DL3 is isomorphic in *D. melanogaster* (Stockinger et al. 2005). This may indicate an altered expression of *Fru* in the olfactory circuitry of *D. suzukii* males and females. This notion of an altered expression pattern in *D. suzukii* is substantiated by the observation that *Fru* is translated in brains of both sexes of *D. suzukii*, whereas in *D. melanogaster* only in males(Yamamoto et al. 2004). However, I did not explore the RNA-seq of *D. suzukii* due to low coverage of libraries available.

My colleagues performed a behavioral experiment to see the effect of cVA in *D. suzukii* and they could confirm an adversary effect of cVA in its courting and mating behavior. It is also shown that the species uses cVA as a short range cue to avoid rotten fruits.

Overall, my phylogenetic analyses of genes involved in the metabolism and regulation of CVA indicated that most of them are conserved in *D. suzukii*, although the loss of an Elongase and a Desaturase may represent cases of ecological adaptation. Further work is needed in functionally characterizing putative isoforms of the analyzed gene families. Overall, results indicates that cVA or cVA related compounds may be used as deterrents for the soft fruit industry, in disrupting the mating and/or oviposition behavior of *D. suzukii* in an integrated pest management.

4.3. The role of chemosensation in the making of a pest

4.3.1. Evolution of chemosensory genes in *D. suzukii:* Increased turnover rates of the ORs, OBPs and GRs

I have extracted, annotated and studied the evolution of the main 5 chemosensory genes (olfactory receptors (OR), gustatory receptors (GR), odorant binding proteins (OBP), chemosensory proteins (CSP) and antennal ionotropic receptors (aIR) in *D. suzukii* and *D. biarmipes*. In *D. suzukii*, the chemosensory repertoire consists of 74 protein coding genes from 67 OR loci, 85 protein coding genes from 77 GR loci, 53 OBPs, 4 CSP and 18 aIRs.

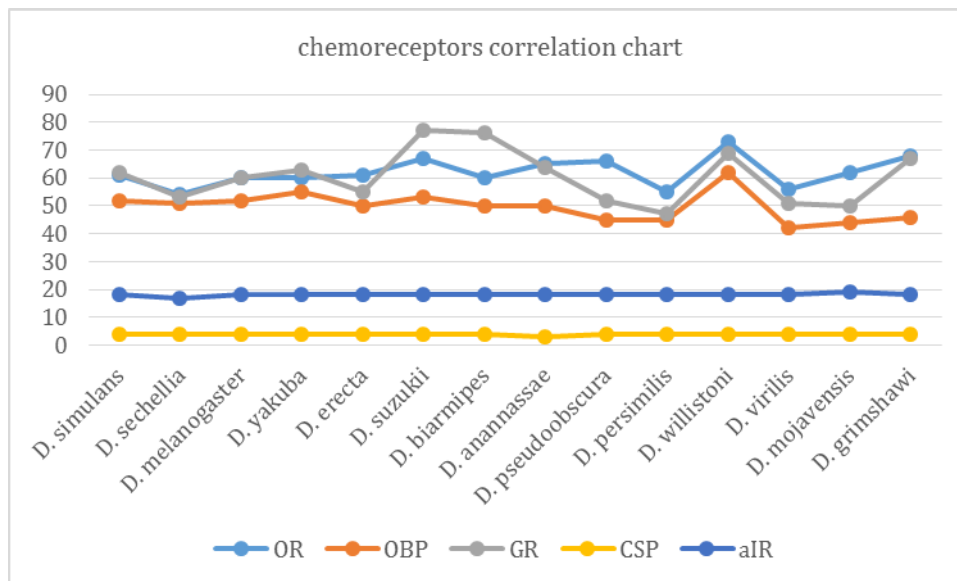


Figure 4.7. Relative count of chemosensory genes in 14 *Drosophila* species. Although a correlation between gene families can be seen mostly in all species, there is a discrepancy in i) high number of GRs in *D. suzukii* - *D. biarmipes* lineages and ii) a reduction of GRs in *D. sechellia* and *D. erecta.*

The OR gene family proved to be extremely dynamic in the branch leading to *D. suzukii,* with eight gene gains (duplications of *Or19a, Or49a, Or59a, Or59c*, *Or67a* and quadruplication of *Or23a*), two genes that likely lost their original function (*Or85a, Or74a;* see below on how we defined a change of function), and two new isoforms (the locus of *Or42a* has three likely transcription start

sites). In the branch leading to *D. suzukii* and *D. biarmipes,* I further identified a loss of function for *Or22a,* a loss of *Or98a,* duplications of *Or65c* and *Or22b*, and a triplication of *Or67a.* In *D. suzukii* all OR duplications arose by tandem replication. The GR family is characterised byan expansion of various genes in the branch leading to *D. suzukii* and *D. biarmipes*: *Gr59c-d* is duplicated 7 times, *Gr92a-93d-93N* thrice, *Gr36a-c* twice, followed by one gain in *Gr22a-f*, *Gr59a-b* and *Gr98b-d* each. *D. suzukii*, in specific, has a gain in *Gr59a-b* and *Gr59c-d* with no loss. Concerning OBPs, I identified three changes in the *D. suzukii* repertoire, namely duplications in *Obp46a* and *Obp47a* and loss of *Obp18a*. Overall, my annotations indicate a fair conservation of the aIR and CSP gene family size among the *Drosophila* species, while the OR, GR and OBP gene families are highly variable.

In general, the annotation process required various rounds of iterative searches, combining automated and manual approaches. One of the interesting, but an expected results is that I could recover more paralogs (particularly frame-disrupted pseudogenes) using the manual annotation, rather than the automated de-novo approach. In the bioinformatic community, compromises in gene annotation approaches between faster predictions using automated pipelines against slower but more accurate predictions using manual approaches is well known (Misra et al. 2002; Yandell et al. 2005). My results show that though the automated approach is most commonly used, it can lead to false negatives, impairing the correct interpretation of functional genomics or applied studies.

The birth-and-death analysis based on the timetree of fourteen *Drosophila* species showed an increase in the turnover rates of OR and OBP genes in *D. suzukii.* Apart from *simulans* and *pseudoobscura* group, *D. suzukii* is the only species to have a turnover rate higher than 1 (overall rate of expansion and normalized beta rate are 1.36 and 0.018 respectively, fig. 4.8*A*). *D. suzukii* clearly falls as an outlier, along with *D. pseudoobscura* and *D. simulans,* in the box plots of figure A). In fact, the total number of events in *D. suzukii* is higher than the other two groups (n=10), but occurred during a longer evolutionary time scale (7.3 mya). These genomic events present a non-random distribution: seven out of the ten events in D. *suzukii* are clustered in a well-supported clade (BS=71) in the gene phylogeny (See grey part in fig. 4.9). This sub family comprises less than a third of the whole OR family (17 out of 61), housing most of the gene gains/losses that characterize *D. suzukii* (refer Ramasamy et al. 2016). Indeed, a Fisher exact test (two tailed, P = 0.25) confirmed a significant departure of the *D. suzukii* lineage from a random distribution of genomic events on the phylogeny. Apart from *D. suzukii*, three other species scored significantly- *D. sechellia, D. virilis* and *D. ananassae.*

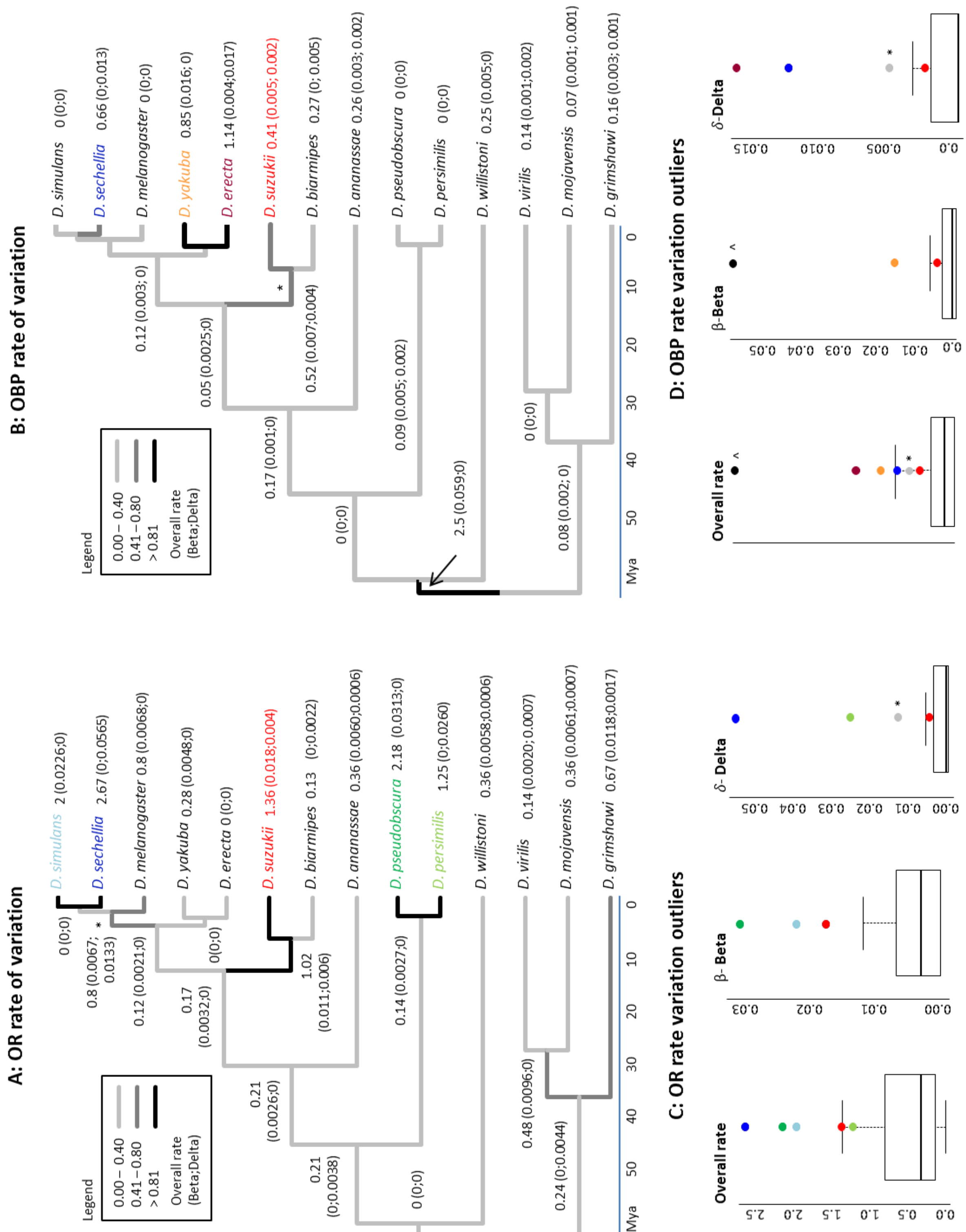
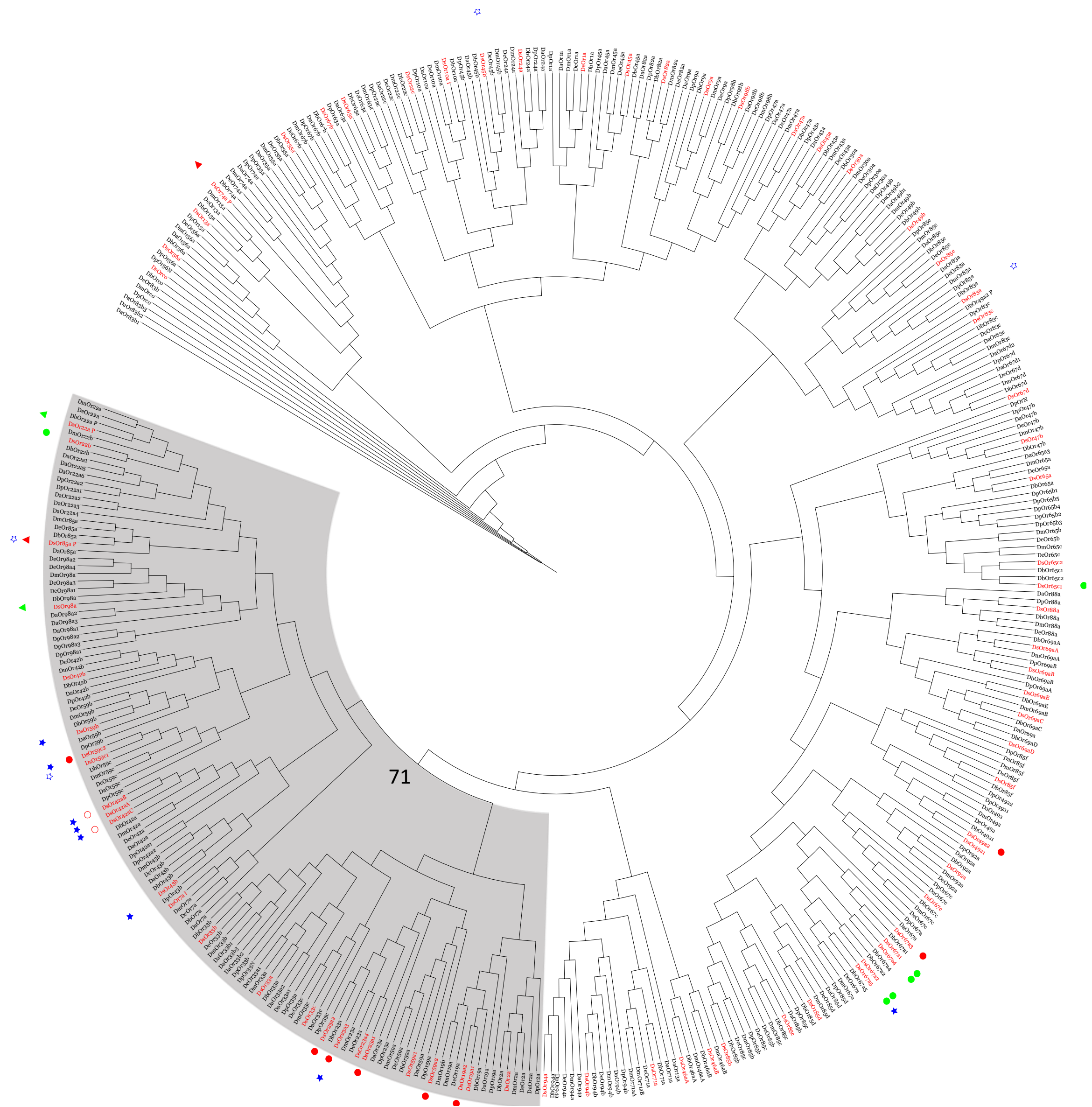


Figure 4.8. Evolution of OR and OBP on the *Drosophila* phylogeny. A: Distribution of the OR gene family size rate variation mapped on a time-tree. Each branch in the tree has overall rate of variation (rate of gain + rate of loss/ (divergence times)) followed by the beta and delta parameters describing respectively birth and death rates from the Badirate analysis using BDI-FR-CML parameters. β and δ values are rounded at the fifth integer. B: Same caption as ORs for OBPs. C: boxplots of overall rate of variation, beta, and delta for OR rate variation. D: Same caption as ORs for OBPs.



