LIÈGE université
Gembloux
Agro-Bio Tech

Identification and functional studies of odorant receptors in aphids

Author: Tianyu Huang
Promoters: Prof. Frédéric Francis & Prof. Guirong Wang

2023

COMMUNAUTÉ FRANÇAISE DE BELGIQUE
UNIVERSITÉ DE LIÈGE – GEMBLOUX AGRO-BIO TECH

Identification and functional studies of odorant receptors in aphids

Tianyu Huang

Dissertation originale présentée en vue de l'obtention du grade de docteur en sciences agronomiques et ingénierie biologique

Promoteurs: Prof. Frédéric Francis & Prof. Guirong Wang
Année civile: 2023

Abstract

Aphids are one of the most destructive pests in agriculture, relying on their highly sensitive chemosensory system to locate host plants, mate and avoid predators. The perception of odors in aphids is associated to various gene families, including odorant receptors (ORs), gustatory receptors (GRs), ionotropic receptors (IRs), odorant-binding proteins (OBPs), and chemosensory proteins (CSPs). During the olfaction process, odors are bound and transported by OBPs and CSPs, then subsequently released to receptors, such as ORs, which are expressed on the dendritic membranes of olfactory sensory neurons. These receptors are critical as they convert chemical cues into electrical signals, which are transmitted by the central nervous system and finally generates a series of corresponding behavioral responses. Although functional characterizations have mainly focused on the OBP family in aphids, less attention has been given to investigating the function of ORs, which has severely limited our understanding of the molecular mechanisms of odor detection in aphids. Here, we identified and functionally characterized three ORs that were significantly conserved among different aphid species, playing essential role in detecting ecologically and environmentally important volatiles.

Green leaf volatiles (GLVs) are chemicals released by plants when they are damaged. Studies have shown that aphids are attracted by GLVs, but little is known on how aphids detect these volatiles. In this study, we identified the OR23 clade, which consists of single-copy orthologs from 8 aphid species, using phylogenetic and sequence analysis. Using a heterologous *Xenopus* expression system, ApisOR23 was found to be tuned to five plant volatiles, with *trans*-2-hexen-1-al eliciting the strongest response from ApisOR23. This chemical is one of the main GLVs released by leguminous plants, and may serve as a cue for aphids to locate their host plants to feed on. These findings suggest that conserved OR23 clade in aphid species likely plays a critical role in host plant detection.

Herbivore-induced plant volatiles (HIPVs) are important cues in the interactions between plants and herbivorous insects. Undamaged plants release lower levels of volatiles. HIPVs are widely used by predators or parasitoids to locate their preys and promote or deter herbivorous insect behavior. In this work, 54 candidate chemosensory genes were identified, including 20 OR genes from antennal transcriptome of *Megoura crassicauda*. Among these ORs, the two displaying the highest sequence conservation were functionally studied, using 11 HIPVs that have been reported to be released by aphid-infested plants. OR20 orthologs of *M. crassicauda* and the pea aphid, *Acyrthosiphon pisum*, were both specifically tuned to *cis*-jasmone, a HIPV that induces repellent behavior in various aphids. This study sheds light on the molecular mechanism of HIPV detection in aphids and provides OR20 as an olfactory target for mediating aphid behavior.

Alarm pheromones are produced by insects to alert their conspecifics of predators. When aphids are attacked by natural enemies, they produce droplets containing alarm pheromone. (*E*)-β-farnesene (EBF) being the most common and well-studied

compound. Previous study has reported that ortholog OR5 is responsible for detecting EBF and mediating repellent behavior in aphids. However, EBF is not found in the alarm pheromone compounds of some species, indicating that alarm pheromone composition diverges across different species. The evolutionary process of EBF-detecting ORs in aphids has received less attention. In this work, we annotated 533 OR genes from genomic data of 13 aphid species and identified 8 ORs that are significantly conserved among EBF-contained aphids. These OR genes were found to be conserved in the amino acid sequence and under strong purifying selection. OR43 was characterized as the second EBF receptor and involved in mediating behavioral response through two-electrode voltage-clamp (TEVC) technique, RNA interference, and behavioral experiments. Genome-wide identification of OR5 and OR43 orthologs within 17 aphid genomes showed that OR43 was exclusively found in the genomes of Aphidinae species, underlying the unique EBF perception mechanism within Aphidinae aphids.

In conclusion, this study has significantly contributed to our understanding of the molecular mechanisms of chemoreception in aphids and has identified three potential olfactory targets for regulating aphid behaviors. Additionally, the perception of EBF in aphids was demonstrated to be mediated by two ORs, both of which are involved in the avoidance behavior induced by EBF. The results also provided insights into the possible origin of EBF receptors in aphids, highlighting the unique evolutionary path of EBF detection in Aphidinae aphids.

Keywords: Aphid; chemosensory system; odorant receptor; green leaf volatiles; herbivore-induced plant volatiles; alarm pheromone.

Résumé

Les pucerons sontles ravageurs très dommagerables en agriculture , s'appuyant sur un système chimiosensoriel très sensible pour localiser les plantes hôtes, se reproduire et éviter les prédateurs. La perception des odeurs chez les pucerons implique diverses familles de gènes, notamment les récepteurs d'odeurss (OR), les récepteurs gustatifs (GR), les récepteurs ionotropiques (IR), les protéines de liaison aux odeurts (OBP) et les protéines chimiosensorielles (CSP). Pendant le processus d'olfaction, les composés odorants sont liés et transportés par les OBP et les CSP, puis libérés vers les récepteurs, tels que les OR, qui sont exprimés sur les membranes dendritiques des neurones sensoriels olfactifs. Ces récepteurs sont critiques car ils convertissent le signal chimique en un signal électrique, qui est transmis par le système nerveux central et génère finalement une série de réponses comportementales correspondantes. Bien que les caractérisations fonctionnelles se soient principalement concentrées sur la famille des OBP chez les pucerons, moins d'attention a été accordée à l'étude de la fonction des OR, ce qui a considérablement limité notre compréhension des mécanismes moléculaires de la détection d'odeurs chez les pucerons. Dans cette thèse de doctorat, nous avons identifié et caractérisé fonctionnellement trois ORs qui sont significativement conservés parmi différentes espèces de pucerons, jouant un rôle essentiel dans la détection des substances volatiles écologiquement et environnementalement importants.

Les composés volatils à notes vertes (GLV) sont des moléculess libérées par les plantes lorsqu'elles subissent des dégâts . Si es pucerons sont attirés par les GLV, les mécanismes moléculaires pour détecter ces substances volatiles restent méconnus. Dans cette étude, nous avons identifié la clade des OR23 de 8 espèces de pucerons suite à des analyses phylogénétiques de séquences. En utilisant un système d'expression hétérologue *Xenopus*, une étude fonctionnelle a révélé l'activité de ApisOR23 à cinq GLV, avec le *trans*-2-hexen-1-al suscitant la réponse la plus forte. Ces résultats suggèrent que la clade OR23 conservée chez les espèces de pucerons joue probablement un rôle critique dans la détection de la plante hôte.

Les substances volatiles de plantes induites par les phytophages (herbivores en anglais) (HIPV) sont des signaux importants dans les interactions entre les plantes et les insectes herbivores. Les plantes non endommagées émettent une faible quantité de GLV, tandis que les HIPV sont significativement induits par les dégâts causés par les ravageurss. Les HIPV sont largement utilisés par les prédateurs et les parasitoïdes pour localiser leur proie/hôte. Dans ce travail, nous avons identifié 54 gènes chimiosensoriels candidats, dont 20 OR, dans le transcriptome antennaire de *Megoura crassicauda*. Parmi ces OR, deux présentant la plus forte conservation de séquence ont été fonctionnellement étudiés à l'aide de 11 HIPV rapportés comme étant émis par des plantes infestées par des pucerons. Les orthologues OR20 de *M. crassicauda* et du puceron du pois, *Acyrthosiphon pisum*, étaient tous deux spécifiquement accordés au *cis*-jasmone, un HIPV qui induit un comportement

répulsif chez divers pucerons. Cette étude éclaire le mécanisme moléculaire de la détection des HIPV chez les pucerons et fournit un gène OR en tant que cible olfactive pour la médiation du comportement des pucerons.

Parmi les phéromones d'alarme produites par les pucerons, le (*E*)-β-farnésène (EBF) est le composé le plus courant et le mieux étudié. Cependant, l'EBF n'est pas retrouvé dans les composésphéromonaux de certaines espèces, indiquant que la composition de la phéromone d'alarme diverge entre espèces. Le processus évolutif des OR détecteurs d'EBF chez les pucerons a reçu moins d'attention. Dans ce travail, nous avons annoté 533 gènes OR à partir de données génomiques de 13 espèces de pucerons et identifié 8 OR qui sont significativement conservés parmi les pucerons contenant de l'EBF. Ces gènes OR étaient conservés dans la séquence d'acides aminés et sous forte sélection purifiante. OR43 a été caractérisé comme le deuxième récepteur de l'EBF et impliqué dans la médiation de la réponse comportementale par la technique de pince à deux électrodes, des ARN interférents et des expériences comportementales. L'identification à l'échelle du génome des gènes orthologues OR5 et OR43 dans 17 génomes de pucerons a montré que OR43 était exclusivement trouvé dans les génomes des espèces de la sous-famille Aphidinae, sous-tendant le mécanisme unique de perception de l'EBF chez les pucerons de la sous-famille Aphidinae.

En conclusion, cette étude a apporté une contribution importante à notre compréhension des mécanismes moléculaires de la chimioréception chez les pucerons, et a identifié trois cibles olfactives potentielles pour réguler les comportements des pucerons. De plus, cette étude a démontré que la perception de l'EBF chez les pucerons est médiée par deux OR, tous deux impliqués dans le comportement d'évitement induit par l'EBF. Les résultats ont également fourni des informations sur l'origine possible des récepteurs de l'EBF chez les pucerons, mettant en évidence la voie évolutive unique de la détection de l'EBF chez les pucerons de la sous-famille des Aphidinae.

Mots-clés: puceron; système chémorécepteur; récepteur olfactif; composés volatils de feuilles vertes; composés volatils de plantes induits par les herbivores; phéromone d'alarme

Acknowledgements

This Ph.D. thesis has been accomplished with the help of many people, and these much to be thankful for:

First and foremost, I would like to express my immense gratitude to my promoter, Dr. Frédéric Francis, for his unwavering support, invaluable guidance, and continuous encouragement throughout my Ph.D. journey. He is one of the kindest individuals I have ever encountered, consistently supporting me in both my research and daily life.

I would also like to extend my deepest appreciation to my co-promoter, Dr. Guirong Wang, for providing me the opportunity to further my research as a joint Ph.D. student between CAAS and ULiege. From the very first days until now these lines are being written, he has granted me the freedom and autonomy to explore my interests, and his expertise and mentorship have been instrumental in shaping my research and professional growth.

My sincere appreciation goes to Dr. Bing Wang, for her fruitful discussions, meticulous guidance on every aspect of my research, and support beyond my academic pursuits.

I would like to thank my colleagues at the Department of Functional and Evolutionary Entomology, Gembloux Agro-Bio Tech, especially Sandra Torsin and Noël Grégoire, whose assistance have made my life in Belgium much easier and enjoyable.

I cannot express enough gratitude to my friends and colleagues at CAAS. Since 2014, I have worked alongside over 100 brilliant and helpful individuals, and I cannot have achieved my degree without their companionship and encouragement. I want to thank Song Cao for his assistance and shared happiness. I thank Mengbo Guo, Erzong Zhang, Kang He, Hetan Chang, Lixiao Du, Yipeng Liu, Ruibin Zhang, Jingjing Song, for their support and guidance, and Xiaxuan Zhang, Shuai Liu, Baiwei Ma, Ke Tian, Bin Li, Lulu Yang, Dongdong Sun, Shang Lei, Wenbiao Liu, Ying Tian, Jinan Wu, Denghai Yang for the help they provided in my life and work and the moments we had.

I am grateful to the Chinese Scholarship Council for funding my research in Belgium. My time here has offered me the opportunity to learn and experience the differences between China and Belgium in the areas of culture and scientific

research. Belgium is a beautiful country, and its people are incredibly kind and warm-hearted, always willing to take the time to enjoy life. Although my stay in Belgium was brief, it has left an indelible mark on my heart.

I would like to thank my parents for their support and encouragement. Finally, I would like to extend my deepest gratitude to my loving and supportive wife Dr. Yao Liu, for her unwavering faith in my abilities, constant encouragement. Her love and companionship have made the difficult times more bearable and the joyful moments even more memorable. We first met during the PhD application stage, and now, as four years have passed and our doctoral journeys are coming to an end, we are also about to celebrate our wedding. It is she who has transformed these four years, which should have been filled with stress, into a colorful and lovely experience.

Tianyu Huang
March, 2023 in Beijing, China

Contents

List of figures

List of tables

List of abbreviations

ABC : ATP-binding cassette
ANOVA : One-way Analysis of Variance
BB : Böhm Bristle
BUSCO : Benchmarking Universal Single-Copy Orthologs
CSP : Chemosensory Protein
CP : Carboxypeptidase
DMSO : dimethyl sulphoxide
EBF : (*E*)-β-farnesene
FPKM : Fragments per kilobase per million fragment
FPKM : Geranyl acetate
GC/MS : Gas chromatography/mass spectrometry
GLV : Green Leaf Volatile
GPCR : G Protein-Coupled Receptor
GR : Gustatory Receptor
HIPV : Herbivore-induced Plant Volatile
IR : Ionotropic Receptor
LH : Lateral Horn
LN : Local Neuron
MB : Mushroom Body
MYA : Million years ago
NCBI : National Center for Biotechnology Information
NGDC : National Genomic Data Center
OBP : Odorant Binding Protein
ODE : Odor Degrading Enzyme
OR : Olfactory Receptor
ORco : Odorant Receptor Co-receptor
ORF : Open Reading Frame
ORN : Olfactory Receptor Neuron
P450 : cytochrome P450
PR : Pheromone Receptor
PN : Projection Neuron

RNAi	:	RNA interference
ST	:	Sensilla Trichodea
SB	:	Sensilla Basiconica
SCo	:	Sensilla Coeloconica
SC	:	Sensilla Chaetica
SCl	:	Sensilla Clavate
SST	:	Sensilla Styloconica
SRA	:	Sequence Read Archive
SP	:	Sensilla Placodea
SAu	:	Sensilla Auricillica
SNMP	:	Sensory Neuron Membrane Protein
TE	:	Transposable element
TEVC	:	Two-electrode Voltage Clamp
TMD	:	Transmembrane Domain
UGT	:	UDP-glucuronosyltransferase
VOC	:	Volatile organic compound

Chapter I

General introduction

1. Olfactory recognition mechanism of insects: research progress

1.1. Prelude

Insects are the most abundant group of animals in the world, and they have developed a highly sensitive and extremely specific olfactory system through long-term natural selection and evolution. This system enables insects to recognize chemical cues related to their own behavioral activities in highly complex natural environments, allowing them to carry out their normal life activities.

1.2. Olfactory recognition mechanism of insects

Detecting the chemical cues in the environment is crucial for the survival of almost all living organisms. Insects, in particular, rely on their powerful olfactory system to locate host plants and preys, find mates and food sources, select suitable habitats and oviposition sites and avoid predators (Bruce, 2015; Elgar et al., 2018; Gadenne, Barrozo, & Anton, 2016; Haverkamp, Hansson, & Knaden, 2018; Renou & Anton, 2020). Insects encounter chemical signals emitted by conspecifics, including sex pheromones, alarm pheromones, or aggregation pheromones (Basu, Clark, Fu, Lee, & Crowder, 2021; X. Guo et al., 2020), and also use compounds released by a diversity of organisms, such as predators and host plants (Fleischer, Pregitzer, Breer, & Krieger, 2018; Guo & Wang, 2019; Wen et al., 2019). The antennae are the crucial sensory organs for insects that primarily identify complex chemical signals in the environment through hair-like or spiny sensory structures, which are specialized from the antenna's cuticular cells. Different insect species have various types of sensory structures on the antennae based on their size, shape, and structure. There are considerable differences in the types, distribution, and number of sensory structures among insect species, and even among sexes of same species. In Diptera, the main types of sensory structures are sensilla trichodea (ST), sensilla basiconica (SB), sensilla coeloconica (SCo) (Jia, Sun, Luo, & Wu, 2019), sensilla chaetica (SC) (Hore, Saha, & Banerjee, 2018), sensilla clavate (SCl) (Smallegange, Kelling, & Den Otter, 2008), sensilla styloconica (SSt) (Hore et al., 2017), sensilla placodea (SP), sensilla auricillica (SAu) (D. Zhang, Li, Liu, Wang, & Pape, 2016), Böhm bristles (BB), microtrichiae (Mt), sensory pits, sensory sacculi, etc. (Gao et al., 2020; Z. Liu, Hu, Guo, Liang, & Cheng, 2021; Oh, Jeong, Kim, & Park, 2019). Typical feature of olfactory sensory structures is the numerous nanoscale pores on the cuticle, which connect the internal cavity of the sensor with the external environment allowing odor molecules to diffuse into the sensor through these pores (Steinbrecht, 1997) consisting in three accessory cells (sheath, membrane and hair cells) and one to several olfactory receptor neurons (ORNs) (Figure 1). The dendrite of a bipolar ORN extends to the top of the sensor, and odor receptors or ionotropic receptors expressed on the dendritic membrane can receive complex chemical signals from outside. The other end of the neuron extends to the base of the sensor, forming the axon, which can transmit the electrical signals generated by external stimuli to the central nervous system, directing insects to exhibit corresponding behaviors to different odors (Bates et al., 2020; Schmidt &

Benton, 2020).