Legend:

- Gene gain in *D. suzukii*
- Loss of function in *D. suzukii*
- Gene gain in *D. suzukii* and *D. biarmipes*
- Loss of fun. in *D. suzukii* and *D.* biarmipes
- Positive selection in *D. suzukii*
- Weak pos. sel. in *D. suzukii*

Figure 4.9. Phylogenetic tree of Odorant Receptor. Most of the genomic events detected in *D. suzukii* (duplications, losses, loss of function, positive selection, see legend) cluster significantly in two subfamilies highlighted with grey shade. The tree is inferred using the protein sequences from the entire gene families of 6 species (*D. melanogaster, D. erecta, D. suzukii, D. biarmipes, D. ananassae, and D. pseudoobscura*). Support at each selected nodes is the bootstrap support from the analysis of 100 pseudo-replicates.

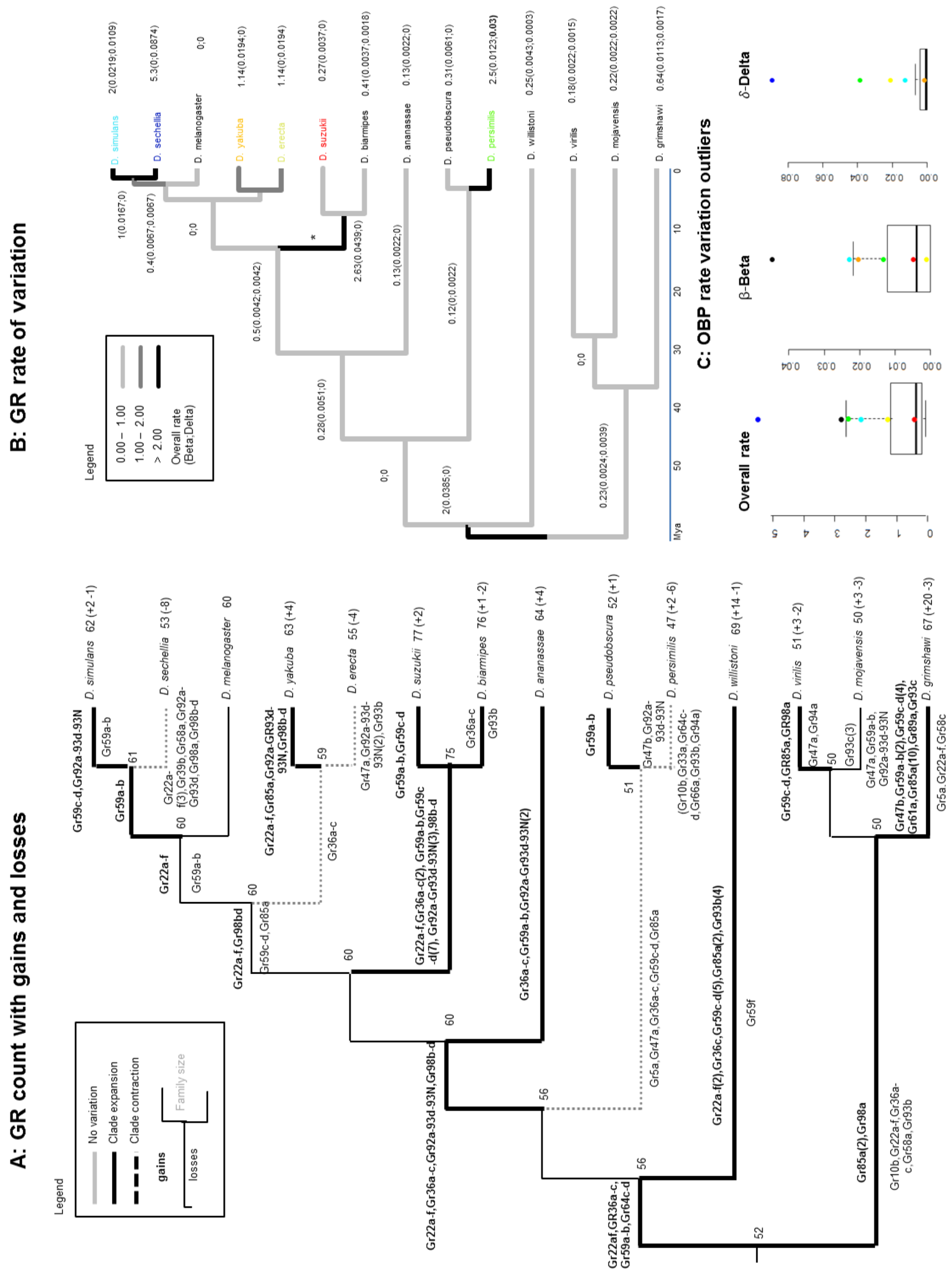


Figure 4.10. Evolution of Gustatory Receptors on the *Drosophila* phylogeny. A: Distribution of gene gains (above branches, in bold) and losses (below branches) on a cladogram depicting phylogeny of 14 *Drosophila* species; values at the right of each terminal or internal nodes are the number of genes calculated by Badirate using BDI-FR-CML model. B: Distribution of the gene family size rate variation mapped on a time-tree. Each branch in the tree has overall rate of variation (rate of gain + rate of loss/ (divergence times)) followed by the beta and delta parameters describing

respectively birth and death rates from the Badirate analysis. β and δ values are rounded at the fifth integer. C: boxplots of overall rate of variation, beta, and delta.

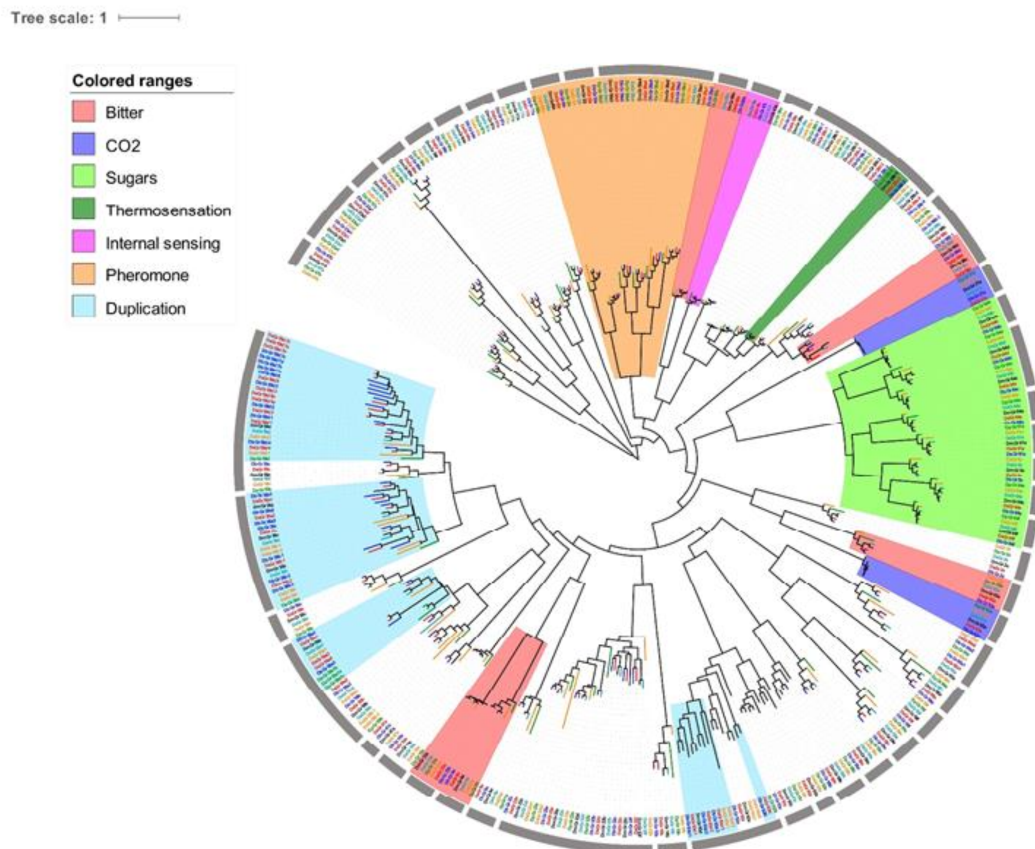


Figure 4.11. Phylogenetic tree of Gustatory Receptor. Clades are coloured based on the functional assays. The expansion of GR clade in *suzukii* subgroup is coloured in blue to indicate the duplication in one part of the tree. The tree is inferred using the protein sequences from the entire gene families of 6 species (*D. melanogaster, D. erecta, D. suzukii, D. biarmipes, D. ananassae, and D. pseudoobscura*). Support at each selected nodes is the bootstrap support from the analysis of 100 pseudo-replicates.

In case of OBPs, the turnover rates are in general lower than ORs for all the species. Again, *D. suzukii*'s overall turnover rate is higher than the average and is one of the highest in the *melanogaster* group. Although *D. suzukii* is not an outlier in the boxplot analysis, the branch leading to *D. suzukii* and *D. biarmipes* is an outlier for what concern the rate of loss (fig. 4.8*D*).

Gustatory receptors similarly show a significant enrichment on the branch grouping *D. suzukii* and *D. biarmipes*, where we could count 15 gains, but no losses. This correspond to a significant gene family enrichment (thick branch in fig. 4.10*A* and boxplot of fig. 4.10*C*), and to one of the highest overall turnover rates in the *Drosophila* phylogeny (2.63), second only to *D. sechellia*. Similar to ORs, most of the duplications/losses that characterise the *suzukii* subgroup are clustered in a well-

supported group of genes (See fig. 4.11): this is again an indication of non-random distribution, although no statistical tests have been yet performed. The chemosensory proteins CSP evolve instead similar to other *Drosophila species*, both in terms of rate of evolution and number of genomic events (data not shown). As for aIRs, both the rate of evolution and gene family size does not vary, except for a gain in *D. mojavensis* and a loss in *D. sechellia* (data not shown).

The process of birth-and-death assumes that chemosensory genes are randomly gained or lost by local genomic events, and that their fate (fixation, or loss/acceptance) is mostly defined by natural selection (Vieira et al. 2007). In the case of *D. suzukii (*or its subgroup)*,* we observe an increase in the birth-and-death rate of the OR and GR (and to a lesser extent OBP) gene families relative to other lineages (figs. 4.8, 4.10). Selection may have played a major role in shaping the duplication pattern in *D. suzukii* receptor genes because duplications and deletions are not randomly distributed along the gene phylogenies (See the grey clade in fig. 4.9). While mutational events (deletions, duplications, consequent positive selection) occur randomly, their fixation is not necessarily random; in the case of *D. suzukii,* selective fixation of certain mutational events may have instead been favoured by natural selection. The observation that such high dynamism occurs within a single clade of ORs and GRs suggests that in *D. suzukii* there has been a shift in the perception of the ligands that characterised such gene clades. The increased rate of GR gain is however shared between *D. suzukii* and *D. biarmipes*; therefore its adaptive role, if any, should be associated with the biology of both species and not only with that of *D. suzukii*. In any case, we can hypothesize that a modification of the chemosensory system, and the associated assortment of receptor genes, accompanied the change in the reproductive lifestyle of *D. suzukii*.

4.3.2. Making sense of the ecological significance of the duplicated/lost genes

To understand the ecological role of duplicated/lost chemosensory genes in *D. suzukii*, we combined our knowledge of their evolution (results outlined above), with behavioural assays, electro-physiological experiments, and information from previous works on *D. melanogaster* (the DoOR v.2 database). We focused mainly on OR and OBP, and evaluate whether certain chemicals or chemical classes are over- or under- represented among those eliciting a response in the duplicated/lost OR genes in *D. suzukii*.

Results show that most of the duplicated/lost/under positive selection ORs in *D. suzukii* respond in *Drosophila melanogaster* to medium sized esters (*Or22a, Or42a, Or59c, Or67a, Or98a,* but also *Obp18a),* to similarly sized fatty alcohols (*Or74a, Or85a*), and to large although chemically unrelated cyclic compounds *(Or19a, Or59a, Or98a);* more details are in fig. 4 of (Ramasamy et al. 2016). These patterns are confirmed by a quantitative screening of the DoOR database which reveals a variety of esters such as ethyl-butyrate, methyl-hexanoate, pentyl-acetate and isopentyl-acetate. Two of the ORs that lost their original function (see next section: 4.3.2) in *D. suzukii* (*Or85a and Or22a)*, bind with high affinity to ethyl 3-hydroxybutyrate and ethyl (and methyl) hexanoate, compounds associated with yeast and bacterial fermentation (Antonelli et al. 1999). Another OBP gene, *Obp57d*, is triplicated in both *D. suzukii* and *D. biarmipes* and is involved in detecting hexanoic and octanoic acids, which are toxic for *Drosophila* in general, but not to *D. sechellia* (Matsuo et al. 2007; Harada et al. 2012b).

It is not straightforward to generalise the biological significance of the many duplications and losses that characterize ORs in *D. suzukii*, as these receptors are elicited by a large assortment of ligands, and because the DoOR database has many biases (see (Ramasamy et al. 2016) for details). Nonetheless, our analyses point toward a role of fatty alcohols, esters, and aromatic compounds which are clearly over-represented as ligands of duplicated/lost genes (compared to all other ORs) in *D. suzukii.* Among esters, the most represented are ethyl butyrate and isopentyl acetate; the latter is present in many ripening soft fruits that host *D. suzukii* (Revadi et al. 2015), and is also released by fermenting materials such as wine and vinegar (Cha et al. 2013). Behavioural assays conducted by colleagues of mine demonstrate that egg-laying females of *D. suzukii* are indeed attracted by lower amount of IPA than *D. melanogaster* are. We speculate an adaptive scenario in which *D. suzukii* has tuned its chemosensory system to better discriminate the odour blend from ripening fresh fruit (for example releasing low amount of IPA), from rotting ones (releasing higher amount of IPA). Our result further point toward three genes *Or19a, Or59a*, *Or67a*, which are duplicated in *D. suzukii*, and that respond to different types of aromatic volatiles in *D. melanogaster*. This may suggest a change in the response to cyclic/aromatic compounds in *D. suzukii*, as also indicated by (Keesey et al. 2015). Any of these duplicated genes are candidates for having replaced other ORs in *D. suzukii* (see 4.3.2).

As for GRs, most of the duplicated ones in *suzukii* subgroup are known to be expressed in bitter-sensitive sensilla in *D. melanogaster* (Weiss et al. 2011). Among them, *Gr59c* is broadly expressed and is hypothesized to work in combination with other GRs in bitter sensing neurons. Other

duplicated GRs- *Gr36a-c, Gr59a, Gr59c-d, Gr22b-f, Gr98b* and *Gr92a*, are expressed in the labellum and/or forelegs, and may be linked with bitter sensing (Weiss et al. 2011; Ling et al. 2014). The fact these duplicated GRs are housed in specific class of sensilla, and cluster in one phylogenetic clade indicates their possible function in binding similar ligands. It is possible that such a rapid expansion of GR underlies differential tolerance for bitter compounds.