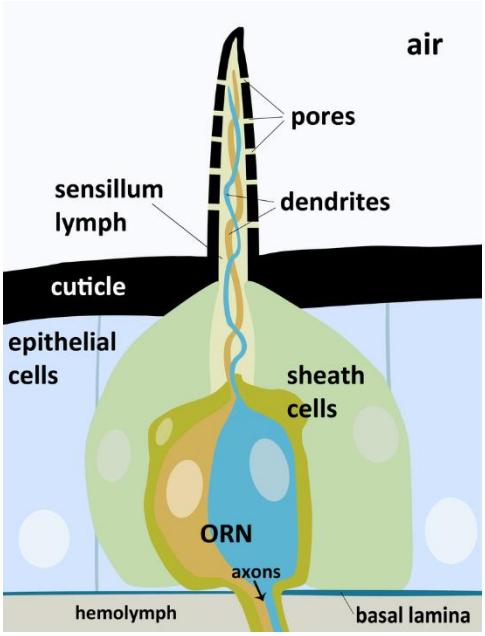


Figure 1. General organization of an insect olfactory sensillum.

The recognition of odor molecules by insects is a complex process, as illustrated in Figure 2. Odor molecules first enter the sensor through pores in the cuticle and must pass through the hydrophilic lymph of the sensor to reach the dendrites of ORNs. As most odor molecules are highly volatile and relatively hydrophobic compounds, they need to bind to odorant binding proteins (OBPs) to pass through the lymph. These OBPs act not only as functional carriers, but also play a role in enhancing the solubility of odors and in the initial selection process of olfactory information (P. Pelosi, I. Iovinella, J. Zhu, G. Wang, & F. R. Dani, 2018; R. Zhang et al., 2017; Zhu et al., 2019). When the odor-OBP complex reaches the ORs on the dendritic membrane surface, it is unclear whether the complex itself binds to the receptor or if the complex dissociates near the receptor before the odor alone stimulates the receptor. At the peripheral olfactory molecular level, in addition to ORs and OBPs, various proteins such as ionotropic receptors (IRs) (Benton, Vannice, Gomez-Diaz, & Vosshall, 2009; Koh et al., 2014; Rytz, Croset, & Benton, 2013) and sensory neuron membrane proteins (SNMPs) (Cassau & Krieger, 2021; S. Liu et al., 2020) are involved in the recognition of odor molecules by the insect peripheral system. In the olfactory process, the deactivation of chemical signals also plays an important role in avoiding continuous stimulation of neurons, restoring the sensitivity of neurons, and preparing for the reception of new signal stimulation. This process is accomplished by various soluble extracellular and intramembrane binding enzymes, cytoplasmic enzymes, and odor degrading

enzymes (ODEs).

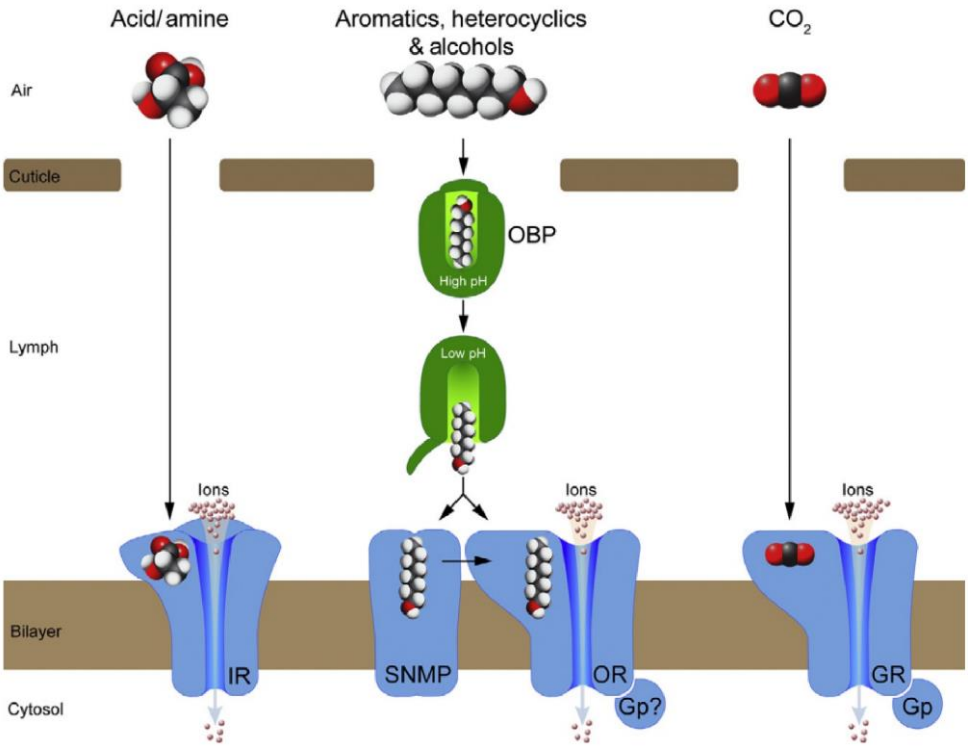
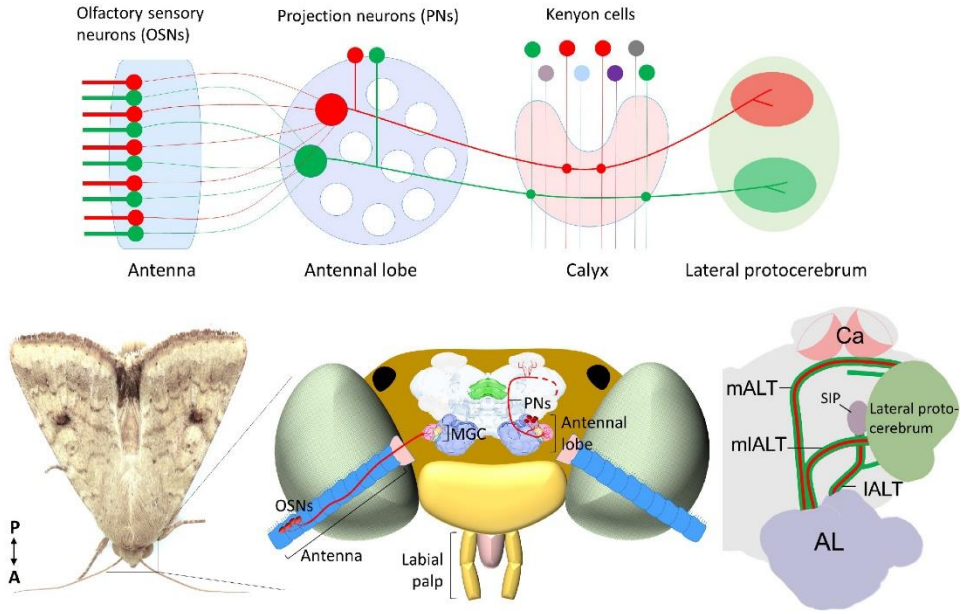


Figure 2. Model pattern of insect peripheral olfactory system for detecting odor molecules (Sparks, Bohbot, & Dickens, 2015).

After the chemical signals from the external environment are converted to electrical signals in the peripheral olfactory system, these electrical signals are transmitted through the axons of olfactory receptor neurons to the primary olfactory center, ----the antennal lobe. The latter contains numerous olfactory glomeruli and local neurons (LNs) connect to the olfactory glomeruli to perform preliminary integration and processing of electrical signals from different sources. The projection neurons (PNs) then project through different parallel bundles to the higher centers, the mushroom bodies (MB) and the lateral horns (LH). The MB are an important center for associative learning and memory formation, not only receiving olfactory but also primary visual and gustatory inputs in networks. Numerous intrinsic Kenyon cells in MB are responsible for integrating and transmitting input information from the mushroom body crown to the mushroom body lobes. A small number of mushroom body output neurons connect to other integration and motor centers in the forebrain. The lateral horns of the forebrain receive direct input from the antennal lobe projection neurons and are closely

connected to the mushroom bodies as well as other sensory centers in the brain. The output neurons of the lateral horns connect to the motor centers in the forebrain. The higher olfactory center integrates and processes these olfactory inputs and transmits them directly or indirectly to downstream neurons, ultimately guiding the insect to make corresponding behavioral responses. Schematic of insect olfactory



system is shown in Figure 3.

Figure 3. Schematic representation of the neurons in the insect olfactory system (Kymre et al., 2021).

1.3. Research progress of insect odorant receptors

1.3.1. Identification of insect odorant receptors

The first OR coding gene was identified in the brown rat, *Rattus norvegicus*. Buck and Axel (1991) discovered that odorants bind to G protein-coupled receptors (GPCRs), which activate G protein-mediated second messenger cascades and mediate the transduction of chemical sensory signals in vertebrate ORNs. Because GPCRs have seven transmembrane domains, Buck and Axel screened for genes encoding 7-transmembrane domain proteins and identified a large number of GPCRs that were specifically expressed in olfactory epithelial cells. Subsequently, based on homology cloning methods, OR gene families were identified in other vertebrate species including humans (Ben-Arie et al., 1994), fish (Ngai et al., 1993), and birds (Nef, Allaman, Fiumelli, De Castro, & Nef, 1996). However, identifying insect odorant receptors has been a challenging study and attempts using sequence homology methods based on vertebrate or previously reported *Caenorhabditis elegans* olfactory receptor sequences have failed (Troemel, Chou, Dwyer, Colbert,

& Bargmann, 1995).

In 1999, with the successful sequencing of the whole genome of the fruit fly *Drosophila melanogaster*, the first insect receptor gene was identified (Adams et al., 2000). The entire OR library of the fruit fly contains 60 functional OR genes encoding 62 different OR proteins (Robertson, Warr, & Carlson, 2003). Subsequently, the genome sequence of the mosquito *Anopheles gambiae* was deciphered and by using the fruit fly sequence for homology search, 79 ORs were identified in mosquito (A. N. Fox, Pitts, Robertson, Carlson, & Zwiebel, 2001; Hill et al., 2002). By comparing the OR families of fruit flies and mosquitoes, even if both belong to the Diptera order, it was found that their OR sequences have significant differences, indicating that the OR subfamilies are not conserved in Diptera. In the past 20 years, with the rapid development of sequencing technology and bioinformatics tools, combined with the significant reduction in sequencing costs, it has become possible to search for OR gene family in genomes of many insect species (Table 1). In addition to genome, transcriptome of insect olfactory tissues has become a suitable resource for identifying ORs. Thousands of candidate insect OR sequences have been identified through methods such as BLAST searches using known OR sequences as queries. The number of OR genes in different insect species varies greatly, with some species having only a few OR genes and others having hundreds. The size of the insect OR gene family may be related to the complexity of the chemical environment in which the insect lives (Robertson, 2019). Comparative phylogenetic analysis of insect OR libraries can provide valuable information on the evolution and expansion of the OR family in the insect lineage.

Table 1. Numbers of intact OR genes, proteins and pseudogenes in select insects

Species	Number of intact OR genes	Number of OR proteins	Number of OR pseudogenes
Drosophila melanogaster	60	60	2
Drosophila sechellia	55	51	6
Musca domestica	78	72	7
Glossina morsitans	46	42	4
Tribolium castaneum	261	222	39
Apis mellifera	165	155	10
Pogonomyrmex barbatus	345	290	55
Nasonia vitripennis	225	149	76

Ceratosolen solmsi	46	44	2
Pediculus humanus	10	10	0
Zootermopsis nevadensis	63	57	6
Blattella germanica	123	105	29
Calopteryx splendens	5	5	0

1.3.2. Structural characteristics of insect odor receptors

In mammals, odor receptors belong to the G protein-coupled receptor superfamily, with typical 7 transmembrane domains, N-terminus extracellular and C-terminus intracellular (Wistrand, Käll, & Sonnhammer, 2006). It has long been thought that insect odor receptors are similar to those in mammals, with 7 transmembrane domains and the same topology. However, in 2006, Benton and colleagues found that the N-terminus of the insect odor receptor protein is intracellular, and the C-terminus is extracellular, which is opposite to the typical G protein-coupled receptor structure (Benton, Sachse, Michnick, & Vosshall, 2006). Subsequently, Lundin et al.(2007) further confirmed this view through glycosylation mapping of the OR83b topology. Currently, it is generally believed that insect ORs are derived from a greatly expanded system evolutionary lineage of gustatory receptors (GRs), different from GPCRs. Therefore, insect ORs do not transduce chemical signals through G proteins, but function as non-selective cation channels by forming heteromeric complexes with OR co-receptors (Sato et al., 2008; Wicher et al., 2008). Research has found that insect odor receptors and mammalian GPCRs lack sequence similarity, likely due to the differentiation of these two kinds of receptors in the process of evolution, forming two different olfactory systems to recognize different types of odorants (H. G. Song, Young Kwon, Soo Han, Bae, & Moon, 2008).

1.3.3. Functional studies on insect odor receptors

The odor receptors of insects can be divided into two types: one is the odorant receptor co-receptor (ORco), generally each insect only has one ORco and is highly conserved across different species of insects (Soffan, Subandiyah, Makino, Watanabe, & Horiike, 2018); the other one is the traditional odor receptor ORx, which varies greatly in number among different species of insects and has low homology. In Lepidoptera, they can be further divided into general odor receptors (ORs) and pheromone receptors (PRs) based on the type of compounds they recognize (Leal, 2013).

1.3.3.1. ORco

Identification of OR genes in *D. melanogaster* and *A. gambiae* resulted a pair of orthologous OR genes among these two insects, named *DmelOR83b* and *AgamOR7*, which encode a unique ORco protein (Larsson et al., 2004; Vosshall & Hansson, 2011). ORco homologous genes have been discovered in many insect orders, such

as Lepidoptera, Coleoptera, Hymenoptera, Hemiptera, Orthoptera, and Phasmatodea (W. D. Jones, Nguyen, Kloss, Lee, & Vosshall, 2005; Krieger, Klink, Mohl, Raming, & Breer, 2003; Missbach et al., 2014; Pitts, Fox, & Zwiebel, 2004; C. Smadja, Shi, Butlin, & Robertson, 2009), and are highly conserved within and between insect orders (with sequence similarity up to ~95%) (Olafson, 2013; Soffan et al., 2018). ORco is expressed in most OR-expressing olfactory receptor neurons (ORNs) in insects, but not in ORNs expressing other types of chemosensory receptor proteins, such as IRs and GRs (Montagné et al., 2012; Montagné, de Fouchier, Newcomb, & Jacquin-Joly, 2015). Therefore, ORco is considered as a marker for OR-expressing ORNs in insects, and these ORNs usually co-express one or several different types of ORs (called typical or classic ORs) together with ORco (Martin, Boto, Gomez-Diaz, & Alcorta, 2013).

The functional research results indicated that ORs and ORco appear to interact through intracellular structural domains to form hetero-oligomeric ligand-gated ion channels and exert their function (Butterwick et al., 2018; Sato et al., 2008; Wicher et al., 2008). In OR/ORco heteromers, the typical OR binds to odorants and determines ligand specificity, while ORco apparently does not participate in ligand binding. However, the loss of ORco leads to impaired olfactory function in insects, and their response to odors is significantly reduced (Paulo et al., 2021; H. Sun, Liu, Ye, Baker, & Zwiebel, 2020). Studies on fruit flies have shown that the olfactory defects caused by the loss of ORco can be rescued by expressing the ORco gene from other insect species (moths or mosquitoes), indicating the important functional conservation and relevance of ORco (W. D. Jones et al., 2005). Additionally, there are studies on the co-expression of ORco in exogenous systems.

1.3.3.2. ORx

Currently, there are three main methods for studying the function of traditional odor receptors. The first method is to use *Xenopus* oocyte expression system combined with two-electrode voltage clamp technology to record the electrophysiological responses of ORs (Hou et al., 2020; Xiao et al., 2020), and PRs (Song Cao, Huang, Shen, Liu, & Wang, 2020; Y. Liu, Liu, Jiang, & Wang, 2018) to different odors, which is the most mature method currently used to study odor receptor function. The principle is to microinject OR and ORco corresponding cRNA into the frog oocytes and then use two-electrode voltage clamp technology to measure the response of the oocytes expressing OR and ORco genes to odors in aqueous solution after a few days of cultivation. In general, hydrophobic odors are usually dissolved in DMSO and added to the liquid medium, and in rare cases, OBPs are used for dissolving (Chang et al., 2015; M. Sun et al., 2013; B. Wang, Cao, Liu, & Wang, 2020). In studies that used odor/OBP solutions to stimulate OR, odor-binding proteins can increase the sensitivity of OR to odor molecules and the specificity of the response (P. Pelosi et al., 2018). However, this system also has some limitations, such as only being able to inject one oocyte at a time with a single OR's cRNA. To overcome this problem, sophisticated semi-automatic or fully automatic oocyte injection systems have been developed (Papke & Stokes, 2010). The application of these advanced technologies helped in high-throughput functional screening of insect ORs.

The second method is the transgenic *D. melanogaster* heterologous expression system. Currently, two types of "empty" *Drosophila* heterologous OR expression systems have been established, which can provide different environments. One mutant strain is the deletion of the endogenous receptor gene Or22a in the cone-shaped sensors of the *Drosophila* antenna (Dobritsa, van der Goes van Naters, Warr, Steinbrecht, & Carlson, 2003). Candidate ORs are expressed in the "empty" neurons through the GAL4/UAS system, which uses the promoter of the endogenous receptor gene Or22a to drive GAL4 expression and then activate the expression of UAS-ORs. The *Drosophila* "empty" neuron system is combined with single-sensor recording to determine the response characteristics of candidate ORs. This method is not only suitable for the study of odor receptor function in *D. melanogaster* (Hallem & Carlson, 2006; Hallem, Ho, & Carlson, 2004; Kreher, Kwon, & Carlson, 2005), but also for other flies (Chahda et al., 2019), mosquitoes (Carey, Wang, Su, Zwiebel, & Carlson, 2010; Speth et al., 2021), and moths (de Fouchier et al., 2017) for the identification of general OR function. The other mutant strain is based on the heterologous expression system of the *Drosophila* T1 sensilla Or67d deletion, which also uses the GAL4/UAS method to express exogenous OR genes and uses single-sensor recording to analyze receptor function, suitable for the study of pheromone receptor function (Syed, Kopp, Kimbrell, & Leal, 2010). The principle is that a neuron in the T1 sensillum expresses the Or67d receptor to detect male pheromones cis-vaccenyl acetate, and contains the SNMP1 receptor, which provides an ORN environment that can detect pheromones (Kurtovic, Widmer, & Dickson, 2007). Studies have shown that the T1 heterologous expression system can express moth PRs well (Bastin-Héline et al., 2019; Cattaneo et al., 2017), and it can also be used to identify the function of OR of other insects, including *Locusta migratoria* (You, Smith, Lv, & Zhang, 2016) and pea aphid (R. Zhang et al., 2017).

The third type of receptor functional research method is to use expression vectors to transfect cell lines to drive the transient expression of OR and ORco. Currently, human embryonic kidney cell line 293 (HEK293 cells) (Miazzi et al., 2019), mammalian HeLa cells (Sato et al., 2008), *Drosophila Schneider* 2 cells (S2 cells, derived from embryos) (Lundin et al., 2007; Smart et al., 2008), as well as cell lines from *Spodoptera frugiperda* (Sf9 cells) (Murugathas et al., 2019), *Plusia ni* (High five cells) (German, van der Poel, Carraher, Kralicek, & Newcomb, 2013), and *Bombyx mori* (Bm5 cell line) ovarian cells (Tsitoura et al., 2010) have been successfully used to identify the function of insect ORs. Generally, organic solvents (such as dimethyl sulfoxide, DMSO, and methanol) dissolved in or odorants combined with odorant-binding proteins are used to stimulate cells expressing ORs for functional identification. Odor-mediated cell responses can be analyzed by calcium imaging or patch clamp techniques (Miazzi et al., 2019).

In conclusion, the cell line and *Xenopus* oocyte expression system are more suitable for *in vitro* OR high-throughput functional screening, while the "empty neuron system" of *Drosophila* is more accurate for in vivo functional analysis (Bing Wang, Liu, He, & Wang, 2016). For some insect ORs, the *Drosophila* empty neuron heterologous expression system may be the preferred method for deciphering their

functions, because the insect *in vivo* expression system can provide appropriate physiological environment and correct upstream processing mechanisms for OR responses.

2. Aphids: important pests on crops

2.1. Basic knowledge on aphids

Aphids are small, soft-bodied insects that belong to the order Hemiptera. Among nearly 5000 species that are widely distributed across the world, many of them are significant pests of many crops and ornamental plants, causing damage by feeding on plant sap and transmitting plant viruses (Pickett, Rasmussen, Woodcock, Matthes, & Napier, 2003). Despite their small size, aphids have unique physiological characteristics that allow them to thrive and adapt to a variety of environments.

Aphids have a distinctive, pear-shaped body that ranges from 1-10 millimeters in length and is composed of three segments: head, thorax and abdomen. The head contains two compound eyes, two simple eyes (ocelli), and a pair of antennae. The thorax has three pairs of legs and two pairs of wings. The abdomen is soft and rounded, and contains the digestive, excretory, and reproductive systems. Aphids have long, slender, needle-like mouthparts called stylets, which they use to penetrate plant tissues and feed on sap. The stylets are composed of two parts: the food canal, which carries sap to the aphid's mouth, and the salivary canal, which secretes saliva containing enzymes that digest plant tissues and help the aphid to obtain nutrients it needs. Cornicles are paired external appendages unique to the family Aphididae which located on the fifth or sixth abdominal segment of aphid. This organ release a diversity of volatiles and nonvolatile compounds which involved in multiple ecological functions (Michaud, 2022). The basic anatomy of aphid is showed in Figure 1.

Aphids are capable of reproducing both sexually and asexually, depending on the species and environmental conditions. Asexual reproduction is the most common form of reproduction in aphids, and it occurs through a process known as parthenogenesis. Female aphids lay young larvae without fertilization that are clones of the mother. Sexual reproduction occurs when environmental conditions become unfavorable, and male aphids are produced. The male and female aphid mate, and the latter lays fertilized eggs that overwinter, giving rise to new generations in the spring.

Some species of aphids are capable of long-distance migration, which allows them to escape from unfavorable environmental conditions, such as drought or cold temperatures. Aphids are able to sense changes in the environment, such as variations in temperature or light, and they respond by moving to new hosts in search of more favorable conditions.

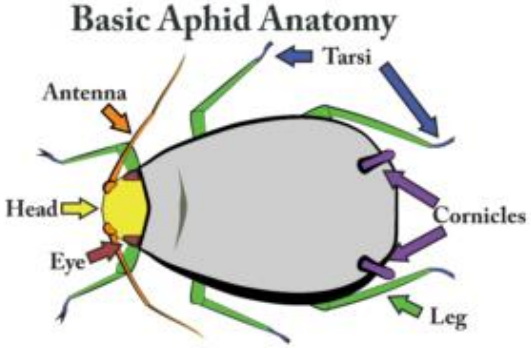


Figure 4. Basic anatomy of aphid. Head is indicated in yellow; Antennae are indicated in orange; Eyes are indicated in red; Legs are indicated in green; Cornicles are indicated in purple; The main body is indicated in grey. (Painted by Jason Thomas)

2.2. Economic impact of aphids on crops

Aphids are one of the most destructive pests in agriculture, causing significant economic impact on crops worldwide. They cause damage on their host plants in diverse ways. Firstly, aphids feed on plant sap using their stylet, causing damage to leaves, stems, and flowers. This feeding can result stunted growth, yellowing of leaves and even death of plant. Secondly, during the feeding phase, they inject various salivary proteins to the plant tissues (Boulain et al., 2019). Watery saliva of aphid is a complex mixture of enzymes and other components capable of modulating host cell's function also the structure for assuring the nutrient availability. Thirdly, aphids transmit numerous plant viruses: around 50% of insect-borne viruses (275 out of 600) are transmitted by aphids (Gray, 2008). Finally, aphids excrete a sticky substance called honeydew, which provides a habitat for mold and other secondary pests.

Yield losses due to aphid damage can be substantial. Depending on the crop and the severity of the infestation, losses can range from a few percent to more than 50%. For example, in soybean crops, the predicted maximum possible yield loss was 75% for soybean aphid infestations starting at the five node stage of soybean (Catangui, Beckendorf, & Riedell, 2009). Similarly, in potato crops, aphid transmit *Potato Virus Y* which can cause the yield reduction ranging from 40% to 70% (Karasev & Gray, 2013). The sugarcane aphid, *Melanaphis sacchari*, has reported to be confirmed on sorghum in 4 states and 38 counties in the United States, resulted in yield loss ranging from 10% to more than 50% (Bowling et al., 2016).

2.3. Pest management strategies for aphid control

2.3.1. Chemical control

Chemical controls are the most widely used method of controlling aphids. The first organochlorinated insecticide, DDT, was discovered in 1939. After Second World War, chemical control of aphids had rapidly progressed due to the production of other organochlorinated compounds, organophosphates, carbamates and pyrethroids from 1940s to 1980s respectively. The latter are still by far the most used for spraying on plants (Dedryver, Le Ralec, & Fabre, 2010). Another type of insecticide that is very efficient for aphid control is neonicotinoids as they can be transmitted through plant by xylem and phloem vessels, consequently becoming a good method for controlling phloem feeding insects.

Although chemical control is of the most efficient method on aphid management, there are increasing limitations. Firstly, it can cause adverse effects on both natural environment and health of non-target species including human. Moreover, the overuse of chemical insecticides can lead to the development of insecticide-resistant populations of aphids, making the pest control more difficult in the future. Indeed, many aphid species have evolved significant resistance of insecticides. For example, the peach potato aphid, *Myzus persicae*, a generalist aphid with a host range of more than 400 plant species (Blackman & Eastop, 2000). Due to the particularly large host range and intensively use of insecticides over many years, *M. persicae* has developed multiple types of resistance, which greatly increased the difficulty of the control of this pest (Bass et al., 2014).

2.3.2 Biological controls

Biological controls offer a more sustainable alternative to chemical insecticides, and they can be used in the place that not suitable for spraying with broad-spectrum pesticides (e.g., organophosphates, carbamates, and pyrethroids) that kill natural enemy species as well as pests. Biological controls usually include the release of natural predators, such as parasitic wasps and ladybugs. Aphid parasitoids, belonging to the Hymenoptera families of Braconidae and Aphelinidae as well as a few species of Diptera family Cecidomyiidae, play a crucial role in regulating aphid populations and are commonly used in biological control programs in both greenhouses and field environments. The entire lifecycle of the aphid parasitoid is spent inside its host aphid, with the female depositing her egg directly within the host, the aphid is killed during the last larval instar (O'Donnell, 1987). Ladybugs and their larvae are the major predator of aphids (Seo & Youn, 2002). So far, various species have been used for biological control of aphids, including *Harmonia axyridis*, *Propylaea japonica*, *Scymnus babai* and so on.

3. Advance in aphid chemical ecology

3.1. Prelude

The resistance of major agricultural pests to insecticides and the consequent environmental problems have worsened due to the widespread use of chemical

pesticides. It is of great importance to develop environment-friendly methods to control aphid populations. Aphid behavior can be influenced by chemical cues from the natural environment, such as attracting them to host plant odors (Hopkins, Cameron, & Butlin, 2017) or repelling them with aphid pheromones (Beale et al., 2006).

Indeed, some VOCs have potential value in the aphid pest management. For example, the main species found on wheat in Belgium, *Metopolophum dirhodum* and *Sitobion avenae*, can be attract by (*Z*)-3-Hexenol, suggesting the potential role as an effective infochemical for managing aphids by luring them away from crop fields. Moreover, the application of (*E*)-β-farnesene (EBF) or a garlic extract (GE) resulted in a notable reduction in the population of these two wheat aphids (Zhou et al., 2016). Slow-release alginate bead of EBF mixed with MeSA has significantly decreased the abundance of aphids with or without wings (Liu et al., 2021).

These features can be utilized as efficient biological control methods for aphid management. Therefore, understanding of how aphids perceive odorants is essential for developing novel aphid attractants and behavior regulators.

3.2. Structure of aphid antennae

Insects heavily rely on their antennae to detect semiochemicals and environmental odors. In aphids, the antennal olfactory sensilla have been categorized into three distinct types based on their external morphology: primary rhinaria, second rhinaria, and trichoid sensilla (Shambaugh, Frazier, Castell, & Coons, 1978; L. Song, Wang, Liu, Sun, & Ban, 2020) (Figure 4). Primary rhinaria comprise several sensillum types and are located on the 5th and 6th segments of the antenna. Also, the second rhinaria, sensilla placoidea, are situated between the 3rd and 5th segments. Two distinct types of trichoid sensilla have been identified based on their morphology. Type I hair is present throughout the entire length of the antenna up to the 6th segment primary rhinarium, while type II hair is found on the processus terminalis and at the peak of antennae.

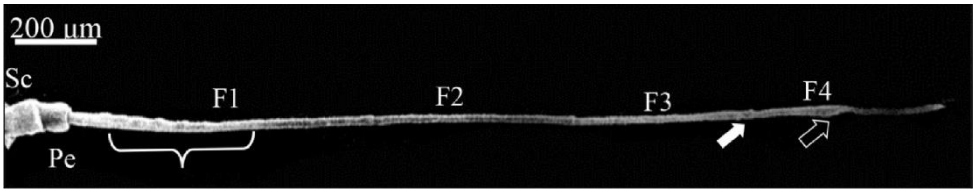


Figure 4. Aphid antenna is composed of six segments, including one scape (Sc), one pedicel (Pe) and four flagella (F1-F4). The white arrow indicates the primary rhinaria located on the fifth segments, and the black arrow indicates those located on the sixth segments; The brace indicates the secondary rhinaria located on the third segement (Song et al., 2020).

3.3. Interaction between aphid and plant volatiles

Plant volatiles play an important role as chemical cues for aphids to locate their host. Consequently, chemicals released by suitable host plants often attract aphids.

For example, bird cherry-oat aphid, *Rhopalosiphum padi*, were attracted to volatiles released by their host plants, wheat (*Triticum aestivum* L., Poaceae) and oats (*Avena sativa* L.) (Quiroz & Niemeyer, 1998). The black bean aphid, *Aphis fabae*, was also found to be attracted to volatiles emitted from its host plant *Vicia faba* L. (Webster et al., 2008). Interestingly, aphids may have different preferences for different strains of the same host plant. Schröder and colleagues (2015) reported that *R. padi* was attracted to odors emitted from maize cultivar 6Q-121, but did not respond to odorants from the other two maize cultivars. This suggested that the olfactory system of aphid is highly sensitive, allowing it to distinguish slight differences between the volatile repertoires of different cultivars. Moreover, the attraction of plants to aphid can be alerted due to plant virus infection. Indeed, the behavioral responses of *M. persicae* to volatiles released by cucumber plants infected and uninfected with cucumber mosaic virus (CMV) were different. The attractiveness of infected plants to aphids has significantly increased due to increased volatile emissions (Mauck, De Moraes, & Mescher, 2010).

Herbivore-induced plant volatiles (HIPVs) are chemical compounds that plants increasingly produced upon herbivory. These compounds are important cues for predator and parasitoids to locate prey or hosts (Vet & Dicke, 1992). Many studies have indicated that HIPVs attract nature enemies of aphids and consequently reduce aphid population. For instance, Mallinger and colleagues (2011) demonstrated that HIPV methyl salicylate significantly increased the number of various natural enemies, resulting in a significant decrease in the abundance of soybean aphid, *Aphis glycines* in the treated plot. In cereal fields, the number of natural enemies and that of their prey (a diversity of common aphid species on cereal crops) are positively correlated, indicating the crucial role of predators in controlling aphid populations on cereal plants (Ramsden, Menendez, Leather, & Wäckers, 2017). Additionally, HIPVs released by cereal plants can also affect the behavior of aphids. For example, significantly fewer *Rhopalosiphum maidis* have been observed in maize fields treated with HIPVs, and a repellent effect of *R. maidis* to HIPVs has been observed (Bernasconi, Turlings, Ambrosetti, Bassetti, & Dorn, 1998).