4.3.3. Altered responses of sensory physiology driven by loss of function of key receptors

In *D. suzukii*, the amino acid sequences of two odorant receptors (*Or22a* and *Or85a*) present deletions or stop codons that compromise their reading frame, but otherwise retained high sequence similarity with their *D. melanogaster* orthologues (fig. 4.12). Because *Or22a* and *Or85a* are transcribed, there is the intriguing possibility that these changes did not cause a pseudogenization of the gene, but rather are associated to a change of function. The aforementioned "deleterious" changes are indeed, found in portions of the exons that are missing in the transcripts (available only for the American strain, Bioproject Accession: PRJNA221549), suggesting new exon structures and novel splicing patterns that resulted in at least one transmembrane region being lost in each of the genes when compared to the *D. melanogaster* proteins. A third receptor, *Or74a*, is more likely a pseudogene because it retains poor similarity with the orthologs in other species. We further validated these differences using PCR analysis: results confirmed that the genomic region covering the first transmembrane helix is completely absent from the American strain (dotted line in fig. 4.12*A*, and indicated two different alleles in Italian population, while only one in American (fig. 4.12*C*). While introns (including the newly formed ones) are fairly divergent between the American and the Italian genome, exons are highly conserved and did not accumulate deleterious mutations. Given that these two ORs- *Or85a* and *Or22a* present an interesting case of possible change of function, my colleagues performed single sensillum recordings from the large basiconic sensilla that house neurons expressing *Or85a* (ab2B) and *Or22a* (ab3A) in *D. melanogaster.* Results demonstrated that the *D. suzukii* cognate neurons have a strongly shifted response profile compared to *D. melanogaster* (see fig. 6 in Ramasamy et al. 2016).

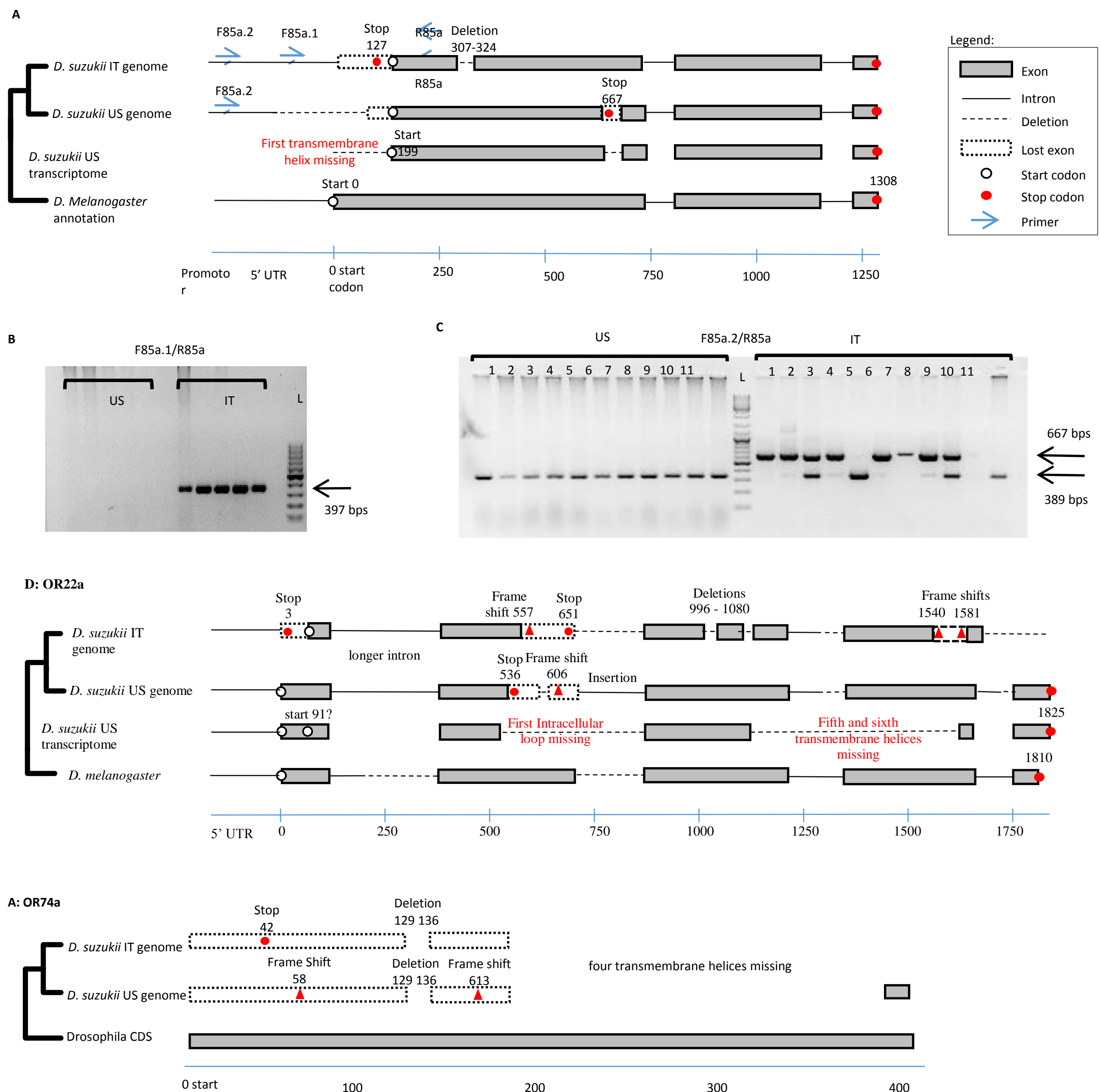


Figure 4.12. Different non-functional Odorant Receptors in American and European population. The structure of the predicted coding sequences (CDS) of *Or85a* (panel A), *Or22a* (panel D) and *Or74a (*panel E) from the genome analysis of the Italian (IT) and American (US) strains of *D. suzukii*. For the American strain, we also provide the CDS from transcriptome (Chiu et al. 2013). Dotted lines in *D. suzukii* indicate that the CDS is missing either from the genomes or the transcriptome. B and C: agarose gel (2%) electrophoresis of different splice variants present in different individuals of American and Italian *D. suzukii* populations: US – American strain, IT – Italian strain, L – Ladder.

In *D. melanogaster,* the ab2B neuron is tuned to oxidized esters typical of rotten fruit like ethyl 3-hydroxybutyrate: our recordings from ab2B demonstrate that *D. suzukii* does not respond to this odour (fig. 6*A*; Ramasamy et al. 2016), but rather had acquired an increased affinity for 2heptanone,

supporting a loss of function of its cognate receptor *Or85a* (see also Keesey et al. 2015, and a likely replacement by another OR. Similarly, while the ab3A neuron responds strongly to ethyl and methyl hexanoate in *D. melanogaster* (see also Andersson et al. 2012), *D. suzukii* has lost its high sensitivity to these compounds and acquired an increased sensitivity for ethyl acetate (fig. 6*B;* Ramasamy et al. 2016). Indeed, these two compounds are typical of ripening fruits (Keesey et al. 2015). Response of neuron expressing *Or74a* was not tested since in *D. melanogaster* this receptor is expressed during larval stage (Kreher et al. 2005).

4.4. Evolution of chemosensory receptors in Arthropoda

4.4.1. Odorant Receptors and new findings in Palaeoptera

I tested the hypothesis that ORs originated concomitantly with the emergence of vascular plants (in the branch leading to Zygentoma plus pterygotes (Missbach et al. 2014)) by searching OR/*Orco* in the recently sequenced basal-insect genomes, Ephermeroptera (Mayfly), Odonata (dragonfly) and Archeognatha (Bristletail).

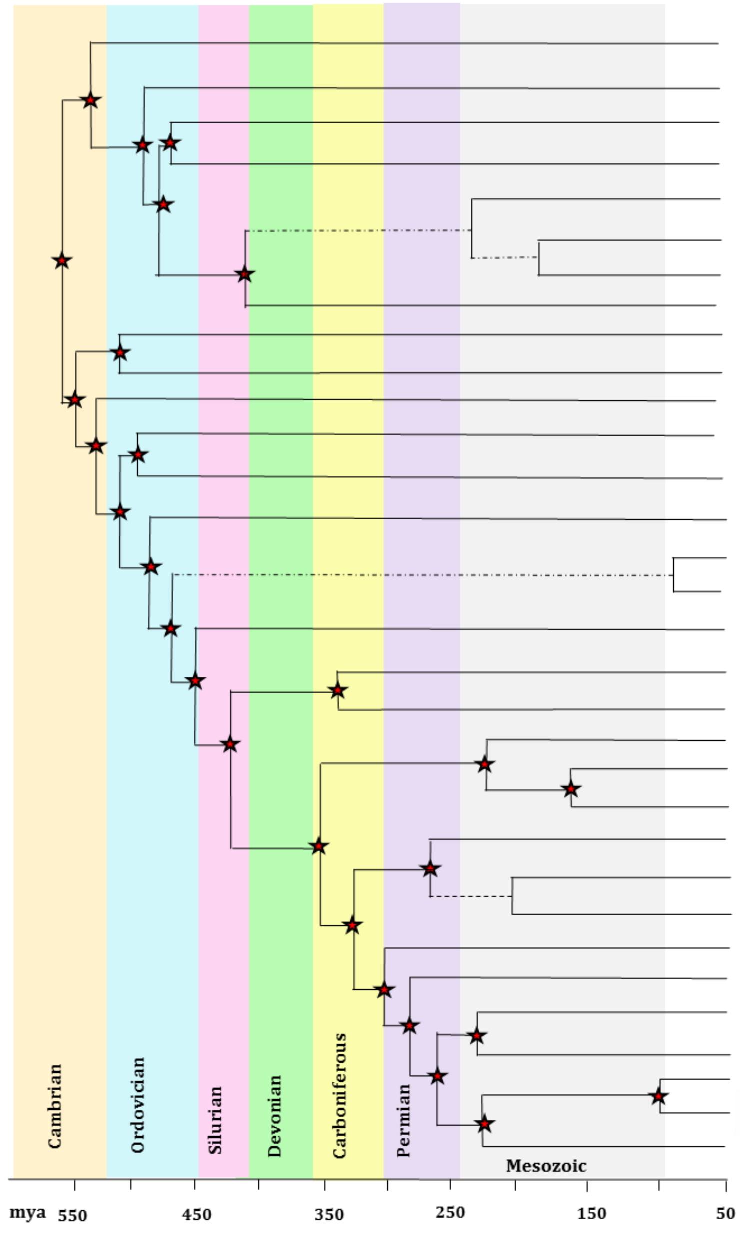
My annotations (fig. 4.13) indicate that Archeognatha (*Machilis hrabei*) completely lack any OR/*Orco*. I could find instead fully formed *Orco* in the Palaeoptera (mayfly and dragonfly), the first extant flying lineages of insects. Palaeoptera have been long considered anosmic because they lack glomerular antennal lobes and mushroom body calyces, involved with olfaction in other insects (Farris 2005); however, recent studies indicated an aerial borne sense of smell for dragon flies (Rebora et al. 2012; Piersanti et al. 2014). Given the low number of olfactory sensilla and smaller sized antennae, previous attempts to identify OR/*Orco* from Palaeoptera RNA-seq proved unsuccessful (Missbach et al. 2014). Unlike Zygentoma that has 6 *Orco*, I could only find a single, although full-length conserved copy of *Orco* in the genomes of *Ladona fulva* (dragonfly) and *Ephemeroptera danica* (mayfly) and no traces of specific ORs. Although I performed extensive searches, this does not completely exclude presence of (false negative) ORs: while Orco is well conserved, ORs are often lineage specific and may differ from the ones in other insect orders,

therefore it is difficult to find ORs in unexplored lineages using homology. I have only performed annotations of ORs; further research will focus on reconstructing their phylogeny.

Overall, my results confirm that *Orco* originated in the common ancestor of Zygentoma and Pterygota (a clade known as Dicondylia). It also suggests that a fully formed OR mechanism evolved quite later, only in the neopterans, as Palaeoptera are characterized by the sole presence of *Orco*. The presence of the sole *Orco* in Zygentoma and Palaeoptera suggests that these groups are characterized by a primordial, not yet developed OR based olfactory system. This is not surprising from an ecological perspective. Palaeoptera for example spend their life as aquatic juveniles: in water, taste plays a pivotal role, and olfaction may not be required. Furthermore, adult mayflies do not feed, while Odonata seems to have a biology strongly driven by sight (Futahashi et al. 2015); smell may be less important in these two orders of insects.

4.4.2. Lineage based, rapid evolution of GRs

I further explored the distribution of GRs and IRs, two chemosensory receptors whose functional role in insects is more devious than ORs (Croset et al. 2010; Montell 2013). GRs are mostly involved in taste, but some of them are tuned to small gases such as CO_2 (Kwon et al. 2007). IRs are broadly expressed in both olfactory (aIR) and gustatory receptor neurons (dIR) in *Drosophila*. The ability of these two types of receptor to sense both water-borne and air-borne chemicals, while OR virtually being only air borne detectors, suggests that they are both more ancient than ORs, and that their evolutionary patterns may be complex. The distribution of these receptors in the arthropod is still however poorly understood. Taking advantage of the increasing availability of arthropod genomes (Consortium 2013, https://www.hgsc.bcm.edu/i5k-pilot-project-summary), I searched these genomes, as well as other available transcriptomes, for both GRs and IRs.



Sub-Phylum	Class	Species	GR	IR
Chelicerara	Pycnogonida	*unknown*	26	52
	Xiphosura	*Limulus polyphemus*	37	39
	Acari	*Tetranychus urticae*	3	21
		Ixodes scapularis	24	41
	Aranae	*Loxosceles reclusa*	3	52
		Latrodectus hesperus	5	16
		Stegodyphus mimosarum	0	63
	Scorpiones	*Centruroides sculpturatus*	36	74
Myriapoda	Chilopoda	*Strigamia maritima*	60	82
	Diplopoda	*Trigoniulus corallinus*	2	17
Crustacea	Malacostraca	*Hyalella azteca*	0	24
	Copepoda	*Eurytemora affinis*	0	52
	Branchiopoda	*Daphnia pulex*	55	93
Hexapoda	Diplura	*Catajapyx aquilonaris*	12	39
	Archeognatha	*Machilis hrabei*	20	NA
		*Lepismachilis y-signata**	7	17
	Zygentoma	*Thermobia domestica**	9	19
	Ephemeroptera	*Ephemera danika*	7	20
	Odonata	*Ladona fulva*	10	16
	Phasmatodea	*Phyllium siccifolium**	6	32
	Blattodea	*Blatella germanica*	7	50
	Isoptera	*Zoomeptera nevadensis*	80	143
	Hemiptera	*A. pisum/A. gossypii*	77	14
	Thysanoptera	*Frankliniella occidentalis*	7	23
	Phthiraptera	*Pediculus humanus*	6	12
	Hymenoptera	*Apis mellifera*	10	10
	Coleoptera	*Tribolium casteneum*	220	23
	Trichoptera	*Limnephilus lunatus*	36	26
	Lepidoptera	*Bombyx mori*	66	18
	Diptera	*Aedes aegypti*	79	95
		Anopheles gamibae	76	46
		Drosophila melanogaster	60	66

Figure 4.13. List of species included in the study and the corresponding number of GRs and IRs identified. Species in blue indicate annotations taken from previous studies and the ones with asterisk are predicted only from respective transcriptomes (See Missbach et al. 2014). The Molecular clock is adapted for higher order species from (Rota-Stabelli et al. 2013b) and for holometabola and chelicerate from (Misof et al. 2014; Sharma and Giribet 2014; Garrison et al. 2016). The dotted lines indicate absence of accurate divergence times. NA indicates the annotation is not yet done.

My annotations indicate a complex pattern of GR evolution in arthropods. Acari and Diplopoda for example are characterized by few GRs, while Merostomata (Horseshoe crab), Pycnogonida (Sea spider), Scorpiones and Chilopoda possess many orthologs. Some of these difference may be, however, due to false negatives, for example in the case of the diplopod *Trigoniulus*, the original annotation was not focused on GRs. The ballooning spider, *Stegodyphus,* is the only arthropod (apart from some Crustacea) to completely lack GRs. In case of *Tetranychus urticae,* even performing various rounds of gene prediction analysis, I could only recover 4 putative chemosensory genes. This is in contrast with the study (Grbic et al. 2011), which found ~135 receptors as putative chemosensory genes (ORs and GRs): my annotation indicates that these may all be false positives. Conversely to GRs, IRs are more homogenously present throughout the sampled lineages, and are in general more abundant than GRs. Apart from Branchiopoda in Crustacea, GRs are completely missing in Malacostraca and Copepoda. I also analysed the RNA-seq of Oniscidea, a terrestrial isopod but could not recover any GRs or IRs; the absence of IRs suggest a poor/incomplete transcriptome, therefore it has not been included in figure (4.13). Within insects, Hemiptera and Isoptera (Termites) have the highest number of GRs, next to Holometabola. Isoptera have the most number of IRs (~150), while the Hymenoptera have the lowest of all, with only 10 (Croset et al. 2010). While I have annotated both GRs and IRs in the genomes, as detailed in the figure (4.14), I have analyzed only the phylogeny of GRs. The phylogenetic analysis of IRs and ORs will be carried out as a part of future work.

The GR dataset was mainly composed of fragmented proteins, in most cases covering only 6th and 7th transmembrane helices. This is because this gene family is characterized by lineage specific expansions and hence gene prediction algorithms and HMM-based profile searches likely failed to extract full length sequences. I validated these findings by performing blast searches against the Uniprot database and found no false positives. The GR alignment showed an interesting pattern: while the last two transmembrane helices are conserved throughout all arthropods, some internal transmembrane (likely helical) domains in Myriapoda show a different alignment structure, from the rest of Hexapoda (Appendix A). The phylogenetic analysis of GRs was performed on two different datasets: a full length alignment encompassing all sites (fig. 4.14), and a dataset including

only the more conserved C-terminal (fig. 4.15). The main difference is the position of the myriapods GR. However, both trees are fairly unresolved at many nodes as indicated by low bootstrap supports (BS). However, it is possible to observe clear local expansions in each of the main arthropod lineages: Hexapoda (green and pink), Crustacea (light blue), Myriapoda (yellowish) and Chelicerata (red). In particular, all myriapod GR seem to have a unique origin and a puzzling direct orthology with a crustacean GR. In both trees is possible to observe a small group of GR close to the root that includes genes from hexapods, branchiopod and chelicerates: this may represent the ancestral set of GR in the arthropods. It is also possible to observe a clear large chelicerate expansion and an equally large pancrustaecan (Hexapoda plus Crustacea) expansion. Intriguingly, both Maximum Likelihood trees depict a clade of insects plus branchiopod (to the exclusion of few other insects and dipluran GRs); this challenge the monophyly of both insects and hexapods (Rota Stabelli et al. 2013), but see for example (Carapelli et al. 2007). Although interesting, any possible speculations over the groupings in figure 4.14 and 4.15 should be taken with care because BS are extremely low and in most cases insignificant, suggestive of a complex, likely contrasting, surely weak, phylogenetic signal in these sequences. The phylogenetic analysis of GR I'm presenting is therefore extremely preliminary and more work is needed, possibly using Bayesian inference, to disentangle the weak phylogenetic signal in the dataset.

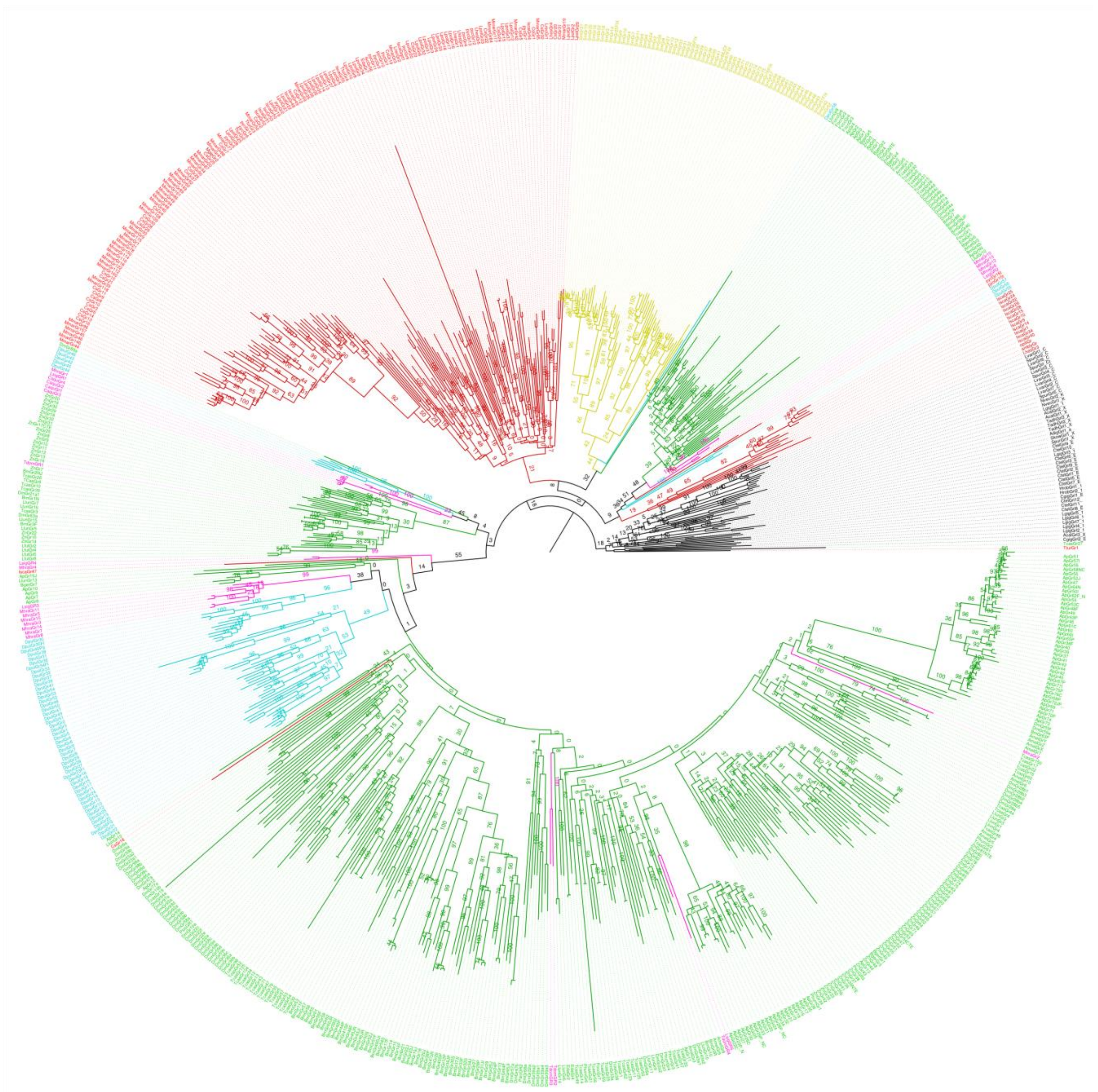


Figure 4.14. Phylogenetic tree based on whole alignment dataset of Arthropoda GR, rooted with GR-like of Cnidaria and Lophotrocozoa (black). Protein alignment made from 523 positions. Color code: flying insects Pterygota (green), other hexapods (Apterygota and Entognatha (pink), Crustacea (sky blue), Myriapoda (Brown/yellow) and Chelicerata (red). The tree is the Maximum Likelihood tree inferred by RAxML using the GTR-G model with bootstrap supports from the analysis of 100 pseudo-replicates.. Highly fragmented and partial sequences have been removed from the analysis.

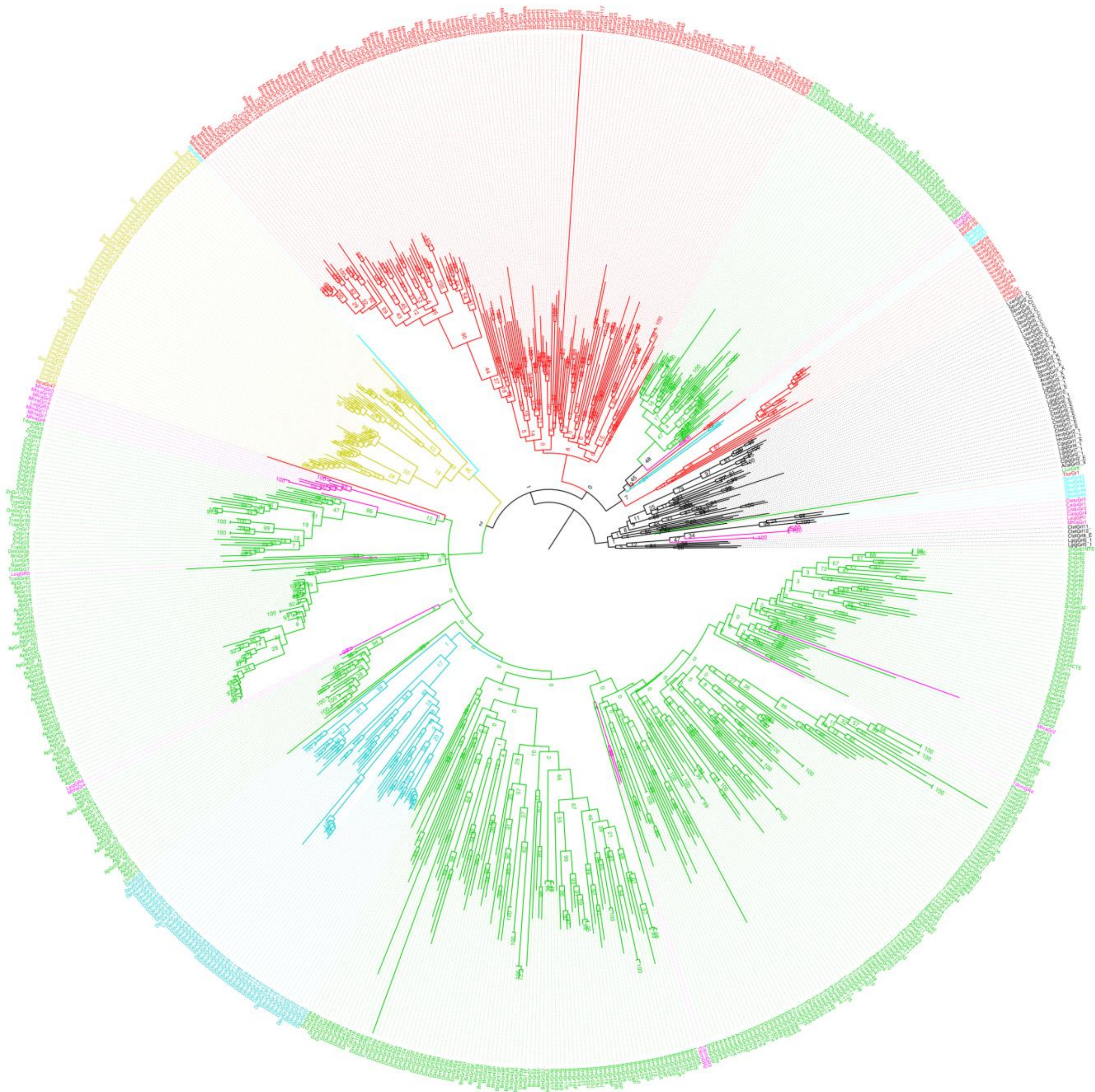


Figure 4.13. Phylogenetic tree based on conserved C-terminal alignment dataset of Arthropoda GR, rooted with GR-like of Cnidaria and Lophotrocozoa (black). Protein alignment made from 201 positions. Color code: flying insects Pterygota (green), other hexapods (Apterygota and Entognatha (pink), Crustacea (sky blue), Myriapoda (Brown/yellow) and Chelicerata (red). The tree is the Maximum Likelihood tree inferred by RAxML using the GTR-G model with bootstrap supports from the analysis of 100 pseudo-replicates. Highly fragmented and partial sequences have been removed from the analysis.

5. Conclusion and future perspectives

D. suzukii genomics

The invasive nature of *D. suzukii* and the consequent damage to soft fruit industry required immediate attention; to tackle this issue in a sustainable and efficient way, my host lab has sequenced its genome and transcriptome, and I have contributed to its analysis (Ometto et al. 2013). The assembly size revealed that *D. suzukii's* genome is comparable to those from most other *Drosophila*, both in term of size and repeat elements. The genome of *D. suzukii* has been extremely useful throughout my thesis as a quick and reliable source for obtaining full length genes; in the absence of a genome, I would have relied on transcriptome with the risk of finding only partial genes and high amount of false negatives, which would have impaired the correct interpretation of comparative studies.

My analyses of mitochondrial genomes indicate a sister relationship of *D. suzukii* with *D. biarmipes.* Although the transcriptomic dataset confirmed such relationship, the two datasets did not converge in the placement of *D. eugracilis*, because of signals concentrated in the third codon positions of mitochondrial dataset. This incongruence has thought me the importance of using different datasets (mitogenomic, transcriptomic), substitution models (homogeneous, heterogeneous models) as well as different phylogenetic frameworks (ML, Bayesian) to identify systematic errors, if any, in phylogenetic studies.

Importantly, *D. suzukii's* likely closer related species, *D. subpulchrella*, has not been included in my analyses: if *D. subpulchrella* is the actual sister species to *D. suzukii,* then some of the evolutionary events I have ascribed to *D. suzukii* may instead be shared by both species. On the other hand, *D. subpulchrella* may in fact have a reproductive biology more similar to the typical *Drosophila* one (Atallah et al. 2014), so that the functional changes that I have found along the *D. suzukii* lineage may be really reflecting the new reproductive habit in this species. The genome of the *D. subpulchrella* is under sequencing in my hosting lab, and annotations of its gene repertoire with that of *D. suzukii* will clarify this issue, and hopefully confirm some of the findings of my research.