3.4. Aphid alarm pheromone

Aphids typically live in large groups on plants, communicating with each other through various chemical signals, including pheromones. One of the most intriguing types of pheromones used by aphids is the alarm pheromone in response to danger, such as the presence of predators or parasitoids. The alarm pheromone is released by cornicle (Figure 1), however, the exact location of alarm pheromone synthesis is still unknown. The release of the alarm pheromone by one aphid can quickly spread to other aphids in the group. When aphids detect the alarm pheromone, they respond by moving away from the source of danger or changing their behavior in some other way.

Bowers and colleagues (1972) were the first to characterized EBF as the primary component of the alarm pheromone in various aphid species that cause significant damage to crops. EBF has been identified as the main or even the only component of alarm pheromones in many aphids, including *M. persicae*, *A. pisum* (Dewhirst et

al., 2008). In a noteworthy study, Francis and colleagues identified and analyzed the released alarm pheromones from 23 aphid species, with EBF identified as the primary compound in 16 of these species, providing important information about aphid alarm pheromones (Francis, Vandermoten, Verheggen, Lognay, & Haubruge, 2005).

Although EBF is the primary component of the alarm pheromone in many aphid species, other chemicals have been identified as the alarm pheromone compound in some aphids. For example, germacrene A, was identified in alfalfa aphids, and showed a repellent effect on aphids, but only in the genus *Therioaphis* (Bowers et al., 1977). Monoterpenes, including α-pinene, β-pinene, and β-limonene, were also characterized as aphid alarm pheromones in some species, such as *Megoura viciae* (Pickett & Griffiths, 1980). Further analysis of terpenoids in *M. viciae* showed that a mixture of four major components (-)-β-pinene (49.74%), EBF (32.64%), (-)-α-pinene (9.42%) and (+)-limonene (5.24%) were identified, and the result emphasized the important role of the mixture of these chemicals as the repellent (Song, Qin, Yin, & Li, 2021). These studies suggest that the alarm pheromone compounds may differ among different aphids.

Chapter II

Objectives and thesis structure

This thesis aims to improve our understanding of how aphids detect important volatiles and the molecular mechanisms of aphid olfaction. This study also focuses on investigating the function of odorant receptors that are specific to EBF, the most well-studied alarm pheromone compound within aphids, and shedding light on the evolutionary process of these EBF receptors.

Chapter 3 aimed to identify odorant receptors (ORs) responsible for detecting green leaf volatiles (GLVs) in aphids, as GLVs play a crucial role in the interactions between plants and aphids. The OR gene sets of three aphids (*A. pisum*, *A. glycines* and *A. gosspyii*) were analyzed to identify conserved OR clades using phylogenetic analysis. Motif analyses and tissue expression patterns were also examined, and the ortholog OR of the pea aphid *A. pisum* was functionally characterized using a stimulation panel of 57 plant-released volatiles. The aim was to determine whether conserved olfaction mechanisms exist among different aphids for the detection of GLVs. The findings from this chapter will provide a better understanding of the molecular mechanisms of aphid olfaction and the role of GLVs in plant-aphid interactions.

In chapter 4, the focus was on understanding the olfaction mechanism of aphids in detecting herbivore-induced plant volatiles (HIPVs). *Megoura crassicauda*, a specialist aphid on legume plants, was selected as the target species. The antennal transcriptome of *M. crassicauda* was sequenced and annotated for chemosensory-related gene families. The aim was to identify odorant receptors (ORs) that tune to HIPVs and have a better understanding of the olfaction mechanism of aphids. The conserved ORs were functionally studied in vitro using 11 HIPVs that are reported to be released by aphid-infested plants.

Performing a comparative analysis between the OR gene families of different aphids often requires genome assemblies as essential resources. While many aphid genomes are publicly available, they mainly focus on Aphidinae species, lacking genomic data of other aphid subfamilies. This limitation may constrain our efforts to investigate the evolutionary process of the aphid OR gene family. Therefore, **in chapter 5, we presented the first chromosome-level genome assembly of the spotted alfalfa aphid, *Therioaphis trifolii***, which belongs to the aphid subfamily Calaphidinae. This assembly aims to enrich the genomic resources of non-Aphidinae species.

Although EBF has been extensively studied as the most common alarm pheromone in aphids, some species use other chemicals as their alarm pheromones. Therefore, we hypothesized that the molecular mechanism of EBF detection has diverged among aphids. **In chapter 6**, **we wanted to study the evolutionary process of EBF-tuned ORs within aphids.** With the availability of aphid genome assemblies of different subfamilies, we were able to annotate and compare OR repertoires between multiple species. We then identified various significantly conserved Aphidinae-specific OR groups, and we perform two-electrode voltage-clamp technique, RNA interference, and behavioral experiments to study the function of a newly identified EBF-tuned OR.

Finally, **in the last chapter**, all previous chapters were **globally discussed**, resulting in the formulation of **conclusions and perspectives**.

Chapter III

Identification and functional characterization of a ApisOr23 in pea aphid *Acyrthosiphon pisum*

Adapted from the following reference: Huang TY, Zhang RB, Yang LL, Cao S, Francis F, Wang B, Wang GR. Identification and functional characterization of ApisOr23 in pea aphid *Acyrthosiphon pisum* [J]. *Journal of Integrative Agriculture*, 2022, 21(5): 1414-1423.

Introduction to chapter III.

In Chapter 1, the olfaction system of aphids and their interaction with host plant volatiles and aphids were briefly introduced. While the chemical ecology of aphids has been extensively studied for decades, the molecular mechanism of olfaction in aphids still remain largely unknown.

Host plant volatiles are crucial chemical cues for aphids to locate their host, making the chemosensory system of aphids, particularly the OR gene family, a valuable molecular target for developing attractants. Additionally, ORs tuned to volatiles that play a significant role in mediating aphid behaviors are hypothesized to be evolutionarily conserved among different species. Therefore, comparative analysis of the OR gene family between various species is essential to identify significantly conserved OR genes, and functional studies on these ORs can shed light on their role in detecting host plant volatiles.

The objective of this is to identify conserved OR genes among different aphid species and perform functional studies to characterize the OR tuned to host plant volatiles.

Abstract: Pea aphid, *Acyrthosiphon pisum*, is a serious pest of many different leguminous plants, and it mainly relies on its odorant receptors (Ors) to discriminate among host species. However, less is known about the role that Ors play in the host plant location. In this study, we identified a novel conserved odorant receptor clade by phylogenetic analysis, and conducted the functional analysis of ApisOr23 in *A. pisum*. The results showed that the homologous Ors from *A. pisum*, *Aphis glycines* and *Aphis gossypii* share 94.28% identity in amino acid sequences. Moreover, conserved motifs were analyzed using the annotated homologous Or23 from eight aphid species, providing further proof of the high conservation level of the Or23 clade. According to the tissue expression pattern analysis, ApisOr23 was mainly expressed in the antennae. Further functional study using a heterologous *Xenopus* expression system revealed that ApisOr23 was tuned to five plant volatiles, namely *trans*-2-hexen-1-al, *cis*-2-hexen-1-ol, 1-heptanol, 4′-ethylacetophenone, and hexyl acetate. Among them, trans-2-hexen-1-al, which is one of the main volatile organic compounds released from legume plants, activated the highest response of ApisOr23. Our findings suggest that the conserved Or23 clade in most aphid species might play an important role in host plant detection.

Keywords: *Acyrthosiphon pisum*, odorant receptor, phylogenetic analysis, two-electrode voltage clamp, trans-2-hexen-1-al.

1. Introduction

Insects rely on their chemosensory organs to detect and decipher a variety of chemical cues in natural environment (Hansson & Stensmyr, 2011). The insect olfactory system plays a vital role in many critical behaviors related to host plant location, natural enemy avoidance, mate interactions and oviposition site selection (L. Chen et al., 2020; Leal, 2013; Y. Liu, Liu, Lin, & Wang, 2013; Wada-Katsumata, Robertson, Silverman, & Schal, 2018). In the process of odor reception, odorants enter into the antennal sensillum lymph through pores, where they are bound and transported by odorant binding proteins (OBPs), and are subsequently released to odorant receptors (ORs) which are expressed on the dendritic membranes of olfactory sensory neurons (OSNs) (Suh, Bohbot, & Zwiebel, 2014). Insect ORs are critical elements in the process of chemical signal transmission, as they convert the chemical signal into electrical signal which are transmitted by the central nervous system and subsequently induce a series of corresponding behavioral responses (Hallem et al., 2004).

One efficient way to understand how the OR repertoires of insects contribute to their adaptation to a particular environmental cue is to identify the specific ligands of these ORs. Therefore, the functions of ORs in insects have been intensively studied by using different expression systems, including many in vitro systems that are performed in *Xenopus* oocytes (G. Wang, Vasquez, Schal, Zwiebel, & Gould, 2011), HEK293 cells (Forstner, Breer, & Krieger, 2009) or Bm5 cells (Tsitoura et al., 2010). Other *in vivo* systems are conducted by transgenic *Drosophila* techniques with the “empty neuron” system (Dobritsa et al., 2003) or the OR67d GAL4 knock-in system (Kurtovic et al., 2007; Bing Wang et al., 2016), RNA interference (RNAi) (Pan, Yang, Romeis, Siegfried, & Zhou, 2020), or the clustered regularly interspaced short palindromic repeat/CRISPR-associated nuclease 9 (CRISPR/Cas9) system (Chang et al., 2017).

Aphids, which constitute a major family of Hemiptera, feed exclusively on plants by inserting their stylet into the sieve elements to suck sap (Moreno et al., 2011). Among 5 000 aphid species, many are agricultural pests and can not only feed on the phloem of plants, but also transmit plant viruses (Hodge & Powell, 2010). *Acyrthosiphon pisum* is the first aphid species with sequenced genome, and its genome was re-sequenced recently (International Aphid Genomics, 2010; Y. Li, Park, Smith, & Moran, 2019); it also serves as a model for studying molecular aspects related to various biological features, such as wing dimorphism (B. Li et al., 2020; Shang et al., 2020), sex chromosome evolution (Jaquiery et al., 2018), horizontal gene transfer (Moran & Jarvik, 2010), symbiont association (Hansen & Moran, 2011; Manzano-Mari et al., 2020), and host plant adaption (Jaquiery et al., 2012). However, few studies have examined the chemosensory mechanisms of aphids. Currently, most studies in this area have focused on the identification and expression profiling of aphid chemoreceptors and OBPs (Robertson, 2019; Q. Wang et al., 2019), but only a few ORs of pea aphid have received complete functional characterization. Our previous studies have shown that ApisOR5 is the receptor of the main alarm pheromone compound (*E*-β-farnesene) and ApisOr4 is

broadly tuned to eight plant volatiles (Francis et al., 2005; R. Zhang et al., 2017; R. B. Zhang, Liu, Yan, & Wang, 2019). Several studies also demonstrated the importance of chemical reception in aphids (Vandermoten, Mescher, Francis, Haubruge, & Verheggen, 2012), such as aphid–plant interactions (Sobhy et al., 2017) and particularly the host plant selection (Dardouri, Gautier, Ben Issa, Costagliola, & Gomez, 2019; de Oliveira et al., 2020). Therefore, uncovering the mechanism of odorant reception in aphids will contribute to the development of new ways to control the aphids.

In this study, we focused on a highly conserved OR clade (ApisOR23) identified from the phylogenetic tree of three aphid species. We cloned the ApisOR23 gene from *A. pisum* antennae and further analyzed the conserved protein motifs of the OR23 clade among eight aphid species. Then, we analyzed tissue expression patterns by semi-quantitative RT-PCR. Moreover, a functional analysis was performed using the *Xenopus* oocyte system, in order to find chemicals able to stimulate ApisOR23. Our results shed light on the molecular mechanisms of those host plant detection in *A. pisum*, and will contribute to the discovery of novel pea aphid attractants or repellents.

2. Materials and methods

2.1. Insect rearing

The pea aphid *A. pisum* was fed on potted broad bean plants (*Vicia faba* L.) at the Institute of Plant Protection, Chinese Academy of Agricultural Sciences, Beijing, China. Clonal rearing was maintained under constant environmental conditions, 21±2°C and 70±5% relatively humidity with a 16 h light: 8 h dark cycle.

2.2. RNA extraction and cDNA synthesis

Different pea aphid tissues, including 600 antennae, 300 heads without antennae, 360 legs and 5 bodies, were collected and immediately frozen in liquid nitrogen and stored at –80°C before RNA extraction. Total RNA was extracted using Trizol reagent (Invitrogen, Carlsbad, CA, USA) according to the manufacturer's instructions, and was exposed to DNase I (Thermo Scientific, USA) to remove genomic DNA. Reverse transcription was performed using RevertAid First Strand cDNA Synthesis Kit (Fermentas, Vilnius, Lithuania). We selected the cDNA sample from antennae as the template for ApisOr23 cloning, and the cDNA samples of the four tissues mentioned above were used for semi-quantitative RT-PCR.

2.3. Identification of homolog ApisOR23 in different aphids

The OR23 genes in five aphid genomes (*Rhopalosiphum maidis* (W. Chen et al., 2019), *Sitobion miscanthi* (Jiang et al., 2019), *Diuraphis noxia*, *Myzus cerasi* (Legeai et al., 2010) and *Myzus persicae* (Mathers et al., 2017) were annotated using ApisOR23 (Smadja et al., 2009) for the query in TBLASTN searches of the genome assembly (with a cutoff 10–5), and the genes obtained were named as *RmaiOR23*, *SmisOR23*, *DnoxOR23*, *McerOR23*, and *MperOR23*, respectively.

Gene models were checked manually. The amino acid sequences of these genes are listed in Figure S1.

2.4. Sequence and phylogenetic analysis

The transmembrane domains of ApisOR23, AglyOR14 and AgosOR23 were predicted by TMHMM version 2.0 (http://www.cbs.dtu.dk/services/TMHMM/). The alignment of the amino acid sequences was generated by DNAMAN version 8 (Lynnon LLC, San Ramon, CA, USA), and carried out using the ORs sequences from three aphid species (*A. pisum*, *Aphis glycines*, and *Aphis gossypii*) (D. Cao, Liu, Walker, Li, & Wang, 2014; Robertson, 2019). AgosOR18, AgosOR31, AgosOR40 and AgosOR44 were excluded because of their depressed annotation quality. The alignment was generated by Mafft version 7.0 (Katoh & Standley, 2013) with default settings, and trimmed by TrimAI version 1.4 with the "gappyout" option (Capella-Gutierrez, Silla-Martinez, & Gabaldon, 2009). The motifs that were conserved among the aphids were identified by the MEME Program (Bailey et al., 2009) with a maximum number of motifs of ten, and decorated by TBtools Program (Chen et al., 2020). The phylogenetic analysis was conducted by MEGA7 (Kumar, Stecher, & Tamura, 2016) using the neighbor-joining method, and node support was assessed using a bootstrap procedure of 1000 replicates. The resultant tree was constructed by Evolview version 2 (He et al., 2016).

2.5. Molecular cloning

The open reading frame (ORF) of ApisOR23 was cloned using a coding sequence identified from the first version of the *A. pisum* genome. The 25 μL PCR reaction system contained 0.25 μL. PrimeSTAR HS DNA polymerase (2.5 units μL^{-1}), 1 μg μL^{-1}cDNA template, 5 μL 5× PrimeSTAR buffer (Mg^{2+} Plus), 2 μL dNTP mixture (2.5 mmol L^{-1} of each), and 10 μmol $L^{-1}$of each primer. The PCR was performed according to the following conditions: 94°C for 5 min; 40 cycles of 94°C for 30 s, 55°C for 30 s, 72°C for 30 s; and 72°C for 10 min. The PCR products were ligated into the cloning vector pEASY-Blunt (TransGen Biotech, Beijing, China) and verified by DNA sequencing. The ORF of ApisOr23 was ligated into the pT7TS expression vector using specific primers with restriction enzyme cutting sites (Table S1).

2.6. Tissue expression pattern of ApisOr23 gene

The expression pattern of ApisOR23 was detected by semi-quantitative RT-PCR. The first cDNA strand was synthesized from the RNA of each tissue, namely antennae (A), heads without antennae (H), legs (L) and bodies (B). The succinate dehydrogenase B (ApisSDHB) gene (GenBank accession number: NM_001162436) (Yang, Pan, Liu, & Zhou, 2014) was selected as a reference. The specific primers used in RT-PCR are listed in Table S1. The RT-PCR reactions were performed using EasyTaq SuperMix (TransGene, Strasbourg, France) under conditions of 95°C for 3 min; 28 cycles of 94°C for 30 s, 55–60°C for 30 s, 72°C for 30 s; and 72°C for 10 min. The experiment was biologically repeated three times.