Chemosensory genes of D. suzukii, their ecological significance and utility in pest control

'The act of smelling something, anything, is remarkably like the act of thinking itself'

-Lewis Thomas

The insect chemosensory system is one of the most dynamic among the animals, given their level of diversity and successful invasions of extreme ecologies. Hence, they are forced to reinvent mechanisms so as to perceive appropriate volatiles, and are subjected to strong selective pressures. How the chemosensory system responds to selective pressures is essential in answering key evolutionary questions. For example, the correlation between host specialization and olfaction has been documented in few *Drosophila* species. The results of such studies immensely encouraged me in taking up the case of *D. suzukii*, a pest that has switched its ovipositing substrate from rotten fruits to fresh fruits. In my study, I have identified and annotated the entire chemosensory system of *D. suzukii*, to understand the role of its chemosensory genes and their evolution in influencing the pest's behaviour. My work (Ramasamy et al. 2016) is one of the very few to combine evolutionary genomics with behavioural analysis. Results show that *D. suzukii* has an accelerated chemosensory evolution among other species of the same genus, particularly the olfactory receptors. Notably, *D. suzukii* is the only species to show both an increase in the OR turn-over rate and a non-random distribution of events in its phylogeny.

I have identified few key genes and my colleagues have found their associated ligands which may have played a role in the peculiar phenotype of *D. suzukii*. I found a burst of duplication in genes that have affinity to a variety of ligands, particularly esters such as IPA (Isopentyl acetate), which my colleagues and I suggested to play a peculiar role in *D. suzukii* (Revadi et al. 2015). On the contrary, the pest have lost few relevant genes such as *Or22a* and *Or85a*, which bind volatiles released from fermenting fruits. Interestingly these two tend to have different isoforms (gene structures) in European and American populations. The physiological recordings by my colleagues confirm this, where the neurons ab3A and ab2B respectively expressing *Or22a* and *Or85a* receptors have shifted their sensitivity towards other compounds (fig. 6; (Ramasamy et al. 2016)).

In case of GRs, the clade leading to *D. suzukii* and *D. biarmipes* shows a high rate of variation (Crava, Ramasamy et al. in prep), with many duplications that can be attributed to one specific clade of receptors known to be expressed in bitter sensing neurons in *D. melanogaster*. Further

studies should concentrate in performing single cell recordings in the neurons expressing these receptors as my colleagues already did for ab2B and ab3A neurons. Whatever might be the case, such duplications are shared also by *D. biarmipes*, and hence their adaptive role, if any, should be associated with the biology of both the species.

I have also focused on class of enzymes and transcription factors involved in the sexual communication and mating behaviour of *Drosophila* mediated by the pheromone cVA (Dekker et al. 2015). *D. suzukii* shows loss of a gene both in desaturase and elongase gene families, both shared by *D. biarmipes.* Odorant receptors, *Or67d* and *Or65a* directly involved in cVA perception and consequently mating behaviour are conserved in *D. suzukii* as well the transcription factors such as *fruitless, transformer, doublesex* and *sexlethal*. Further work is needed in functionally characterizing the putative isoforms of Elongases as they may be responsible for the loss of cVA In *D. suzukii.*

Overall, my analysis indicates that chemosensory genes contributed in shaping, at least partly, the adaptive behaviour in *D. suzukii*: we have found duplications in genes with affinity for fresh fruit volatiles and loss of function in genes that in *D. melanogaster* are found to respond ligands emitted from decaying substrates. The above mentioned important genes along with the ligands they bind, as given in fig.4 of (Ramasamy et al. 2016), are promising targets for further functional analysis, for example expression in a heterologous empty neuron system. It is my hope that the results presented here will help direct research efforts in the development of more targeted odour-based trapping and control methods. Future works should test those ligands for which there has been a shift in chemosensation, particularly 1-hexanol, 2-heptanone, and beta-cyclocitral, the two latter being putatively new ligands of respectively ab2B and ab3A neurons in *D. suzukii.*

Evolution of chemosensory genes on a birth-death process

Chemosensory genes are some of the many gene families known to evolve through a process of gene gain and loss, conveniently described by the birth and death model of evolution (Nei and Rooney 2005). Some of my analyses provided results that did not match previous ones (conducted on similar datasets but with different taxon sampling). For example, the distribution of gene gains/losses along my OR phylogeny slightly differs from inferences made on a more restricted sample of *Drosophila* (Guo and Kim 2007b; McBride and Arguello 2007): some of the gains that were previously located on the branch subtending the *melanogaster* subgroup, in our analysis are located on the branch subtending the whole *melanogaster* group. Furthermore, my estimate of OBP

overall birth rate in *Drosophila* (β = 0.0028) differed from that of the whole arthropods (β = 0.0049, (Vieira and Rozas 2011)), indicating that *Drosophila* OBP turnover rate is lower than in most other arthropods.

On the origin and radiation of chemoreceptors in Arthropods

For the very first time, I could identify and annotate an entire *Orco* gene in the Palaeoptera, which have been long considered anosmic. In accordance with (Missbach et al. 2014) I could not find any OR/*Orco* in Diplura and Archeognatha, confirming the origin of insect-specific ORs in the branch leading to pterygotes and Zygentoma. My screening of GRs in basal Hexapoda, Crustacea, Myriapoda and Chelicerata revealed a strong lineage specific evolutionary dynamics. I have found that except for Branchiopoda, the sampled Crustaceans completely lack GRs. This might be an additional clue in solving the Pancrustacean phylogeny, supporting the sister relationship of Hexapoda and Branchiopoda. However, future works on the genome of Remipedia and Oligostraca are needed to validate my findings.

Final remarks

In conclusion, my work on the chemosensory receptors has helped in exploring some fundamental questions pertained to the origin and evolutionary dynamics of these multi-gene families. Science has come a long way in analysing the evolutionary origin of smell and taste in insects, owing to the intriguing and complex nature of the proteins underlying these senses. With my work, I could confirm that ORs are specific to insects and GRs are functionally lineage specific. As for *Drosophila*, my work further pushes forward our understanding of the evolutionary dynamics of these genes in this important biological model. In the case of *D. suzukii*, the list of candidate genes and ligands that I have identified are good candidates for downstream applied physiological and behavioural experiments.

References

Abascal F, Zardoya R, Telford MJ. 2010. TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. Nucleic Acids Research 38:W7-W13.

Ader L, Jones DNM, Lin H. 2010. Alcohol Binding to the Odorant Binding Protein LUSH: Multiple Factors Affecting Binding Affinities. Biochemistry 49:6136-6142.

Ai M, Min S, Grosjean Y, Leblanc C, Bell R, Benton R, Suh GSB. 2010. Acid sensing by the Drosophila olfactory system. Nature 468:691-695.

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Research 25:3389-3402.

Anderson PSL, Friedman M, Ruta M. 2013. Late to the Table: Diversification of Tetrapod Mandibular Biomechanics Lagged Behind the Evolution of Terrestriality. Integrative and Comparative Biology.

Andersson MN, Schlyter F, Hill SR, Dekker T. 2012. What Reaches the Antenna? How to Calibrate Odor Flux and Ligand–Receptor Affinities. Chemical Senses 37:403-420.

Angeli S, Ceron F, Scaloni A, Monti M, Monteforti G, Minnocci A, Petacchi R, Pelosi P. 1999. Purification, structural characterization, cloning and immunocytochemical localization of chemoreception proteins from Schistocerca gregaria. European Journal of Biochemistry 262:745-754.

Antonelli A, Castellari L, Zambonelli C, Carnacini A. 1999. Yeast Influence on Volatile Composition of Wines. Journal of Agricultural and food Chemistry 47:1139-1144.

Arn H, Louis F. 1997. Mating Disruption in European Vineyards. In: RT Cardé, AK Minks, editors. Insect Pheromone Research: New Directions. Boston, MA: Springer US. p. 377-382.

Asplen MK, Anfora G, Biondi A, Choi D-S, 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. Journal of Pest Science 88:469-494.

Atallah J, Teixeira L, Salazar R, Zaragoza G, Kopp A. 2014. The making of a pest: the evolution of a fruit-penetrating ovipositor in Drosophila suzukii and related species. Proceedings of the Royal Society of London B: Biological Sciences 281.

Benton R, Sachse S, Michnick SW, Vosshall LB. 2006. Atypical Membrane Topology and Heteromeric Function of *Drosophila* Odorant Receptors In Vivo. PLoS Biol 4:e20.

Benton R, Vannice KS, Gomez-Diaz C, Vosshall LB. 2009. Variant Ionotropic Glutamate Receptors as Chemosensory Receptors in Drosophila. Cell 136:149-162.

Biémont C, Cizeron G. 1999. Distribution of transposable elements in Drosophila species. Genetica 105:43-62.

Billeter J-C, Rideout EJ, Dornan AJ, Goodwin SF. 2006. Control of Male Sexual Behavior in Drosophila by the Sex Determination Pathway. Current Biology 16:R766-R776.

Bray S, Amrein H. 2003. A Putative Drosophila Pheromone Receptor Expressed in Male-Specific Taste Neurons Is Required for Efficient Courtship. Neuron 39:1019-1029.

Buck L, Axel R. 1991. A novel multigene family may encode odorant receptors: A molecular basis for odor recognition. Cell 65:175-187.

Calvello M, Brandazza A, Navarrini A, Dani FR, Turillazzi S, Felicioli A, Pelosi P. 2005. Expression of odorant-binding proteins and chemosensory proteins in some Hymenoptera. Insect Biochemistry and Molecular Biology 35:297-307.

Calvello M, Guerra N, Brandazza A, D'Ambrosio C, Scaloni A, Dani FR, Turillazzi S, Pelosi P. 2003. Soluble proteins of chemical communication in the social wasp Polistes dominulus. Cellular and Molecular Life Sciences CMLS 60:1933-1943.

Carapelli A, Liò P, Nardi F, van der Wath E, Frati F. 2007. Phylogenetic analysis of mitochondrial protein coding genes confirms the reciprocal paraphyly of Hexapoda and Crustacea. BMC Evolutionary Biology 7:1-13.

Cattel J, Kaur R, Gibert P, Martinez J, Fraimout A, Jiggins F, Andrieux T, Siozios S, Anfora G, Miller W, et al. 2016. <italic>Wolbachia</italic> in European Populations of the Invasive Pest <italic>Drosophila suzukii</italic>: Regional Variation in Infection Frequencies. PLoS ONE 11:e0147766.

Cha DH, Hesler SP, Cowles RS, Vogt H, Loeb GM, Landolt PJ. 2013. Comparison of a Synthetic Chemical Lure and Standard Fermented Baits for Trapping *Drosophila suzukii* (Diptera: Drosophilidae). Environmental Entomology 42:1052-1060.

Chertemps T, Duportets L, Labeur C, Ueda R, Takahashi K, Saigo K, Wicker-Thomas C. 2007. A female-biased expressed elongase involved in long-chain hydrocarbon biosynthesis and courtship behavior in Drosophila melanogaster. Proceedings of the National Academy of Sciences 104:4273-4278.

Chertemps T, Duportets L, Labeur C, Ueyama M, Wicker-Thomas C. 2006. A female-specific desaturase gene responsible for diene hydrocarbon biosynthesis and courtship behaviour in Drosophila melanogaster. Insect Molecular Biology 15:465-473.

Chipman ADFerrier DEKBrena CQu JHughes DSTSchröder RTorres-Oliva MZnassi NJiang HAlmeida FC, et al. 2014. The First Myriapod Genome Sequence Reveals Conservative Arthropod Gene Content and Genome Organisation in the Centipede <italic>Strigamia maritima</italic>. PLoS Biol 12:e1002005.

Cini A, Ioriatti C, Anfora G. 2012. A review of the invasion of Drosophila suzukii in Europe and a draft research agenda for integrated pest management. Bulltein of Insectology. p. 149-160.

Clyne PJ, Warr CG, Freeman MR, Lessing D, Kim J, Carlson JR. 1999. A Novel Family of Divergent Seven-Transmembrane Proteins: Candidate Odorant Receptors in Drosophila. Neuron 22:327-338.

Consortium DG. 2007. Evolution of genes and genomes on the Drosophila phylogeny. Nature 450:203-218.

Consortium iK. 2013. The i5K Initiative: Advancing Arthropod Genomics for Knowledge, Human Health, Agriculture, and the Environment. Journal of Heredity 104:595-600.

Croset V, Rytz R, Cummins SF, Budd A, Brawand D, Kaessmann H, Gibson TJ, 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, Lei Y-T, Kwon JY, Carlson JR. Two GR Genes Underlie Sugar Reception in Drosophila. Neuron 56:503-516.

de Bruyne M, Foster K, Carlson JR. 2001. Odor Coding in the Drosophila Antenna. Neuron 30:537-552.

de Bruyne M, Smart R, Zammit E, Warr C. 2010. Functional and molecular evolution of olfactory neurons and receptors for aliphatic esters across the Drosophila genus. Journal of Comparative Physiology A 196:97-109.

Dekker T, Ibba I, Siju KP, Stensmyr MC, Hansson BS. 2006. Olfactory Shifts Parallel Superspecialism for Toxic Fruit in Drosophila melanogaster Sibling, D. sechellia. Current Biology 16:101-109.

Demir E, Dickson BJ. 2005. fruitless Splicing Specifies Male Courtship Behavior in Drosophila. Cell 121:785-794.

Drysdale RA, Crosby MA, Consortium TF. 2005. FlyBase: genes and gene models. Nucleic Acids Research 33:D390-D395.

Dunipace L, Meister S, McNealy C, Amrein H. 2001. Spatially restricted expression of candidate taste receptors in the Drosophila gustatory system. Current Biology 11:822-835.

Dweck Hany KM, Ebrahim Shimaa AM, Kromann S, Bown D, Hillbur Y, Sachse S, Hansson Bill S, Stensmyr Marcus C. Olfactory Preference for Egg Laying on Citrus Substrates in Drosophila. Current Biology 23:2472-2480.
Dweck HKM, Ebrahim SAM, Thoma M, Mohamed AAM, Keesey IW, Trona F, Lavista-Llanos S, Svatoš A, Sachse S, Knaden M, et al. 2015. Pheromones mediating copulation and attraction in Drosophila. Proceedings of the National Academy of Sciences 112:E2829-E2835.
Earl DA, Bradnam K, St. John J, Darling A, Lin D, Faas J, Yu HOK, Vince B, Zerbino DR, Diekhans M, et al. 2011. Assemblathon 1: A competitive assessment of de novo short read assembly methods. Genome Research.
Edgar RC. 2004a. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32.
Edgar RC. 2004b. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research 32:1792-1797.
Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, Peluso P, Rank D, Baybayan P, Bettman B, et al. 2009. Real-Time DNA Sequencing from Single Polymerase Molecules. Science 323:133-138.
Farris SM. 2005. Evolution of insect mushroom bodies: old clues, new insights. Arthropod Structure & Development 34:211-234.
Fernández MdlP, Chan Y-B, Yew JY, Billeter J-C, Dreisewerd K, Levine JD, Kravitz EA. 2010. Pheromonal and Behavioral Cues Trigger Male-to-Female Aggression in <italic>Drosophila</italic>. PLoS Biol 8:e1000541.
Firestein S. 2001. How the olfactory system makes sense of scents. Nature 413:211-218.
Fujii S, Yavuz A, Slone J, Jagge C, Song X, Amrein H. 2015. Drosophila Sugar Receptors in Sweet Taste Perception, Olfaction, and Internal Nutrient Sensing. Current Biology 25:621-627.
Futahashi R, Kawahara-Miki R, Kinoshita M, Yoshitake K, Yajima S, Arikawa K, Fukatsu T. 2015. Extraordinary diversity of visual opsin genes in dragonflies. Proceedings of the National Academy of Sciences of the United States of America 112:E1247-E1256.
Gao Q, 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, Ritchie MG. 2008. Drosophila chemoreceptor gene evolution: selection, specialization and genome size. Molecular Ecology 17:1648-1657.
Garrison NL, Rodriguez J, Agnarsson I, Coddington JA, Griswold CE, Hamilton CA, Hedin M, Kocot KM, Ledford JM, Bond JE. 2016. Spider phylogenomics: untangling the Spider Tree of Life. PeerJ 4:e1719.
Getahun MN, Olsson SB, Lavista-Llanos S, Hansson BS, Wicher D. 2013. Insect Odorant Response Sensitivity Is Tuned by Metabotropically Autoregulated Olfactory Receptors. PLoS ONE 8:e58889.
Grassi, A.; Anfora, G.; Maistri, S.; Maddalena, G.; De Cristofaro, A.; Savini, G.; Ioriatti, C. 2014. Development and efficacy of Droskidrink, a food bait for trapping *Drosophila suzukii*. In: IOBC VIII Workshop on Integrated Soft Fruit Production, Vigalzano di Pergine, (TN), Italy, 26-28 May 2014.
Grbic M, Van Leeuwen T, Clark RM, Rombauts S, Rouze P, Grbic V, Osborne EJ, Dermauw W, Thi Ngoc PC, Ortego F, et al. 2011. The genome of Tetranychus urticae reveals herbivorous pest adaptations. Nature 479:487-492.
Gulia-Nuss M, Nuss AB, Meyer JM, Sonenshine DE, Roe RM, Waterhouse RM, Sattelle DB, de la Fuente J, Ribeiro JM, Megy K, et al. 2016. Genomic insights into the Ixodes scapularis tick vector of Lyme disease. Nat Commun 7.
Guo S, Kim J. 2007a. Molecular Evolution of Drosophila Odorant Receptor Genes. Molecular Biology and Evolution 24:1198-1207.
Guo S, Kim J. 2007b. Molecular evolution of Drosophila odorant receptor genes. Mol Biol Evol 24.

Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, et al. 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat. Protocols 8:1494-1512.

Hahn MW, Han MV, Han S-G. 2007. Gene Family Evolution across 12 <italic>Drosophila</italic> Genomes. PLoS Genet 3:e197.

Hall JC. 2002. Courtship lite: a personal history of reproductive behavioral neurogenetics in Drosophila. Journal of Neurogenetics 16:136-163.

Hallem EA, Carlson JR. 2004. Coding of Odors by a Receptor Repertoire. Cell 125:143-160.

Hallem EA, Dahanukar A, Carlson JR. 2006. INSECT ODOR AND TASTE RECEPTORS. Annual Review of Entomology 51:113-135.

Harada E, Nakagawa J, Asano T, Taoka M, Sorimachi H, Ito Y, Aigaki T, Matsuo T. 2012a. Functional Evolution of Duplicated Odorant-Binding Protein Genes, <italic>Obp57d</italic> and <italic>Obp57e</italic>, in <italic>Drosophila</italic>. PLoS ONE 7:e29710.

Harada E, Nakagawa J, Asano T, Taoka M, Sorimachi H, Ito Y, Aigaki T, Matsuo T. 2012b. Functional Evolution of Duplicated Odorant-Binding Protein Genes,*Obp57d* and *Obp57e* in Drosophila. PLoS ONE 7:e29710.

Hildebrand JG, Shepherd GM. 1997. MECHANISMS OF OLFACTORY DISCRIMINATION:Converging Evidence for Common Principles Across Phyla. Annual Review of Neuroscience 20:595-631.

Hiroi M, Marion-Poll F, Tanimura T. 2002. Differentiated Response to Sugars among Labellar Chemosensilla in Drosophila. Zoological Science 19:1009-1018.

Holt C, Yandell M. 2011. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. BMC Bioinformatics 12:491.

Hopf TA, Morinaga S, Ihara S, Touhara K, Marks DS, Benton R. 2015. Amino acid coevolution reveals three-dimensional structure and functional domains of insect odorant receptors. Nat Commun 6.

Horner DS, Pavesi G, Castrignanò T, De Meo PDO, Liuni S, Sammeth M, Picardi E, Pesole G. 2010. Bioinformatics approaches for genomics and post genomics applications of next-generation sequencing. Briefings in Bioinformatics 11:181-197.

Huang XaM, A. 1999. A DNA sequence assembly program. Genome Res., :868-877.

Ioriatti C, Walton V, Dalton D, Anfora G, Grassi A, Maistri S, Mazzoni V. 2015b. Drosophila suzukii (Diptera: Drosophilidae) and its Potential Impact to Wine Grapes During Harvest in Two Cool Climate Wine Grape Production Regions. Journal of Economic Entomology 108:1148-1155.

Jannin J. 1999. Tsetse biology and ecology: their role in the epidemiology and control of trypanosomiasis. Bulletin of the World Health Organization 77:450-450.

Jiang Y, Matsunami H. 2015. Mammalian odorant receptors: functional evolution and variation. Current Opinion in Neurobiology 34:54-60.

Jones WD, Cayirlioglu P, Grunwald Kadow I, Vosshall LB. 2007. Two chemosensory receptors together mediate carbon dioxide detection in Drosophila. Nature 445:86-90.

Jones WD, Nguyen T-AT, Kloss B, Lee KJ, Vosshall LB. 2005. Functional conservation of an insect odorant receptor gene across 250 million years of evolution. Current Biology 15:R119-R121.

Joseph RM, Carlson JR. 2015. Drosophila Chemoreceptors: A Molecular Interface Between the Chemical World and the Brain. Trends in Genetics 31:683-695.

Kain P, Boyle SM, Tharadra SK, Guda T, Pham C, Dahanukar A, Ray A. 2013. Odour receptors and neurons for DEET and new insect repellents. Nature 502:507-512.

Kanzawa T. 1939. Studies on Drosophila suzukii Mats. Review of Applied Entomology 24:315.

Keesey I, Knaden M, Hansson B. 2015. Olfactory Specialization in Drosophila suzukii Supports an Ecological Shift in Host Preference from Rotten to Fresh Fruit. Journal of Chemical Ecology:1-8.
Kirkness EF, Haas BJ, Sun W, Braig HR, Perotti MA, Clark JM, Lee SH, Robertson HM, Kennedy RC, Elhaik E, et al. 2010. Genome sequences of the human body louse and its primary endosymbiont provide insights into the permanent parasitic lifestyle. Proceedings of the National Academy of Sciences 107:12168-12173.
Klun JA, Strickman D, Rowton E, Williams J, Kramer M, Roberts D, Debboun M. 2004. Comparative Resistance of Anopheles albimanus and Aedes aegypti to N,N-Diethyl-3-methylbenzamide (Deet) and 2-Methylpiperidinyl-3-cyclohexen-1-carboxamide (AI3-37220) in Laboratory Human-Volunteer Repellent Assays. Journal of Medical Entomology 41:418-422.
Koh T-W, He Z, Gorur-Shandilya S, Menuz K, Larter Nikki K, Stewart S, Carlson John R. 2014. The Drosophila IR20a Clade of Ionotropic Receptors Are Candidate Taste and Pheromone Receptors. Neuron 83:850-865.
Korenromp E. 2005. MALARIA INCIDENCE ESTIMATES AT COUNTRY LEVEL FOR THE YEAR 2004
– PROPOSED ESTIMATES AND DRAFT REPORT - for the RBM Monitoring and Evaluation Reference Group & MERG Task Force on Malaria Morbidity: World Health Organization, Geneva.
Korf I. 2004. Gene finding in novel genomes. BMC Bioinformatics 5:59.
Kreher SA, Kwon JY, Carlson JR. 2005. The Molecular Basis of Odor Coding in the Drosophila Larva. Neuron 46:445-456.
Kruse SW, Zhao R, Smith DP, Jones DNM. 2003. Structure of a specific alcohol-binding site defined by the odorant binding protein LUSH from Drosophila melanogaster. Nat Struct Mol Biol 10:694-700.
Kuhar TP, Mori K, Dickens JC. 2006. Potential of a synthetic aggregation pheromone for integrated pest management of Colorado potato beetle. Agricultural and Forest Entomology 8:77-81.
Kurtovic A, Widmer A, Dickson BJ. 2007. A single class of olfactory neurons mediates behavioural responses to a Drosophila sex pheromone. Nature 446:542-546.
Kurtz S, Narechania A, Stein JC, Ware D. 2008. A new method to compute K-mer frequencies and its application to annotate large repetitive plant genomes. BMC Genomics 9:1-18.
Kwon JY, Dahanukar A, Weiss LA, Carlson JR. 2007. The molecular basis of CO2 reception in Drosophila. Proceedings of the National Academy of Sciences 104:3574-3578.
Le SQ, Gascuel O. 2008. An Improved General Amino Acid Replacement Matrix. Molecular Biology and Evolution 25:1307-1320.
Leal WS. 2013. Odorant reception in insects: roles of receptors, binding proteins, and degrading enzymes. Annu Rev Entomol 58.
Leal WS, Nikonova L, Peng G. 1999. Disulfide structure of the pheromone binding protein from the silkworm moth, Bombyx mori. FEBS Letters 464:85-90.
Lebreton S, Grabe V, Omondi AB, Ignell R, Becher PG, Hansson BS, Sachse S, Witzgall P. 2014. Love makes smell blind: mating suppresses pheromone attraction in Drosophila females via Or65a olfactory neurons. Scientific Reports 4:7119.
Lee Y, Moon SJ, Montell C. 2009. Multiple gustatory receptors required for the caffeine response in Drosophila. Proceedings of the National Academy of Sciences 106:4495-4500.
Leite NR, Krogh R, Xu W, Ishida Y, Iulek J, Leal WS, Oliva G. 2009. Structure of an Odorant-Binding Protein from the Mosquito <italic>Aedes aegypti</italic> Suggests a Binding Pocket Covered by a pH-Sensitive ?Lid? PLoS ONE 4:e8006.
Li W, Godzik A. 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 22:1658-1659.

Librado P, Vieira FG, Rozas J. 2011. BadiRate: Estimating Family Turnover Rates by Likelihood-Based Methods. Bioinformatics.

Ling F, Dahanukar A, Weiss LA, Kwon JY, Carlson JR. 2014. The Molecular and Cellular Basis of Taste Coding in the Legs of Drosophila. The Journal of Neuroscience 34:7148-7164.

Liu W, Liang X, Gong J, Yang Z, Zhang Y-H, Zhang J-X, Rao Y. 2011. Social regulation of aggression by pheromonal activation of Or65a olfactory neurons in Drosophila. Nat Neurosci 14:896-902.

Loman NJ, Misra RV, Dallman TJ, Constantinidou C, Gharbia SE, Wain J, Pallen MJ. 2012. Performance comparison of benchtop high-throughput sequencing platforms. Nat Biotech 30:434-439.

M.H. Schulz DRZ, M. Vingron and Ewan Birney. 2012. Oases: Robust de novo RNA-seq assembly across the dynamic range of expression levels. Bioinformatics.

Matsuo T. 2008. Rapid Evolution of Two Odorant-Binding Protein Genes, Obp57d and Obp57e, in the Drosophila melanogaster Species Group. Genetics 178:1061-1072.

Matsuo T, Sugaya S, Yasukawa J, Aigaki T, Fuyama Y. 2007. Odorant-binding proteins OBP57d and OBP57e affect taste perception and host-plant preference in Drosophila sechellia. PLoS Biol 5.

McBride CS. 2007. Rapid evolution of smell and taste receptor genes during host specialization in Drosophila sechellia. Proceedings of the National Academy of Sciences 104:4996-5001.

McBride CS, Arguello JR. 2007. Five Drosophila Genomes Reveal Nonneutral Evolution and the Signature of Host Specialization in the Chemoreceptor Superfamily. Genetics 177:1395-1416.

McKenzie SK, Oxley PR, Kronauer DJ. 2014. Comparative genomics and transcriptomics in ants provide new insights into the evolution and function of odorant binding and chemosensory proteins. BMC Genomics 15:1-14.

Miller MB, Bassler BL. 2001. Quorum Sensing in Bacteria. Annual Review of Microbiology 55:165-199.

Min S, Ai M, Shin SA, Suh GSB. 2013. Dedicated olfactory neurons mediating attraction behavior to ammonia and amines in Drosophila. Proceedings of the National Academy of Sciences 110:E1321–E1329.

Misof BLiu SMeusemann KPeters RSDonath AMayer CFrandsen PBWare JFlouri TBeutel RG, et al. 2014. Phylogenomics resolves the timing and pattern of insect evolution. Science 346:763-767.

Misra S, Crosby MA, Mungall CJ, Matthews BB, Campbell KS, Hradecky P, Huang Y, Kaminker JS, Millburn GH, Prochnik SE, et al. 2002. Annotation of the Drosophila melanogaster euchromatic genome: a systematic review. Genome Biology 3:1-22.

Missbach C, Dweck HK, Vogel H, Vilcinskas A, Stensmyr MC, Hansson BS, Grosse-Wilde E. 2014. Evolution of insect olfactory receptors. eLIFE 3.

Mistry J, Finn RD, Eddy SR, Bateman A, Punta M. 2013. Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. Nucleic Acids Research 41:e121.

Mitsui H, Beppu K, Kimura MT. 2010. Seasonal life cycles and resource uses of flower- and fruit-feeding drosophilid flies (Diptera: Drosophilidae) in central Japan. Entomological Science 13:60-67.

Montell C. 2013. Gustatory Receptors: Not Just for Good Taste. Current Biology 23:R929-R932.

Moon SJ, Lee Y, Jiao Y, Montell C. 2009. A Drosophila Gustatory Receptor Essential for Aversive Taste and Inhibiting Male-to-Male Courtship. Current Biology 19:1623-1627.

Muench D, Galizia CG. 2015. DoOR 2.0 - Comprehensive Mapping of Drosophila melanogaster Odorant Responses. bioRxiv.