2.7. Chemical compounds

The 57 representative compounds used in this study are listed in Table 2 and Table S2. These compounds include common host plant volatiles and aphid alarm pheromones.

2.8. OR expression in Xenopus oocytes and electrophysiological recordings

The ORF of *ApisOR23* was subcloned into the pT7TS vector based on the restriction enzyme digestion sites. The cRNA was synthesized by mMESSAGE mMACHINE T7 Kit (Ambion, Austin, TX, USA). Mature healthy oocytes were treated according to a previous study (R. B. Zhang et al., 2019). Oocytes were microinjected with 27.6 ng of ApisOR23 cRNA and 27.6 ng ApisOrco cRNA, then cultured for 4-7 days at 18°C. The cell currents induced by the odorants were recorded with a two-electrode voltage clamp (TEVC). Data acquisition and analysis were performed with Digidata 1440 A and Pclamp10.0 Software (Axon Instruments Inc., Union City, CA, USA). Each odorant used in this study (Table S2) was prepared as a 1 mol L^{-1} stock solution in dimethyl sulphoxide (DMSO) and stored at −20°C. Before the experiments, stock solutions were diluted in 1× Ringer's buffer to a final concentration of 10^{-4}mol L^{-1}. Data were analyzed using software SAS 9.1, by the one-way analysis of variance (ANOVA) followed by the Duncan's multiple range test. Statistical significance was determined at the α=0.05 level.

3. Results

3.1. Phylogenetic and conserved motif analysis of the OR23 clade

Previous studies have shown that conserved ORs might ensure a number of crucial biological functions in aphids, such as alarm pheromone detection (Zhang et al., 2019). Therefore, we selected ORgenes from three aphid species, including *A. pisum*, *A. glycines* and *A. gossypii*, that were annotated from previous genome studies (see methods). The amino acid sequences of these ORs were used for phylogenetic analysis in order to discover the conserved OR clade. Intriguingly, ApisOR23, AgosOR23 and AglyOR14 were clustered together and showed a highly homologous relationship among these three aphids, indicating that this clade is relatively well-conserved (Figure 5).

3.2. Gene cloning and sequence analysis

The sequence of ApisOR23 was obtained from published data (Robertson et al. 2019). Specific primers were designed for cloning the full-length ORF of

ApisOR23 from antennal cDNA. The ORF of the *ApisOR23* was 1242 bp, encoding 414 amino acids. The alignment of amino acid identity showed that ApisOR23 shared 94.28% sequence identity with its ortholog AglyOR14 and AgosOR23, and possessed seven transmembrane domains (Figure 6).

To further confirm whether this clade is conserved among aphid species, we annotated the corresponding orthologs from five other aphids (*R. maidis*, *S. miscanthi*, *D. noxia*, *M. cerasi* and *M. persicae*). The amino acid sequences of the eight homolog OR23s from all eight aphids mentioned above were included in the conserved motif analysis by MEME Program (the full sequences are listed in Supplementary materials). A total of ten conserved motifs were predicted by the MEME Program (Figure S1). These ORs shared a highly conserved motif pattern, as each gene included all ten motifs, and the motifs were in almost the same order (motif order: 7-6-4-1-5-10-3-8-2-9) and locations (Figure 7). Furthermore, seven of the ten motifs possessed extremely high conservation, with P-values less than 10^{-190}. Such highly conserved amino acid sequence patterns indicated that these OR23s might tune to the same ligand spectrum, as the motifs covered almost all the receptor sequences, and consequently demonstrated the conservation of most of the functional amino acid sites.

3.3. Tissue expression pattern of ApisOR23

In order to investigate the expression pattern of ApisOR23, we selected SDHB as the reference gene, and carried out RT-PCR on the tissues of antennae, heads without antennae, legs and remaining bodies. The high expression level of ApisOR23 was found in the antennae, while considerably lower expression level was detected in the legs. No expression was found in the tissues from the heads and bodies (Figure 8).

3.4. Functional characterization of ApisOR23/Orco

The *Apis*OR*23*/*Orco* co-expressing *Xenopus* oocytes were used for functional characterization by two-electrode voltage clamps. A total of 57 plant volatiles were tested (listed in Table S2). ApisOR23 mainly tuned to five of the chemicals, including aromatic ketone (4′-ethylacetophenone) and aliphatic compounds (*cis*-2-hexen-1-ol, 1-heptanol, hexyl acetate and *trans*-2-hexen-1-al) (Figure 9-A and B; Table 1). However, there were no measurable responses to the other tested chemicals. The highest response of ApisOR23/Orco was induced by *trans*-2-hexen-1-al (207.81±17.63 nA), while 4′-ethylacetophenone and hexyl acetate activated the low responses with current values of 44.18±5.03 and 45.12±6.71 nA, respectively. Oocytes co-expressing ApisOR23/Orco had moderate responses to *cis*-2-hexen-1-ol and 1-heptanol (87.40±8.61 and 28.20±7.82 nA, respectively)

(Figure 9C).

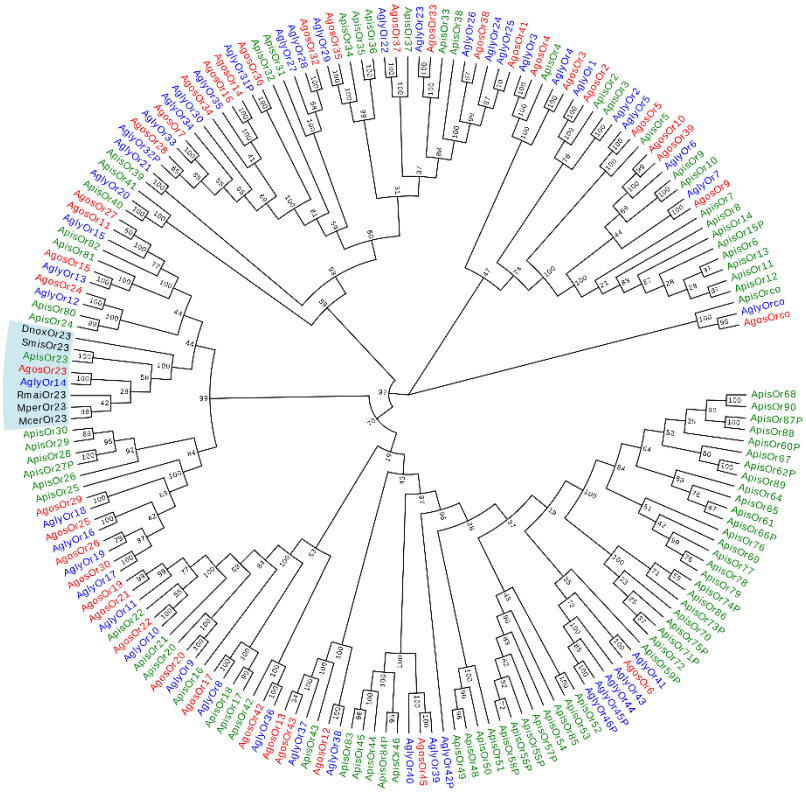


Figure 5. Phylogenetic analysis of odorant receptors (ORs) in *Acyrthosiphon pisum*, *Aphis gossypii* and *Aphis glycines* in addition to five OR23s from *Rhopalosiphum maidis*, *Sitobion miscanthi*, *Diuraphis noxia*, *Myzus cerasi* and *Myzus persicae*. The predicted amino acid sequences of the ORs were aligned using the Mafft V7.0 Program. The phylogenetic tree was constructed using the neighbor-joining (NJ) method with 1 000 bootstrap replicates by MEGA 5.0. This neighbor-joining tree was rooted with the Orco proteins, and indicated one highly conserved OR clade within many insect species. Abbreviations of *A. pisum*, *A. gossypii* and *A. glycines* are shown respectively as Apis in green, Agos in red, Agly in blue, and the five species listed above as Rmai, Smis, Dnox, Mcer, and Mper in black. The ApisOR23 clade is masked with light blue shadow.

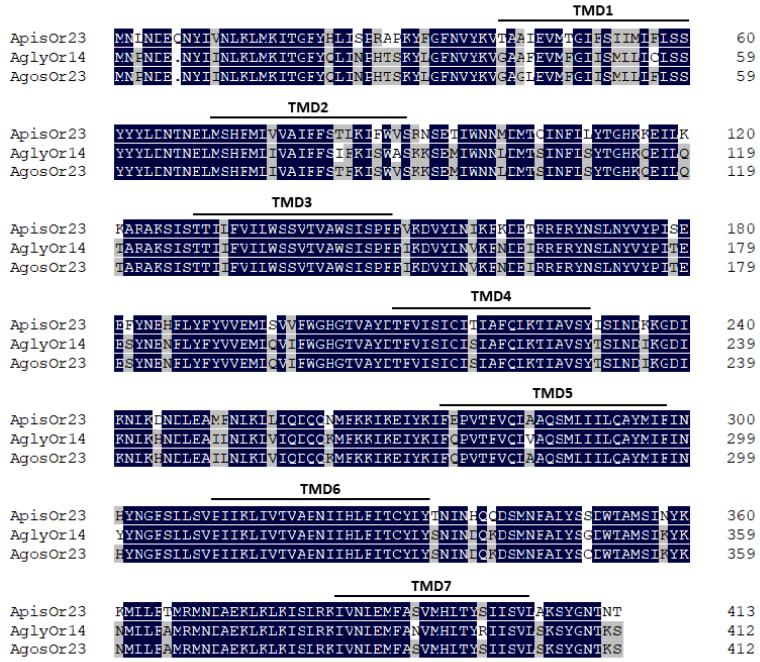


Figure 6. Sequence alignment of ApisOr23, AglyOr14, and AgosOr23. The amino acid identity of the three sequences is 94.28%. Conserved amino acids are covered in black boxes while the unique amino acid sites are represented by grey and white boxes. Seven transmembrane domains (TMD) are predicted and marked with black lines.

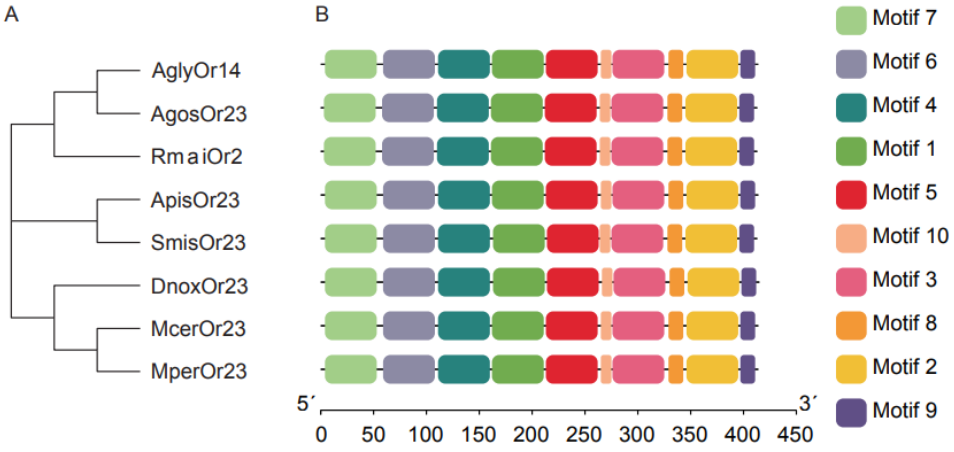


Figure 7. Phylogenetic analysis and conserved motifs of the OR23 clade of eight aphid species. A, phylogenetic tree of the eight species in the OR23 clade. Agly, *Aphis glycines*; Agos, *Aphis gossypii*; Rmai, *Rhopalosiphum maidis*; Apis, *Acyrthosiphon pisum*; Smis, *Sitobion miscanthi*; Dnox, *Diuraphis noxia*; Mcer, *Myzus cerasi*; Mper, *Myzus persicae*. B, schematic distribution of conserved motifs in the OR23 clade. Motif analysis was carried out using MEME Software. The colored boxes represent conserved motifs that were located in the corresponding location of each OR.

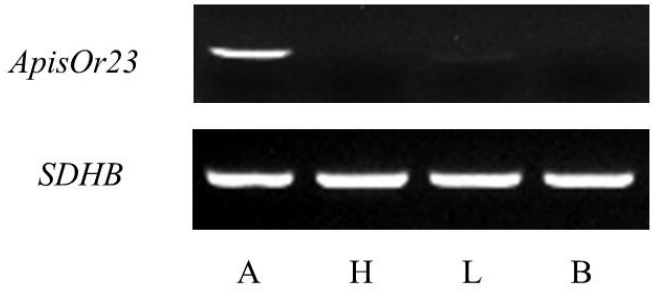


Figure 8. Tissue expression pattern of ApisOR23 using semi-quantitative RT-PCR. The succinate dehydrogenase B (*SDHB*) gene (GenBank: NM_001162436) was selected as reference. A, antennae; H, heads (antennae removed); L, legs; B, bodies.

4. Discussion

ORs play an important role in the process of host plant volatile detection among various insect species. The functions of ORs from the model species *Drosophila melanogaster*, as well as many other species from Lepidoptera, Hemiptera and Diptera, have been studied in recent years (Cui et al., 2018; Dweck, Ebrahim, Farhan, Hansson, & Stensmyr, 2015; Khashaveh et al., 2020; Y. Liu, Cui, Wang, Zhou, & Liu, 2020; C. Wang et al., 2020; Wicher et al., 2008). Nevertheless, only few studies were conducted on the function of ORs in aphids. The pea aphid has a complex plant-specialized population, displaying a highly adaptive evolution (Duvaux et al., 2015; Eyres et al., 2016). It can feed on multiple legumes, while many other aphids are reported to be specialists (Ragsdale, Voegtlin, & O'Neil, 2004). With the support of increasingly available genomics and transcriptomics data, more than 70 OR genes have been identified from the *A. pisum* genome (Robertson, 2019), and most OR genes have experienced recent and rapid expansion, which might indicate that such gene expansion is essential for host plant acceptance (Caillaud & Via, 2000; Smadja et al., 2009; 2012).

The odorant receptor gene family evolves under a birth-and-death process, which means ORs genes undergo many evolutionary events, including duplications, deletions, pseudogenizations and positive selection (McBride, Arguello, & O'Meara, 2007). Comparisons of the OR gene family members from diverse insect species have revealed striking differences in gene family size (Robertson 2019). Although such a feature is remarkably common, even in closely related species, we

still noticed that many receptors remain quite conserved among different aphid species. These conserved ORs not only showed sequence similarity, but also possessed highly consistent odorant response profiles (S. Cao, Liu, Guo, & Wang, 2016). Moreover, the highly conserved ORs may play an important role in the key life processes of insects. For example, the major component of aphid alarm pheromone, (*E*)-β-farnesene, was detected by two highly conserved odorant receptors from *A. pisum* and *A. gossypii* (Zhang et al., 2017). Here, we identified another odorant receptor of *A. pisum*, named ApisOR23. The amino acid sequences are conserved among three different aphid species, which utilize relatively distinctive host species. In order to further confirm that the OR23 clade is conserved among different aphids, we annotated ApisOR23 homologs from five other aphids, and performed the conserved motif analysis. The OR23 clade was shown to be considerably conserved among the different species, suggesting that the OR23 genes of various aphid species might play an essential role in host plant location or other behaviors, such as oviposition site-selection. Future works on the functions of other OR23 clade members would provide more evidences for this hypothesis.