Murphy EJ, Booth JC, Davrazou F, Port AM, Jones DNM. 2013. Interactions of Anopheles gambiae Odorant-binding Proteins with a Human-derived Repellent: IMPLICATIONS FOR THE MODE OF ACTION OF N,N-DIETHYL-3-METHYLBENZAMIDE (DEET). Journal of Biological Chemistry 288:4475-4485.
Nagarajan N, Pop M. 2013. Sequence assembly demystified. Nat Rev Genet 14:157-167.
Nei M, Rooney AP. 2005. Concerted and Birth-and-Death Evolution of Multigene Families. Annual Review of Genetics 39:121-152.
Nguyen N-pD, Mirarab S, Kumar K, Warnow T. 2015. Ultra-large alignments using phylogeny-aware profiles. Genome Biology 16:1-15.
Niimura Y, Matsui A, Touhara K. 2014. Extreme expansion of the olfactory receptor gene repertoire in African elephants and evolutionary dynamics of orthologous gene groups in 13 placental mammals. Genome Research.
Ometto L, Cestaro A, Ramasamy S, Grassi A, Revadi S, Siozios S, Moretto M, Fontana P, Varotto C, Pisani D, et al. 2013. Linking Genomics and Ecology to Investigate the Complex Evolution of an Invasive Drosophila Pest. Genome Biology and Evolution 5:745-757.
Park SK, Shanbhag SR, Wang Q, Hasan G, Steinbrecht RA, Pikielny CW. 2000. Expression patterns of two putative odorant-binding proteins in the olfactory organs of Drosophila melanogaster have different implications for their functions. Cell Tissue Res 300.
Pellegrino M, Nakagawa T. 2009. Smelling the difference: controversial ideas in insect olfaction. Journal of Experimental Biology 212:1973-1979.
Pelosi P, Iovinella I, Felicioli A, Dani FR. 2014. Soluble proteins of chemical communication: an overview across arthropods. Frontiers in Physiology 5:320.
Pelosi P, Zhou JJ, Ban LP, Calvello M. 2006. Soluble proteins in insect chemical communication. Cell Mol Life Sci 63.
Penalva-Arana DC, Lynch M, Robertson H. 2009. The chemoreceptor genes of the waterflea Daphnia pulex: many Grs but no Ors. BMC Evolutionary Biology 9:79.
Piersanti S, Frati F, Conti E, Gaino E, Rebora M, Salerno G. 2014. First evidence of the use of olfaction in Odonata behaviour. Journal of Insect Physiology 62:26-31.
Pimentel D. 1991. Diversification of biological control strategies in agriculture. Crop Protection 10:243-253.
Ramasamy S, Ometto L, Crava CM, Revadi S, Kaur R, Horner D, Pisani D, Dekker T, Anfora G, Rota-Stabelli O. 2016. The evolution of olfactory gene families in *Drosophila* and the genomic basis of chemical-ecological adaptation in *Drosophila suzukii*. Under Review.
Rebora M, Salerno G, Piersanti S, Dell'Otto A, Gaino E. 2012. Olfaction in dragonflies: Electrophysiological evidence. Journal of Insect Physiology 58:270-277.
Revadi S, Vitagliano S, Rossi Stacconi MV, Ramasamy S, Mansourian S, Carlin S, Vrhovsek U, Becher PG, Mazzoni V, Rota-Stabelli O, et al. 2015. Olfactory responses of Drosophila suzukii females to host plant volatiles. Physiological Entomology 40:54-64.
Ribeiro FJ, Przybylski D, Yin S, Sharpe T, Gnerre S, Abouelleil A, Berlin AM, Montmayeur A, Shea TP, Walker BJ, et al. 2012. Finished bacterial genomes from shotgun sequence data. Genome Research 22:2270-2277.
Rice MJ. 1977. Blowfly ovipositor receptor neurone sensitive to monovalent cation concentration. Nature 268:747-749.
Rideout EJ, Dornan AJ, Neville MC, Eadie S, Goodwin SF. 2010. Control of sexual differentiation and behavior by the doublesex gene in Drosophila melanogaster. Nat Neurosci 13:458-466.
Robertson HM, Gadau J, Wanner KW. 2010. The insect chemoreceptor superfamily of the parasitoid jewel wasp Nasonia vitripennis. Insect Molecular Biology 19:121-136.
Robertson HM, Warr CG, Carlson JR. 2003. Molecular evolution of the insect chemoreceptor gene superfamily in Drosophila melanogaster. Proc. Natl. Acad. Sci:14537–14542.

Robinson GE, Hackett KJ, Purcell-Miramontes M, Brown SJ, Evans JD, Goldsmith MR, Lawson D, Okamuro J, Robertson HM, Schneider DJ. 2011. Creating a Buzz About Insect Genomes. Science 331:1386-1386.
Roelofs WL, Rooney AP. 2003. Molecular genetics and evolution of pheromone biosynthesis in Lepidoptera. Proceedings of the National Academy of Sciences 100:9179-9184.
Rossi-Stacconi MV, Kaur R, Mazzoni V, Ometto L, Grassi A, Gottardello A, Rota-Stabelli O, Anfora G. 2016. Multiple lines of evidence for reproductive winter diapause in the invasive pest Drosophila suzukii: useful clues for control strategies. Journal of Pest Science:1-12.
Rota-Stabelli O, Blaxter M, Anfora G. 2013a. Drosophila suzukii. Current Biology 23:R8-R9.
Rota-Stabelli O, Daley Allison C, Pisani D. 2013b. Molecular Timetrees Reveal a Cambrian Colonization of Land and a New Scenario for Ecdysozoan Evolution. Current Biology 23:392-398.
Rota-Stabelli O, Lartillot N, Philippe H, Pisani D. 2013c. Serine Codon-Usage Bias in Deep Phylogenomics: Pancrustacean Relationships as a Case Study. Systematic Biology 62:121-133.
Sanchez-Gracia A, Vieira FG, Rozas J. 2009. Molecular evolution of the major chemosensory gene families in insects. Heredity 103:208-216.
Sargsyan V, Getahun MN, Lavista Llanos S, Olsson SB, Hansson BS, Wicher D. 2011. Phosphorylation via PKC regulates the function of the Drosophila odorant coreceptor. Frontiers in Cellular Neuroscience 5.
Sato K, Pellegrino M, Nakagawa T, Nakagawa T, Vosshall LB, Touhara K. 2008. Insect olfactory receptors are heteromeric ligand-gated ion channels. Nature 452:1002-1006.
Scott K, Brady Jr R, Cravchik A, Morozov P, Rzhetsky A, Zuker C, Axel R. 2001. A Chemosensory Gene Family Encoding Candidate Gustatory and Olfactory Receptors in Drosophila. Cell 104:661-673.
Shanbhag S, Hekmat-Scafe D, Kim M, Park S, Carlson J, Pikielny C, Smith D, Steinbrecht R. 2001. Expression mosaic of odorant-binding proteins in Drosophila olfactory organs. Microsc Res Tech. 55:297-306.
Shanbhag SR, Müller B, Steinbrecht RA. 1999. Atlas of olfactory organs of Drosophila melanogaster: 1. Types, external organization, innervation and distribution of olfactory sensilla. International Journal of Insect Morphology and Embryology 28:377-397.
Sharma P, Giribet G. 2014. A revised dated phylogeny of the arachnid order Opiliones. Frontiers in Genetics 5.
Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJM, Birol İ. 2009. ABySS: A parallel assembler for short read sequence data. Genome Research 19:1117-1123.
Slater GSC, Birney E. 2005. Automated generation of heuristics for biological sequence comparison. BMC Bioinformatics 6:1-11.
Smit A, Hubley R, Green P. 1996-2010. RepeatMasker Open-3.0.

Stamatakis A. 2014. RAxML Version 8: A tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies. Bioinformatics.
Stanke M, Steinkamp R, Waack S, Morgenstern B. 2004. AUGUSTUS: a web server for gene finding in eukaryotes. Nucleic Acids Research 32:W309-W312.
Stanke M, Waack S. 2003. Gene prediction with a hidden Markov model and a new intron submodel. Bioinformatics 19:ii215-ii225.
Steiger S, Schmitt T, Schaefer HM. 2011. The origin and dynamic evolution of chemical information transfer. Proceedings of the Royal Society of London B: Biological Sciences.
Stensmyr Marcus C, Dweck Hany KM, Farhan A, Ibba I, Strutz A, Mukunda L, Linz J, Grabe V, Steck K, Lavista-Llanos S, et al. 2012. A Conserved Dedicated Olfactory Circuit for Detecting Harmful Microbes in Drosophila. Cell 151:1345-1357.

Stocker RF. 1994. The organization of the chemosensory system in Drosophila melanogaster: a rewiew. Cell and Tissue Research 275:3-26.

Stockinger P, Kvitsiani D, Rotkopf S, Tirián L, Dickson BJ. 2005. Neural Circuitry that Governs Drosophila Male Courtship Behavior. Cell 121:795-807.

Su C-Y, Menuz K, Carlson JR. 2009. Olfactory Perception: Receptors, Cells, and Circuits. Cell 139:45-59.

Symonds MRE, Wertheim B. 2005. The mode of evolution of aggregation pheromones in Drosophila species. Journal of Evolutionary Biology 18:1253-1263.

Tegoni M, Pelosi P, Vincent F, Spinelli S, Campanacci V, Grolli S, Ramoni R, Cambillau C. 2000. Mammalian odorant binding proteins. Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology 1482:229-240.

Thode AB, Kruse SW, Nix JC, Jones DNM. 2008. The Role of Multiple Hydrogen-Bonding Groups in Specific Alcohol Binding Sites in Proteins: Insights from Structural Studies of LUSH. Journal of Molecular Biology 376:1360-1376.

Thorne N, Amrein H. 2008. Atypical expression of Drosophila gustatory receptor genes in sensory and central neurons. The Journal of Comparative Neurology 506:548-568.

van der Goes van Naters W, Carlson JR. 2006. Insects as chemosensors of humans and crops. Nature 444:302-307.

Vieira F, Sanchez-Gracia A, Rozas J. 2007. Comparative genomic analysis of the odorant-binding protein family in 12 Drosophila genomes: purifying selection and birth-and-death evolution. Genome Biology 8:R235.

Vieira FG, 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 Biology and Evolution.

Vogt RG, Rogers ME, Franco M-d, Sun M. 2002. A comparative study of odorant binding protein genes: differential expression of the PBP1-GOBP2 gene cluster in Manduca sexta (Lepidoptera) and the organization of OBP genes in Drosophila melanogaster (Diptera). Journal of Experimental Biology 205:719-744.

Vosshall LB, Amrein H, Morozov PS, Rzhetsky A, Axel R. 1999. A Spatial Map of Olfactory Receptor Expression in the Drosophila Antenna. Cell 96:725-736.

Vosshall LB, Stensmyr MC. 2005. Wake Up and Smell the Pheromones. Neuron 45:179-181.

Waddell PJ, Steel MA. 1997. General Time-Reversible Distances with Unequal Rates across Sites: Mixing Γ and Inverse Gaussian Distributions with Invariant Sites. Molecular Phylogenetics and Evolution 8:398-414.

Walsh DB, Bolda MP, Goodhue RE, Dreves AJ, Lee J, Bruck DJ, Walton VM, O'Neal SD, Zalom FG. 2011. Drosophila suzukii (Diptera: Drosophilidae): Invasive Pest of Ripening Soft Fruit Expanding Its Geographic Range and Damage Potential. Journal of Integrated Pest Management 2:1-7.

Wang L, Anderson DJ. 2010. Identification of an aggression-promoting pheromone and its receptor neurons in Drosophila. Nature 463:227-231.

Wang L, Han X, Mehren J, Hiroi M, Billeter J-C, Miyamoto T, Amrein H, Levine JD, Anderson DJ. 2011. Hierarchical chemosensory regulation of male-male social interactions in Drosophila. Nat Neurosci 14:757-762.

Watanabe K, Toba G, Koganezawa M, Yamamoto D. 2011. Gr39a, a Highly Diversified Gustatory Receptor in Drosophila, has a Role in Sexual Behavior. Behavior Genetics 41:746-753.

Waterhouse RM, Tegenfeldt F, Li J, Zdobnov EM, Kriventseva EV. 2013. OrthoDB: a hierarchical catalog of animal, fungal and bacterial orthologs. Nucleic Acids Research 41:D358-D365.

Weiss LA, Dahanukar A, Kwon JY, Banerjee D, Carlson JR. 2011. The Molecular and Cellular Basis of Bitter Taste in Drosophila. Neuron 69:258-272.

Wicher D, Schafer R, Bauernfeind R, Stensmyr MC, Heller R, Heinemann SH, Hansson BS. 2008. Drosophila odorant receptors are both ligand-gated and cyclic-nucleotide-activated cation channels. Nature 452:1007-1011.

Wistrand M, Käll L, Sonnhammer ELL. 2006. A general model of G protein-coupled receptor sequences and its application to detect remote homologs. Protein Science 15:509-521.

Wyatt TD. 2003. Pheromones and Animal Behaviour: Communication by Smell and Taste: Cambrige University Press.

Xu P, Atkinson R, Jones DNM, Smith DP. 2005. Drosophila OBP LUSH Is Required for Activity of Pheromone-Sensitive Neurons. Neuron 45:193-200.

Yamamoto D, Usui-Aoki K, Shima S. 2004. Male-Specific Expression of the Fruitless Protein is not Common to all Drosophila Species. Genetica 120:267-272.

Yandell M, Bailey AM, Misra S, Shu S, Wiel C, Evans-Holm M, Celniker SE, Rubin GM. 2005. A computational and experimental approach to validating annotations and gene predictions in the Drosophila melanogaster genome. Proceedings of the National Academy of Sciences of the United States of America 102:1566-1571.

Yandell M, Ence D. 2012. A beginner's guide to eukaryotic genome annotation. Nat Rev Genet 13:329-342.

Zerbino DR, Birney E. 2008. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. Genome Research 18:821-829.

Zhang H-J, Anderson AR, Trowell SC, Luo AR, Xiang Z-H, Xia Q-Y. 2011. Topological and Functional Characterization of an Insect Gustatory Receptor. PLoS ONE 6:e24111.

Zhou J-J, Robertson G, He X, Dufour S, Hooper AM, Pickett JA, Keep NH, Field LM. 2009. Characterisation of Bombyx mori Odorant-binding Proteins Reveals that a General Odorant-binding Protein Discriminates Between Sex Pheromone Components. Journal of Molecular Biology 389:529-545.

Appendix A – Supporting Materials

Table S1. Annotation information where A) and B) are the list of incomplete chemosensory genes and proteins in Trentino strain of *D. suzukii* replaced with sequences from Californian strain of *D. suzukii;* C) Possible allelic variants identified from *D. suzukii*, ignored due to heterologous nature of the genome; D) Putative isoforms in *D. suzukii* that have shared exons as a result of alternative splicing are ignored from Birth – Death analysis.

Annotation	Olfactory Receptor	Odorant Binding Protein
A)Partial genes	*Or33a,Or33b,Or42a, Or43a,Or46aA,Or46aB, Or47a,Or47b,Or49a2*	*Obp47b,Obp5,Obp73a*
B)Missing genes	-	*Obp50b,Obp50d,Obp51a, Obp56c,Obp56g, Obp57e*
C)Allelic variants removed	-	*Obp19b,Obp56a, Obp56h,Obp59a*
D)possible isoforms ignored in bd	*Or42a(B,C),Or46a(B) , Or69a (B,C,D,E)*	

Table S2. List of odorant genes (OR, OBP, aIR) in 14 *Drosophila* species. The list for all but *D. suzukii* and *D. biarmipes* is taken from previous studies as mentioned in the Methods section (Ramasamy et al. 2016) Cases of pseudogenes are assumed as gene loss here.