It has reported that *cis*-2-hexen-1-ol, hexyl acetate and trans-2-hexen-1-al are the most common green leaf volatiles (GLVs) from plants. Specially the latter one is the main volatile released from legumes (Pareja, Mohib, Birkett, Dufour, & Glinwood, 2009). The attractiveness of GLVs (including trans-2-hexen-1-al) has been reported in the black bean aphid *Aphis fabae* (Webster et al., 2008). In addition, *trans*-2-hexen-1-al showed a significant attractiveness to the tea aphid *Toxoptera aurantia* (Bian, Sun, Cai, & Chen, 2014; B. Han, Zhang, & Byers, 2012), suggesting that *trans*-2-hexen-1-al is involved in the attraction of various aphid species. Therefore, we hypothesized that trans-2-hexen-1-al may also attract the pea aphid. In this study, we found that *trans*-2-hexen-1-al was the best ligand for ApisOR23, and the OR23 clade is significantly conserved among the eight aphid species, indicating that OR23 in aphids is involved in signal discrimination of *trans*-2-hexen-1-al. However, we cannot exclude the possibility that there may be other odorant receptors which also respond to *trans*-2-hexen-1-al, leading to some kind of combinatorial coding that could affect the aphid's behavior. For example, numerous ORs from *D. melanogaster* showed responses to *trans*-2-hexen-1-al, including Or7a, OR35a, OR42a and OR67b, and others, among which DmelOR7a was the main receptor tuned to *trans*-2-hexen-1-al, and it is involved in aggregation behavior and oviposition site-selection (Kreher et al., 2005; Kreher, Mathew, Kim, & Carlson, 2008; Lin, Prokop-Prigge, Preti, & Potter, 2015). Therefore, future work should systematically study the peripheral coding map of aphids to odorant detection and host selection behavior in order to elucidate the molecular mechanisms of the detection of host plant volatiles.

Interestingly, *trans*-2-hexen-1-al has also proven to be one of the main herbivore-induced plant volatiles (HIPVs) induced by a chewing herbivore, the beet armyworm caterpillar, *Spodoptera exigua* (Schwartzberg, Boroczky, & Tumlinson, 2011). So, it may act as an indirect defensive signal of plants by repelling pests or attracting natural enemies (Allmann & Baldwin, 2010). When caterpillars and

aphids co-occur on the same plant, caterpillar-induced *trans*-2-hexen-1-al may act as a negative signal reducing the aggregation of aphids because they prefer undamaged plants rather than caterpillar-infested plants (Ray et al., 2020). Therefore, we hypothesize that *trans*-2-hexen-1-al acting as common GLVs (at low concentration) may attract *A. pisum*. However, *trans*-2-hexen-1-al acting as HIPVs are induced in quantity (at high concentration) displaying repellent effect on the aphids when caterpillars and aphids co-occur on the same niche. This phenomenon has been proved recently that attractions and aversions of *D. melanogaster* to alcohol are mediated by three compounds.

5. Conclusions

In this study, we annotated the OR23 genes from five aphid genomes, and the subsequent phylogenetic and conserved motif analysis showed that the OR23 clade was highly conserved among the different aphids. By using a heterologous expression system in *Xenopus* oocytes, the response of ApisOR23/Orco was activated by five plant volatiles, of which *trans*-2-hexen-1-al released from legume plants presented the highest response level. This result indicated that *trans*-2-hexen-1-al might act as an important chemical cue for host selection of aphid.

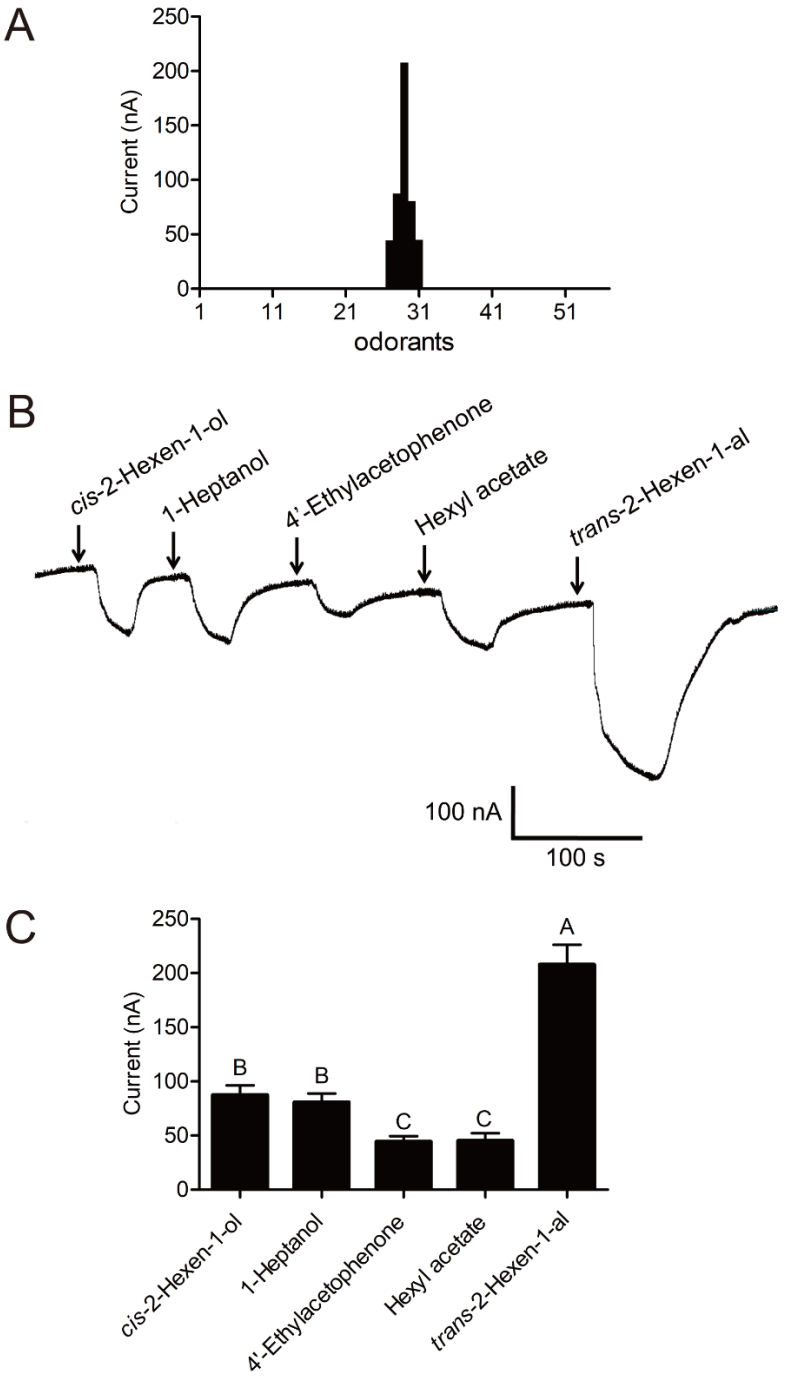
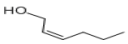
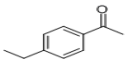
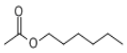
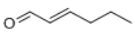

A
Current (nA)
250
200
150
100
50
0
1
11
21
31
41
51
odorants
B
cis-2-Hexen-1-ol
1-Heptanol
4'-Ethylacetophenone
Hexyl acetate
trans-2-Hexen-1-al
100 nA
100 s
C
Current (nA)
250
200
150
100
50
0
B
B
C
C
A
cis-2-Hexen-1-ol
1-Heptanol
4'-Ethylacetophenone
Hexyl acetate
trans-2-Hexen-1-al

Figure 9. Functional characterization of ApisOr23/Orco response to odorants in the *Xenopus* oocyte system. A, tuning curves of ApisOR23 to 57 individual plant volatiles. The x-axis shows the number of the tested odorants (Table S2). The y-axis shows the strength of current values (nA) of the ApisOR23/Orco upon exposure to the odorants, with the strongest response in the center. B, inward current responses of ApisOR23/Orco stimulated with five odorants as 10^{-4} mol L^{-1} solution. C, response profile of ApisOR23/Orco. Error bars indicate mean±SEM (n=13). Different uppercase letters indicate significant differences of the current values stimulated by different odorants ($P<0.05$).

Table 2 The odorants tuning to ApisOR23

Name	CAS	Chemical formula	Structural formula
cis-2-Hexen-1-ol	928-94-9	$C_6H_{12}O$	
1-Heptanol	111-70-6	$C_7H_{16}O$	
4′-Ethylacetophenone	937-30-4	$C_{10}H_{12}O$	
Hexyl acetate	142-92-7	$C_8H_{16}O_2$	
trans-2-Hexen-1-al	6728-26-3	$C_6H_{10}O$	

separate neural pathways, as DmelOR42b and DmelOR59b are necessary for the attraction to alcohol at low concentration, while aversion behavior to high concentration level of alcohol is detected by DmelOR42a (Keesey et al., 2020). Therefore, we speculate that other ORs in aphids, besides ApisOR23, may involve in mediating the attraction or aversion to *trans*-2-hexen-1-al in aphids.

6. Acknowledgments

This work was funded by the National Natural Science Foundation of China

(31572072 and 31725023), the Intergovernmental International Science, Technology and Innovation Cooperation Key Project, China (2019YFE0105800), and the Shenzhen Science and Technology Program, China (KQTD20180411143628272.). The funders had no role in study design, data collection or analysis, decision to publish or preparation of the manuscript.

Chapter IV

A conserved odorant receptor identified from antennal transcriptome of *Megoura crassicauda* that specifically responds to *cis*-jasmone

Adapted from the following reference: Wang B, Huang TY, Yao Y, Francis F, Yan CC, Wang GR, WANG B. A conserved odorant receptor identified from antennal transcriptome of *Megoura crassicauda* that specifically responds to *cis*-jasmone. Journal of Integrative Agriculture, 2022, 21(7): 2042-2054.

Introduction to chapter IV

Herbivore-induced plant volatiles (HIPVs) are essential chemical cues for natural enemies to locate their preys, as well as to mediate the behavioral responses of herbivorous insects like aphids. HIPVs have been widely used for biological control in the crop fields. However, the olfaction mechanism of HIPVs in aphids remains unknown. In this chapter, we wanted to identify ORs responsible for detecting HIPVs in aphids. Previous aphid OR functional studies mostly used the pea aphid *A. pisum* as research model, but *A. pisum* maybe not the ideal model for mining HIPV-tuned ORs due to its significantly wide host range. As a generalist, *A. pisum* can feed on a diversity range of legume plants, a significant part of its ORs may be involved in sensing host plant volatiles that not influenced by herbivory, which makes it difficult to identify HIPVs-specific ORs. Therefore, specialist aphid possessing a minimalist OR repertoire may be more helpful in identifying such ORs. Our research provided a detailed annotation of the antennae transcriptome of *Megoura crassicauda*, which also updated valuable omics resources for aphids.

Abstract: While aphids are common and serious phloem-feeding pests in farmland ecosystems, little is known about how aphids use their sensitive olfactory system to detect HIPVs. In this study, the antennal transcriptomes of the aphid species *M. crassicauda* were sequenced, and expression level analyses of *M. crassicauda* odorant receptors (ORs) were carried out. To investigate the chemoreception mechanisms that *M. crassicauda* uses to detect HIPVs, we performed *in vitro* functional studies of the ORs using 11 HIPVs reported to be released by aphid-infested plants. In total, 54 candidate chemosensory genes were identified, among which 20 genes were ORs. McraOR20 and McraOR43 were selected for further functional characterization because their homologs in aphids were quite conserved and their expression levels in antennae of *M. crassicauda* were relatively high. The results showed that McraOR20 specifically detected *cis*-jasmone, as did its ortholog ApisOR20 from the pea aphid *A. pisum*, while McraOR43 did not respond to any of the HIPV chemicals that were tested. This study characterized the ability of the homologous OR20 receptors in the two aphid species to detect HIPV *cis*-jasmone, and provides a candidate olfactory target for mediating aphid behaviors.

Keywords: *Megoura crassicauda*, transcriptome, chemosensory genes, odorant receptors, *cis*-jasmone.

1. Introduction

Insects rely on their highly sensitive sensory system to distinguish various chemical signals in a complex natural environment and trigger relevant behavioral responses, such as host plant location, mate selection and predator avoidance (Bruce, Wadhams, & Woodcock, 2005; Fleischer & Krieger, 2018). The accurate

detection of chemical signals by insects is mainly accomplished by the products of several gene families, including odorant receptors (ORs), gustatory receptors (GRs), ionotropic receptors (IRs), odorant-binding proteins (OBPs), and chemosensory proteins (CSPs) (Benton et al., 2009; Clyne, Warr, & Carlson, 2000; Clyne et al., 1999; Hendry, 2001; Pelosi, Zhou, Ban, & Calvello, 2006). During olfactory reception, OBPs and CSPs are responsible for binding hydrophobic odorants, surrounding olfactory sensory neurons (OSNs) and transporting the odorants to chemosensory receptors located on the dendrites of OSNs (Sandler, Nikonova, Leal, & Clardy, 2000; J. S. Sun, Xiao, & Carlson, 2018). The OR, GR and IR gene families are the key receptors in the chemosensory system (Jacquin-Joly & Merlin, 2004). OR genes usually co-express with odorant receptor co-receptor (Orco), forming a ligand-gated channel that plays a role in the transduction of chemical signals (Joseph & Carlson, 2015). They are widely tuned to various chemicals that exist in the environment, such as alcohols, esters, and ketones (Song Cao et al., 2020; Y. Liu, Z. Cui, G. Wang, et al., 2020). GRs are mainly expressed in gustatory organs such as the mouthparts and are involved in taste and contact stimuli (Chyb, 2004). They are responsible for the detection of non-volatile compounds such as sugars, salts, and bitter compounds, and even for the detection of carbon dioxide (CO2) (Dahanukar, Lei, Kwon, & Carlson, 2007; Jiao, Moon, & Montell, 2007; Walton D. Jones, Cayirlioglu, Grunwald Kadow, & Vosshall, 2006; Sung et al., 2017). IRs constitute a relatively newly described chemosensory receptor gene family (Benton et al. 2009). Recent functional studies of IRs have revealed that they are capable of perceiving a wide range of environmental factors, including odorants, humidity and temperature (Budelli et al., 2019; Y. Chen & Amrein, 2017; Hassan et al., 2016; Knecht et al., 2017).

Aphids are phytophagous hemipteran insects, with around 5 000 species that are globally distributed. Some of them are primary crop pests in many regions, causing major economic losses (Pickett et al., 2003). Like other insects, aphids use their chemosensory system to accurately locate host plants and detect their intraspecific alarm signals (J. Fan et al., 2017; Z. Q. Li et al., 2018). With the development of sequencing technology, numerous aphid genomes and transcriptomes have been sequenced, providing essential resources for mining chemosensory genes. OBPs and CSPs have been identified by genome and transcriptome analyses in many aphid species, including *Aphis gossypii* (Gu et al., 2013), *Acyrthosiphon pisum* (Zhou et al., 2010), *Myzus persicae* (Wang et al., 2019), *Daktulosphaira vitifoliae* (J. J. Zhao, Zhang, Fan, & Feng, 2017), *Megoura viciae* (Bruno et al., 2018), and *Sitobion avenae* (Xue et al., 2016). Most functional studies focus on investigating the detection of plant volatiles and aphid alarm pheromones such as (*E*)-β-farnesene (EBF) and EBF derivatives (J. Fan et al., 2017; Northey et al., 2016; Qiao et al., 2009; Qin et al., 2020; Zhong, Yin, Deng, Li, & Cao, 2012).

ORs, GRs, and IRs have been identified from the genomes of *A. gossypii*, *A. pisum*, and *Aphis glycines*, each showing significant differences in the numbers of genes. The OR gene family of *A. pisum* (87 ORs) is almost twice as large as those of *A. glycines* and *A. gossypii* (47 and 45 ORs, respectively), resulting from the recent expansions of particular gene lineages in *A. pisum*. IRs are presented as simple orthologs between different aphid species. For example, 14 IRs are found in *A. gossypii*, while 19 IRs are found in both *A. pisum* and *A. glycines* (D. Cao et al., 2014; Robertson, Robertson, Walden, Enders, & Miller, 2019; C. Smadja et al., 2009). Although chemosensory receptors have been reported extensively, only a few have been functionally characterized. Previous studies have shown that ApisOR5 is responsible for EBF detection (Zhang et al., 2017) and ApisOR4 could respond to various plant volatiles (Zhang et al., 2019). SaveOrco is not only involved in the olfactory response to plant volatiles and EBF, but also in wing differentiation triggered by EBF (Fan et al., 2015).

When aphids feed on plants, plant defense responses are triggered by the systemic release of various secondary metabolites, some of which are volatile compounds commonly referred to as herbivore-induced plant volatiles (HIPVs) (Turlings & Tumlinson, 1992). HIPVs usually consist of green leaf volatiles (GLVs), which are six-carbon chemical compounds, and terpene volatiles, as well as a few other commonly released volatiles (e.g., methyl salicylate) (Abuin et al., 2011; Kroes, Weldegergis, Cappai, Dicke, & van Loon, 2017; T. C. J. Turlings & Erb, 2018). Studies have shown that a series of volatiles, including ocimene, α-pinene, *trans*-2-hexen-1-ol, *cis*-3-hexen-1-ol, linalool, and (–)-trans-caryophyllene, could be released from the broad bean plant *Vicia faba* after infestation with *A. pisum* and *Aphis craccivora* (Schwartzberg et al., 2011; Takemoto & Takabayashi, 2015); and potatoes infested by *Macrosiphum euphorbiae* produced *cis*-jasmone, (*E*,*E*)-4,8,12-trimethyl-1,3,7,11-tridecatetraene (TMTT), (*E*)-4,8-dimethyl-1,3,7-nonatriene (DMNT), *trans*-2-hexenal and methyl salicylate (Sobhy et al., 2017). Numerous HIPV compounds are also reported to be associated with aphid repellence (Birkett et al., 2000; Hegde et al., 2012), illustrating the important role of HIPVs in mediating plant-aphid interactions. However, the olfactory coding of aphids to candidate HIPVs in the peripheral nervous system remains unknown.

In this study, we selected the legume specialist aphid species *Megoura crassicauda*, a close relative of *M. viciae* (Kim & Lee, 2008), to perform antennal transcriptome sequencing. Chemosensory-related genes including ORs, GRs, IRs, OBPs, and CSPs were identified in the transcriptomes. Expression level analyses were carried out for all identified OR genes. Assuming that the interactions of ORs with structurally related HIPVs are conserved across aphid species, the sequences of homolog ORs from different aphid species were selected for functional testing. We therefore selected the full-length McraOR20 and McraOR43 genes, which

share high sequence homology with those of *A. pisum* and/or exhibited high expression levels, for further analysis. To verify our hypothesis, we cloned the full-length OR20 and OR43 genes, based on sequences from the *M. crassicauda* transcriptome, and characterized the functions of both McraOR20 and McraOR43 with 11 HIPVs using a two-electrode voltage clamp technique. This study preliminarily reveals the molecular mechanism of HIPV detection in aphids.

2. Materials and methods

2.1. Insect rearing and tissue collection

Megoura crassicauda aphids were fed on broad bean (*Vicia faba* L.) plants and maintained at 21±2°C with 70±5% relative humidity and a photoperiod of 16 h light:8 h dark. Clonal populations were reared at the Institute of Plant Protection, Chinese Academy of Agricultural Sciences, Beijing. Antennae and legs were dissected from apteral parthenogenetic adults, then instantly frozen and stored in liquid nitrogen.

2.2. cDNA library construction and Illumina sequencing

Total RNA was extracted using Quick-RNA Microprep Kit (Beijing Tianmo Biotech Company Limited, China). Total RNA was dissolved in RNase-free water, and the integrity of the RNA was determined by gel electrophoresis. RNA purity and concentration were determined on a Nanodrop ND-2000 spectrophotometer (NanoDrop Products, Wilmington, DE, USA). Three replicates containing tissues excised from 250 and 150 parthenogenetic adult females were generated for antennae and legs, respectively. A total of 2 μg total RNA from each sample (antenna (A1-A3), leg (L1-L3)) were used to construct the cDNA libraries. Library construction and Illumina sequencing were carried out by the Beijing Genome Institute (Shenzhen, China). The libraries were sequenced on the Illumina NoveSeq 6000 (Illumina, San Diego, CA, USA) platform, generating 200-bp long, paired-end reads.

2.3. Assembly and annotation of chemosensory-related genes

The clean reads were generated by removing low-quality reads and adaptors, and are available in the NCBI SRA database (Project number PRJNA674404). The clean reads were de novo assembled using Trinity v2.4.0 (Grabherr et al., 2011). TGICL v2.1 (Pertea et al., 2003) was used to filter duplicate and highly similar sequences from each sample (A1-A3 and L1-L3) to obtain the final assembly. For functional annotation, all transcripts were selected as queries for BLASTX searches against a pooled database of non-redundant (NR) and SwissProt protein sequences with an E-value cutoff set at 1E-5 (Y. Liu, Z. Cui, P. Si, et al., 2020; B. Wang, Liu, & Wang, 2017). To identify chemosensory-related genes, we integrated the

traditional BLAST-based method with a domain-based search approach. First, we selected transcripts of chemosensory-related genes from the functional annotation results according to specific keywords. For example, when screening for ORs, we extracted transcripts that were annotated as 'odorant receptor' or 'olfactory receptor'. Next, we compiled the chemosensory-related genes of other aphid species from previous studies (Robertson et al., 2019; Wang et al., 2019), used these genes as queries for BLASTN searches against all transcripts with E-values<1E-10, and selected the transcripts with the most hits. Furthermore, we used the HMMER v3.1 Software (Mistry, Finn, Eddy, Bateman, & Punta, 2013) to search for genes containing functional domains of ORs or GRs from the transcript datasets. Characteristic domains of ORs (ID no. PF02949) and GRs (ID no. PF08395) were downloaded from the Pfam protein family database (El-Gebali et al., 2019). Finally, candidate chemosensory genes identified using the above methods were merged and redundancies were removed, and the remaining transcripts were selected for analyses in the next step.

2.4. Sequence and phylogenetic analysis

The SignalP v4.0 server was used for predicting putative N-terminal signal peptides of OBPs and CSPs (Petersen, Brunak, von Heijne, & Nielsen, 2011). Amino acid sequence alignment was executed using MAFFT v7 (Katoh & Standley, 2013) with default parameters. The phylogenetic trees of *M. crassicauda* chemosensory genes were constructed by RAxML v8 using the Jones-Taylor-Thornton (JTT) amino acid substitution model (Stamatakis, 2014). Branch support was assessed by a bootstrap method based on 1 000 replicates. The dataset submitted for phylogenetic analysis consisted of the annotated *M. crassicauda* chemosensory genes, as well as previously reported sequences of ORs and GRs from *A. pisum* and *A. glycines* (Robertson et al. 2019); IRs from *Drosophila melanogaster* (Benton et al., 2009; Croset et al., 2010), *A. pisum* and *A. glycines* (Robertson et al. 2019); and OBPs and CSPs from *Myzus persicae* (Wang et al., 2019), *Sitobion avenae* (Xue et al., 2016), *A. pisum* (Zhou et al., 2010), *A. glycines* (Robertson et al.2019) and *A. gossypii* (Gu et al. 2013). DNAMAN v8 (Lynnon LLC, San Ramon, CA, USA) was used to obtain amino acid sequence alignments of the orthologous ORs between *A. pisum* and *M. crassicauda*. Putative chemosensory-related genes of *M. crassicauda* were named based on their homologs in the pea aphid *A. pisum*.

2.5. Tissue-specific expression analysis

Bowtie2 v2.2.5 was used to map the clean reads to the chemosensory-related transcripts (Langmead & Salzberg, 2012). RSEM v1.2.12 was applied to count the fragments per kilobase per million fragments (FPKM) values of each transcript (B.

Li & Dewey, 2011). Semi-quantitative reverse transcription PCR (RT-PCR) was also employed to analyze the expression level of putative OR genes. Tissue samples from three biological replicates were collected from the antennae and legs of asexual females. Total RNA was extracted using the Quick-RNA Microprep Kit (Beijing Tianmo Biotech Company Limited, China). The cDNA was synthesized using the Revert Aid First Strand cDNA Synthesis Kit (Thermo Scientific, Waltham, MA, USA) based on the manufacturer's protocol. Gene-specific primers were designed by Primer 3 (Untergasser et al. 2012) and synthesized by Sangon Biotech Co., Ltd. (Shanghai, China) (Table S3). EasyTaq PCR SuperMix was used for PCR reactions, using cycling conditions of 33 cycles of 95°C for 30 s, 55°C for 30 s, and 72°C for 30 s. The nicotinamide adenine dinucleotide (NADH) dehydrogenase gene of *A. pisum* (GenBank accession no. NM_001162436) (Yang et al., 2014) was used as the query for identifying its ortholog in *M. crassicauda*, and the identified NADH gene was named as *McraNADH*, which was then used as a control gene.

2.6. Molecular cloning

The open reading frames (ORFs) of *McraOR20*, *McraOR43*, *McraOrco*, *ApisOrco* and *ApisOR20* were used for cloning. The 50-μL PCR reaction system consisted of 0.25 μL TaKaRa EX Taq (5 U μL^{-1}), 2 μL cDNA template, 5 μL 10× EX Taq buffer (Mg2+ plus). (20 mmol L^{-1}), 4 μL dNTP mixture (2.5 mmol L^{-1} of each dNTP), and 10 μmol L^{-1} of each primer (Table S3). PCR conditions were as follows: 94°C for 3 min; 35 cycles of 98°C for 10 s, 55°C for 30 s, 72°C for 90 s, and 72°C for 5 min. PCR products were ligated into the pEASY-T3 vector (TransGen Biotech, Beijing, China), then sequenced by the Beijing Genome Institute (Shenzhen, China).

2.7. Chemical compounds

The 11 representative compounds used in this study are listed in Table S4. All compounds are HIPVs released by *V. faba*, which was used as the food resource for *M. crassicauda* in this study (Schwartzberg et al., 2011; Takemoto & Takabayashi, 2015). These HIPVs consist of three types of chemical compounds, including GLVs (*trans*-2-hexenal, *cis*-3-hexen-1-ol, *trans*-2-hexen-1-ol), terpenes (ocimene, (-)-trans-caryophyllene, linalool, α-pinene, DMNT, TMTT), and aliphatic compounds (*cis*-jasmone and methyl salicylate).

2.8. Receptor expression in Xenopus oocytes and electrophysiological recordings

The ORFs of *McraOR20*, *McraOR43*, *McraOrco*, *ApisOR20*, and *ApisOrco* were subcloned into the pT7TS vector using the ClonExpressII One Step Cloning Kit (Vazyme Biotech Co., Ltd., China). Specific primers with Kozak consensus

sequences are listed in Table S3. cRNAs were synthesized using the mMESSAGE mMACHINE T7 Kit (Ambion, Austin, TX, USA). Mature, healthy *Xenopus* oocytes were pre-treated in line with a previous study (Zhang et al. 2017). A mixture of 27.6 ng of OR cRNA and 27.6 ng of Orco cRNA was microinjected into the oocytes, which were then cultured for 4–7 days at 18°C in Ringer's solution (Y. Liu, Z. Cui, P. Si, et al., 2020). Currents triggered by the odorants were recorded by a two-electrode voltage clamp (TEVC). The data generated from TEVC were collected and analyzed using Digidata 1440A and pCLAMP v10.2 Software (Axon Instruments Inc., Union City, CA, USA).

Each odorant used for recording was prepared as a 1 mol L^{-1} stock solution in dimethyl sulphoxide (DMSO) and stored at −20°C. The stock solutions were diluted in 1× Ringer's buffer to a final concentration of 1×10^{-4} mol L^{-1} for response profile experiments with six replicates. For dose-response recording, serial dilutions of odorant stock solution were made at 1×10^{-8}, 1×10^{-7}, 1×10^{-6}, 3×10^{-6}, 1×10^{-5} and 3×10^{-5} mol L^{-1}, and six replicates were recorded. Data were analyzed using GraphPad Prism 7 (GraphPad Software Inc., San Diego, CA, USA).

3. Results

3.1. Antennal transcriptome sequencing and assembly

The three libraries of antennae of *M. crassicauda* (Mcra_Al_1, Mcra_Al_2 and Mcra_Al_3) were sequenced on the Illumina NoveSeq 6000 Platform. After filtering out low-quality and adaptor-polluted reads, 37.95, 40.33, and 39.82 million clean reads of each library were generated, respectively. The datasets generated for this study can be found in NCBI BioProject PRJNA674404. *De novo* assemblies led to the generation of 15 566, 16 765, and 18 047 transcripts in each replicate, respectively. Sequences from the three libraries were then merged and clustered into one set containing 15 984 transcripts. The final dataset had a total length of 47 860 781, average length of 2 994, N50 of 3 711 bp, and GC content of 34.44% (Table S5).

3.2. Identification of candidate chemosensory genes, phylogenetic analysis, and homology analysis

A total of 54 candidate chemosensory genes were identified from the transcriptomes of *M. crassicauda*. We identified 20 candidate ORs in the transcriptomes, among which six ORs and one Orco contained complete ORFs, ranging from 369 to 463 amino acids (aa) (Table S6). Phylogenetic analysis revealed that some clades which consisted of the ORs from *M. crassicauda*, *A. pisum* and *Aphis glycines* were highly conserved, such as the Orco, OR2, OR4,

OR5, and OR20 clades. The species-specific expansion was observed in the OR subfamilies of *A. pisum* and *Aphis glycines*. However, we did not find any OR expansion in *M. crassicauda* (Figure 10). To investigate the homology between the ORs of *M. crassicauda* and A. pisum, seven full-length McraORs were selected for performing amino acid alignment with their *A. pisum* orthologous genes. Besides Orco, four of the six ORs showed significantly high identities (greater than 80%) with their *A. pisum* orthologs, including OR43 (90.05%), OR20 (89.29%), OR37 (88.94%) and OR5 (86.65%) (Table 4). Four GRs were annotated from the transcriptomes, but only McraGR1 produced a full-length ORF of 419 aa, while McraGR5 and McraGR21 lacked the 5´ end of the ORF, and McraGR20 lacked both ends of the ORF (Table S7). Phylogenetic analysis using the GRs of A. glycines, A. pisum and M. crassicauda showed that McraGR1 and McraGR5 were clustered with putative sugar receptors (Robertson 2015) (Figure 11). We identified nine putative IRs, and the complete ORFs of three IRs were predicted (Table S8). Based on a phylogenetic analysis of the IRs with the fly species *D. melanogaster* and other aphid species, we annotated two putative co-receptors, McraIR8a and McraIR25a, the conserved IR clades McraIR40a, McraIR75d, McraIR75j, and McraIR93a; as well as divergent IRs in the aphid, McraIR323 and McraIR325. We failed to identify any IRs from the IR21a, IR68a, IR76b, IR 100a, IR322 and IR324 subsets (Figure 12).

A total of 12 OBPs were identified in *M. crassicauda*, all of which contained full-length ORFs. Eight putative OBPs in *M. crassicauda* (McraOBP2–10) showed high identities with the OBPs of *M. viciae*. A signal peptide was predicted in every OBP, except for McraOBP6 (Table S9). The phylogenetic tree showed that 10 of the 12 McraOBPs had clear orthologues in other species as the amino acid identities of OBPs among aphids ranged from 70.95 to 96.71%, while McraOBP14 and McraOBP15 did not display any homologs (Figure 13). Ten identified OBPs belonged to the classic OBP subfamily that typically contains six conserved cysteine residues, while McraOBP5 and McraOBP6 belonged to the plus-C OBP subfamily containing one additional cysteine as well as a conserved proline next to the sixth cysteine (Table S10). Nine candidate CSP genes containing full-length ORFs were identified. Among all the CSPs in *M. crassicauda*, only McraCSP1 lacked a signal peptide (Table S11). A phylogenetic tree was constructed using the amino acid sequences of CSPs from six aphid species, and the results showed that all the McraCSPs were clustered with CSPs from other aphids and did not show any expansion in this subfamily (Figure 14). The sequence alignment results showed that all the identified CSPs possess four conserved cysteines (Figure S2).