FAM_ID	*SIM*	*SEC*	*MEL*	*YAK*	*ERE*	*SUZ*	*BIA*	*ANA*	*PSE*	*PER*	*WIL*	*VIR*	*MOJ*	*GRI*
OR10A	1	1	1	1	1	1	1	1	1	1	1	1	1	1
OR13A	1	0	1	1	1	1	1	1	1	1	1	1	1	1
OR19A	1	1	2	1	1	2	1	1	1	1	1	1	1	1
OR1A	1	1	1	1	1	1	1	1	1	1	1	1	1	0
OR22A	1	1	1	1	1	0	0	4	2	2	1	1	1	2
OR22B	1	0	1	0	0	1	1	0	0	0	0	0	0	0
OR22C	1	1	1	1	1	1	1	1	1	1	1	1	1	2

Or23a	1	1	1	1	1	4	1	1	1	1	1	1	2	1
Or24a	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Or2a	1	1	1	1	1	1	1	1	1	1	1	1	1	0
Or30a	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Or33a	1	1	1	1	1	1	1	2	1	1	1	0	0	0
Or33b	1	1	1	1	1	1	1	3	2	2	1	0	0	0
Or33c	1	1	1	1	1	1	0	1	1	1	1	1	1	1
Or35a	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Or42a	1	1	1	1	1	1	1	1	2	2	1	2	2	1
Or42b	1	1	1	1	1	1	1	1	1	1	1	1	1	8
Or43a	1	1	1	1	1	1	1	1	1	0	1	1	1	1
Or43b	1	1	1	1	1	1	1	1	1	1	1	0	0	0
Or45a	1	1	1	1	1	1	1	1	1	1	1	0	0	0
Or45b	1	1	1	1	1	1	1	1	1	1	0	1	1	1
Or46a	1	1	1	1	1	1	1	1	1	1	1	1	1	4
Or47a	1	1	1	1	1	1	1	1	1	1	1	0	0	0
Or47b	1	1	1	1	1	1	1	1	1	1	1	1	2	1
Or49a	1	1	1	1	1	2	1	1	2	2	1	1	1	2
Or49b	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Or56a	1	1	1	1	1	1	1	1	2	1	1	1	1	1
Or59a	1	1	1	1	1	2	1	1	1	1	2	3	1	2
Or59b	1	1	1	1	1	1	1	1	1	1	1	2	1	2
Or59c	1	1	1	1	1	2	1	1	1	1	1	0	0	0
Or63a	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Or65a	1	1	1	1	1	1	1	3	0	0	2	0	0	0
Or65b	1	1	1	1	1	0	0	0	5	2	0	0	1	0
Or65c	1	0	1	2	1	2	2	0	0	0	0	0	0	0
Or67a	3	1	1	2	2	4	4	1	1	0	3	2	2	2
Or67b	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Or67c	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Or67d	1	1	1	1	1	1	1	2	1	1	1	1	1	5
Or69a	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Or71a	1	1	1	1	1	1	1	1	1	1	1	1	1	2
Or74a	1	0	1	1	1	0	1	1	1	1	1	1	1	1
Or7a	1	1	1	1	1	1	1	1	0	0	1	0	0	0
Or82a	1	1	1	1	1	1	1	1	1	1	4	1	1	1
Or83a	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Or83b	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Or83c	1	1	1	1	1	1	1	1	1	1	2	2	2	1
Or85a	1	1	1	1	1	0	1	0	0	0	1	0	2	0
Or85b	1	1	1	1	1	1	1	1	1	0	0	0	0	0
Or85c	1	1	1	1	1	1	1	1	1	0	1	1	2	2
Or85d	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Or85e	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Or85f	1	1	1	1	1	1	1	1	1	1	6	1	1	1
Or88a	1	1	1	1	1	1	1	0	1	1	1	1	1	1
Or92a	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Or94a	1	1	1	1	1	1	1	1	1	1	1	0	1	1
Or94b	1	1	1	1	1	1	1	1	1	0	1	1	1	1
Or98a	1	1	1	2	2	1	1	3	3	2	4	3	3	2
Or98b	1	0	1	1	1	1	1	1	1	1	1	1	0	0
Or9a	1	1	1	1	1	1	1	1	1	1	1	1	1	1
OrN1	0	0	0	0	0	0	0	0	1	0	3	1	1	1
OrN2	0	0	0	0	0	0	0	0	0	0	0	2	5	1
lush	1	1	1	2	1	1	1	1	1	1	1	1	1	1
obp18a	1	0	1	1	0	0	1	1	0	0	7	0	1	1
obp19a	1	1	1	1	1	1	1	1	1	1	1	1	1	1
obp19b	1	1	1	1	1	1	1	1	1	1	1	1	1	1
obp19c	1	1	1	1	1	1	1	1	1	1	1	1	1	1
obp19d	1	1	1	1	1	1	1	1	1	1	1	1	1	1

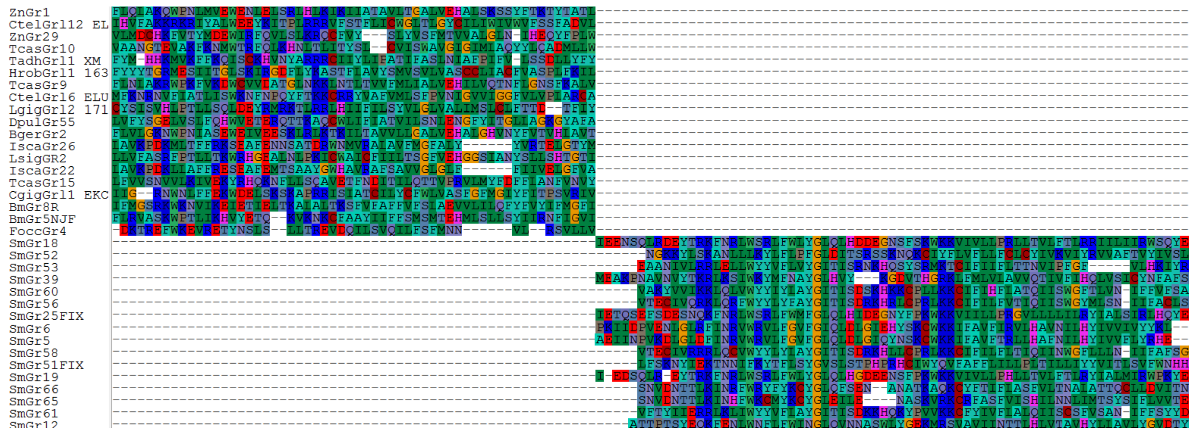
obp22a	1	1	1	1	2	1	0	0	0	0	0	0	0	0
obp28a	1	1	1	1	1	1	1	1	1	1	1	1	1	1
obp44a	1	1	1	1	1	1	1	1	1	1	1	1	1	1
obp46a	1	1	1	1	1	2	1	1	1	1	1	1	1	1
obp47a	1	1	1	1	1	2	1	1	1	1	1	0	1	1
obp47b	1	1	1	1	1	1	1	1	1	1	1	1	1	1
obp49a	1	1	1	1	1	1	1	1	1	1	1	1	1	1
obp50a	1	1	1	1	1	1	1	1	1	1	1	1	1	1
obp50b	1	1	1	1	1	1	1	1	1	1	1	0	0	0
obp50c	1	1	1	1	1	1	1	1	1	1	3	1	1	1
obp50d	1	1	1	1	1	1	1	1	1	1	2	0	1	1
obp50e	1	1	1	1	1	1	1	1	1	1	1	1	1	1
obp51a	1	1	1	2	0	1	1	0	0	0	1	0	0	0
obp56a	1	1	1	1	1	1	1	1	1	1	1	1	1	1
obp56b	1	1	1	1	1	1	1	1	1	1	1	1	1	1
obp56c	1	1	1	1	1	1	1	1	1	1	1	1	1	1
obp56d	1	1	1	1	1	1	1	1	1	1	1	1	1	1
obp56e	1	1	1	1	1	1	0	1	1	1	1	1	2	0
obp56f	1	1	1	1	1	0	0	3	0	0	0	1	0	0
obp56g	1	1	1	1	1	1	1	1	2	2	5	1	1	1
obp56h	1	1	1	1	1	1	1	1	1	1	1	1	1	1
obp56i	1	1	1	1	1	0	0	0	0	0	0	0	0	0
obp57a	1	1	1	1	1	1	1	1	0	0	1	0	0	0
obp57b	1	1	1	1	1	1	1	0	0	0	0	0	0	0
obp57c	1	1	1	1	1	1	1	2	1	1	3	2	2	2
obp57d	1	1	1	1	1	3	3	1	1	1	1	0	0	0
obp57e	1	1	1	1	1	1	1	0	0	0	0	0	0	0
obp58b	1	1	1	1	1	1	1	1	1	1	1	1	1	3
obp58c	1	1	1	1	1	1	1	1	1	1	1	1	1	3
obp58d	1	1	1	1	1	1	1	1	1	1	1	1	1	1
obp59a	1	1	1	1	1	1	1	0	1	1	1	1	1	1
obp69a	1	1	1	1	1	1	1	1	1	1	1	1	1	1
obp73a	1	1	1	1	1	1	1	1	1	1	1	1	1	1
obp83a	1	1	1	1	1	1	1	1	1	1	1	2	2	2
obp83b	1	1	1	1	1	1	1	1	1	1	1	0	0	0
obp83cd	1	1	1	1	1	1	1	1	1	1	1	1	0	1
obp83ef	1	1	1	1	1	1	1	1	1	1	1	1	1	1
obp83g	1	1	1	1	1	1	1	1	1	1	1	1	1	1
obp84a	1	1	1	1	1	1	1	1	1	1	1	1	1	1
obp85a	1	1	1	1	1	1	1	1	1	1	1	1	1	1
obp8a	1	1	1	1	1	1	1	0	1	1	1	1	1	0
obp93a	1	1	1	1	1	1	1	1	1	1	1	1	1	1
obp99a	1	1	1	2	1	1	1	1	1	1	1	1	1	1
obp99b	1	1	1	1	0	1	1	3	1	1	1	1	1	1
obp99c	1	1	1	1	1	1	1	1	1	1	1	1	1	1
obp99d	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Ir8a	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Ir21a	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Ir25a	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Ir31a	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Ir40a	1	1	1	1	1	1	1	1	1	1	1	1	1	
Ir41a	1	1	1	1	1	1	1	1	1	1	1	1	1	1
IR60a	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Ir75a	1	0	1	1	1	1	1	1	1	1	1	1	1	1
Ir75b	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Ir75c	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Ir75d	1	1	1	1	1	1	1	1	1	1	1	1	2	1
Ir76a	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Ir76b	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Ir64a	1	1	1	1	1	1	1	1	1	1	1	1	1	1

IR68A	1	1	1	1	1	1	1	1	1	1	1	1	1	1
IR84A	1	1	1	1	1	1	1	1	1	1	1	1	1	1
IR92A	1	1	1	1	1	1	1	1	1	1	1	1	1	1
IR93A	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Table S3. Overrepresentation of ligands in duplicates/lost ORs in *D. suzukii.*

Ligand	Type	10 ORs of figure 4	27 other ORs	Normalised Skew
1-Hexanol	alcohol	4	6	0.3
E3-hexenol	alcohol	3	3	0.5
3-octanol	alcohol	3	7	0.1
heptan-2-ol	alcohol	2	0	1.0
3-methyl-2-buten-1-ol	alcohol	2	2	0.5
E2-hexenol	alcohol	2	3	0.3
1-octen-3-ol	alcohol	2	6	0.0
E2-hexenal	aldehyde	2	7	-0.1
acetophenone	aromatic	3	3	0.5
phenylacetone	aromatic	2	0	1.0
methyl-benzoate	aromatic	2	1	0.7
benzaldehyde	aromatic	2	3	0.3
ethyl-benzoate	aromatic	2	3	0.3
2-methyl-phenol	aromatic	2	5	0.1
ethyl-butyrate	ester	4	4	0.5
Iso pentyl acetate	ester	3	6	0.2
pentyl acetate	ester	3	6	0.2
butyl-proponoate	ester	2	1	0.7
methyl-hexanoate	ester	2	2	0.5
butyl-acetate	ester	2	5	0.1
ethyl-3-hydroxy-butyrate	ester	2	5	0.1
Isobutyl acetate	ester	2	5	0.1
propyl-acetate	ester	2	6	0.0
6-methyl-5-heptenone	ketone	3	3	0.5
2-Heptanone	ketone	2	9	-0.2
linalool	terpene	2	0	1.0

Figure S1. A part of the whole GR alignment showing different structures of the internal domains (likely helices) between Myriapoda (below) and the rest of Arthropoda sequences.



Appendix B - Articles

My thesis work is based on the following articles.

Ramasamy S, Ometto L, Revadi S, Kaur R, Crava C.M, Horner D, Pisani D, Dekker T, Anfora G, Rota-Stabelli O. 2016. The evolution of olfactory gene families in *Drosophila* and the genomic basis of chemical-ecological adaptation in *Drosophila suzukii.* Genome Biology and Evolution. Accepted with revisions.

Dekker T, Revadi S, Mansourian S, **Ramasamy S**, Lebreton S, Becher P, Angeli S, Rota-Stabelli O, Anfora G. 2015. Loss of Drosophila pheromone reverses its role in sexual communication in *Drosophila suzukii*. Proceedings of the Royal Society B: Biological Sciences. 282.

Revadi S, Vitagliano S, Rossi Stacconi MV, **Ramasamy S**, Mansourian S, Carlin S, Vrhovsek U, Becher PG, Mazzoni V, Rota-Stabelli O, et al. 2015. Olfactory responses of *Drosophila suzukii* females to host plant volatiles. Physiological Entomology 40:54-64.

Ometto L, Cestaro A, **Ramasamy S**, Grassi A, Revadi S, Siozios S, Moretto M, Fontana P, Varotto C, Pisani D, et al. 2013. Linking Genomics and Ecology to Investigate the Complex Evolution of an Invasive Drosophila Pest. Genome Biology and Evolution 5:745-757.

Crava C M[*], **Ramasamy S[*],** Ometto L, Anfora G, Rota-Stabelli O. Evolutionary insights into taste perception of the invasive pest, *Drosophila suzukii*. Manuscript in preparation.

Ramasamy S, Rota-Stabelli O, Pisani D. On the evolutionary origins of chemosensation in Arthropods. Manuscript in preparation.

Acknowledgements

‘பயன்தூக்கார் செய்த உதவி நயன்தூக்கின்

நன்மை கடலின் பெரிது’

- திருக்குறள்

(Kindness shown by those who weigh not what the return may be:
When you ponder right its merit, 'Tis vaster than the sea)

- Thirukural

I would like to thank:

Gianfranco and Omar to have given me this wonderful opportunity to work with insects.

Omar, my Jedi master, for teaching me all cool things about evolution and giving me the right space to explore. And for his supervision, support, and for being quite critical about my writing skills. Someday, I hope to straighten that out with a nice article. But I will keep disliking heavy metals.

David Horner to have accepted and co-ordinated me at the University of Milan.

Davide Pisani for introducing me to the Palaeo world and for his guidance throughout.

Al and Richard, for their incredible support regarding informatics and server.

Santosh, for his fair share of motivations in insect science.

Mark and Armin for their philosophical (too German) lessons and constant encouragements.

Fede and his loving family, for all their affection and care through the most difficult times, and hence teaching me altruism.

My parents for letting me free, and specially my kid brother, who finds my Biology incredibly boring to his Physics.

Finally, Italy, for its flavoursome wine, pizzas and gelato.