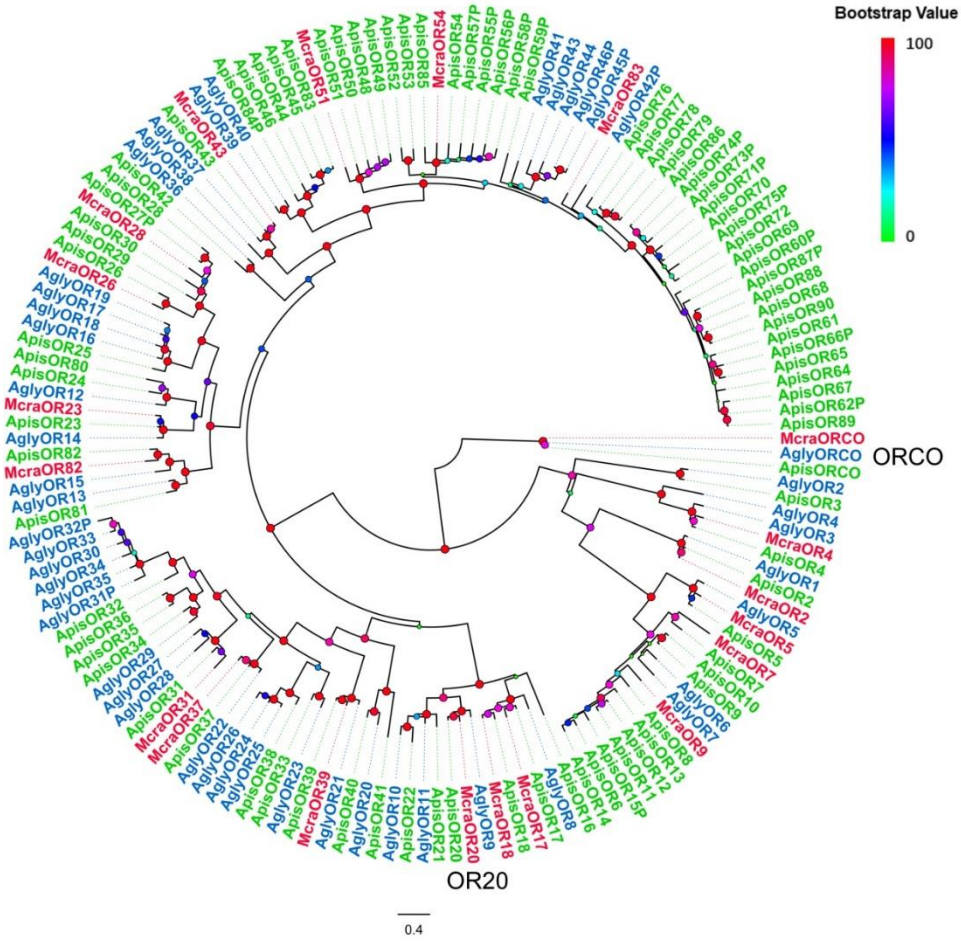


Figure 10. Phylogenetic tree of candidate odorant receptors (ORs) in *Aphis glycines* (Agly), *Acyrthosiphon pisum* (Apis) and *Megoura crassicauda* (Mcra). The distance tree was rooted by the conserved lineage of Orco.

Table 3. Homology analysis of odorant receptors between *Megoura crassicauda* and *Acyrthosiphon pisum*

M. crassicauda	*A. pisum*	Identity

McraOrco	ApisOrco	95.68%
McraOR43	ApisOR43	90.05%
McraOR20	ApisOR20	89.29%
McraOR37	ApisOR37	88.94%
McraOR5	ApisOR5	86.65%
McraOR51	ApisOR51	68.51%
McraOR7	ApisOR7	49.34%

3.3. *Tissue-specific expression patterns of candidate OR genes from Megoura crassicauda*

Expression profiles of the candidate OR genes in the transcriptomes of antennae and legs were assessed using FPKM values. All the ORs presented significantly high expression levels in antennae, with *McraOrco*, *McraOR4*, *McraOR7*, *McraOR20*, *McraOR26* and *McraOR28* possessing relatively high transcript abundances. We also found that some of the ORs were expressed at significantly lower levels in legs, including *McraOR4*, *McraOR18*, *McraOR26* and *McraOR54* (Figure S3 and S4). The tissue expression patterns of ORs were further investigated by RT-PCR. A total of 20 putative OR genes, as well as the control gene *McraNADH*, were examined. High expression level of *McraOrco*, *McraOR7*, *McraOR20* and *McraOR26* was observed, which is generally matched.

3.4. *Functional characterization of McraOR20/Orco and McraOR43/Orco in the Xenopus oocyte expression system*

We cloned *McraOR20* and *McraOR43* for functional analysis, as they produced full-length ORFs, displayed strong homology with their orthologs in *A. pisum*, and also showed relatively high expression levels in the antennae. These characteristics indicated that McraOR20 and McraOR43 may undertake important functions in chemoreception by *M. crassicauda*. In total, eleven HIPV compounds (Table S4), including DMNT, TMTT, ocimene, etc., were used to test the functions of McraOR20 and McraOR43 using TEVC technology. The results demonstrated that oocytes co-expressing McraOR20/Orco gave strong responses to 1×10^{-4} mol L^{-1} *cis*-jasmone, with an average current of 280.5±49.54 nA, but none of the other tested chemicals evoked responses of McraOR20/Orco at the same concentration (Figure 16-A and B). Moreover, the McraOR20/Orco response to *cis*-jasmone was concentration-dependent, from a threshold concentration of 1×10^{-8} mol L^{-1} to a

final concentration of 3×10^{-5} mol L^{-1}, and with an EC50 value of 1.116×10^{-6} mol L^{-1}(Figure 16-C and D). We also recorded responses of ApisOR20, the ortholog of McraOR20 in *A. pisum*. As expected, among the tested chemicals, only *cis*-jasmone elicited a response from ApisOR20/Orco (410.6±96.24 nA; Figure 16-B). A TEVC test for McraOR43/Orco was carried out in parallel. Surprisingly, McraOR43/Orco did not respond to any of the tested HIPV chemicals, suggesting that McraOR43 may not be involved in HIPV-detection (Figure S5).

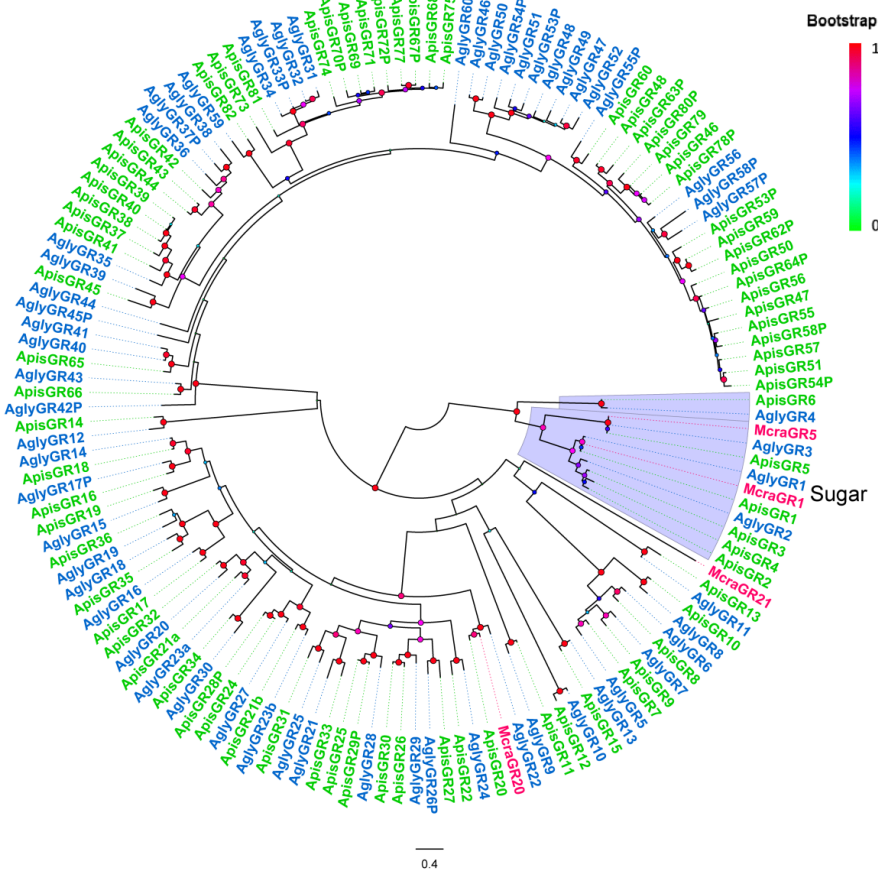


Figure 1. Phylogenetic tree of candidate gustatory receptors (GRs) in *Aphis glycines* (Agly), *Acyrthosiphon pisum* (Apis) and *Megoura crassicauda* (Mcra). The distance tree was rooted by the lineage of putative sugar receptors.

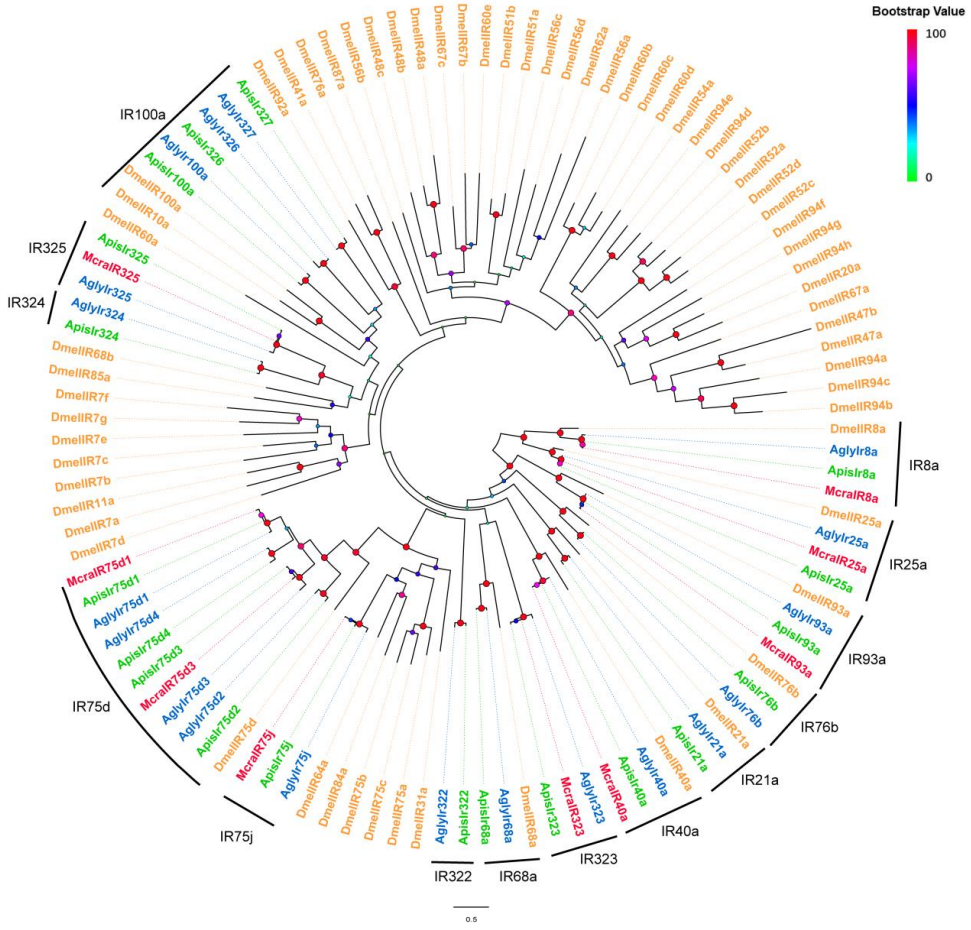


Figure 12. Phylogenetic tree of candidate ionotropic receptors (IRs) in *Aphis glycines* (Agly), *Acyrthosiphon pisum* (Apis), *Megoura crassicauda* (Mcra) and *Drosophila melanogaster* (Dmel). The distance tree was rooted by the IR8a and IR25a clade.

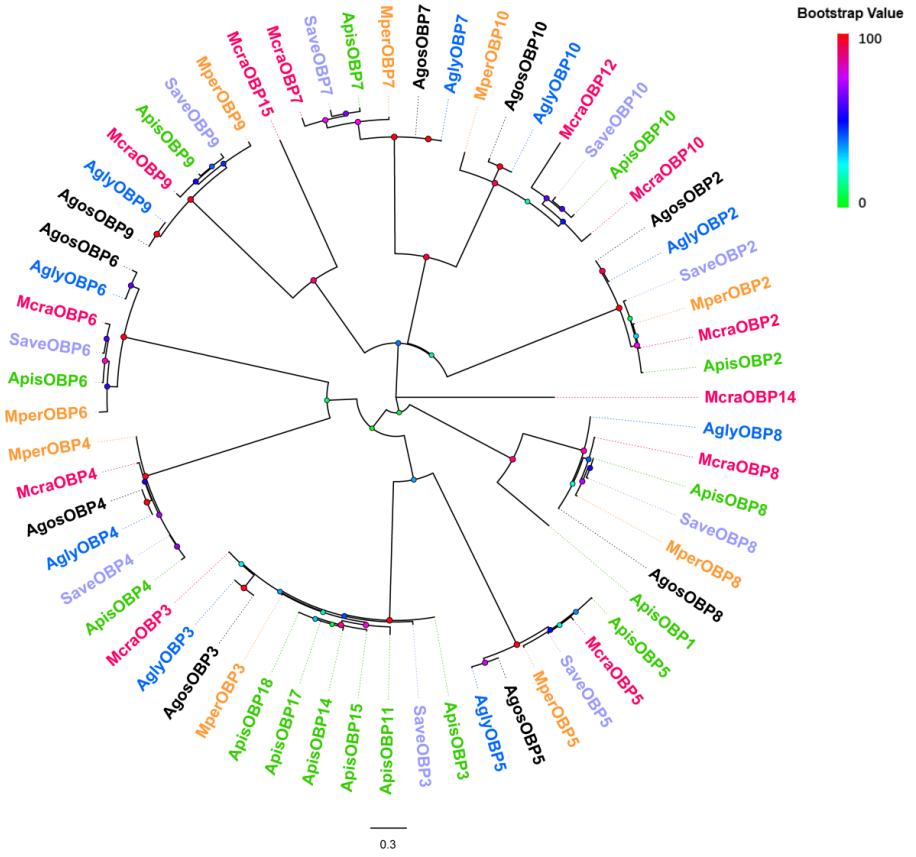


Figure 13. Phylogenetic tree of candidate odorant binding proteins (OBPs) in *Aphis glycines* (Agly), *Aphis gossypii* (Agos), *Acyrthosiphon pisum* (Apis), *Megoura crassicauda* (Mcra), *Myzus persicae* (Mper) and *Sitobion avenae* (Save).

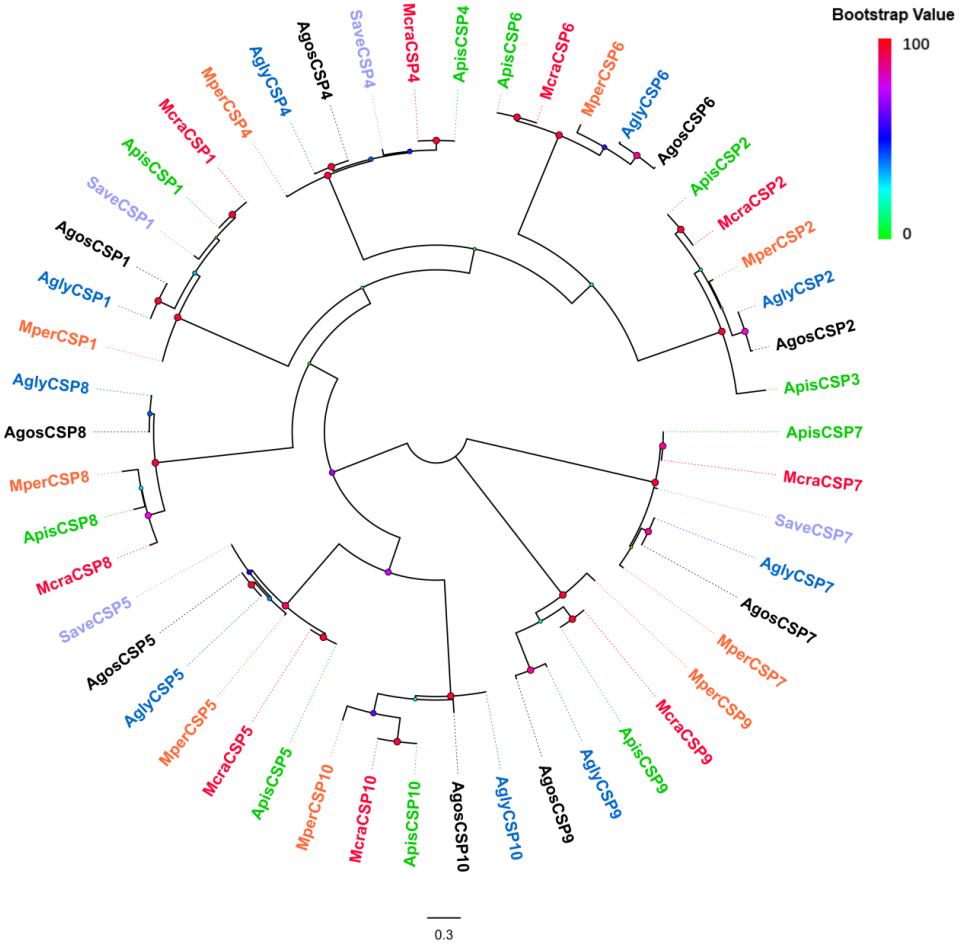


Figure 14. Phylogenetic tree of candidate chemosensory proteins (CSPs) in *Aphis glycines* (Agly), *Aphis gossypii* (Agos), *Acyrthosiphon pisum* (Apis), *Megoura crassicauda* (Mcra), *Myzus persicae* (Mper) and *Sitobion avenae* (Save).

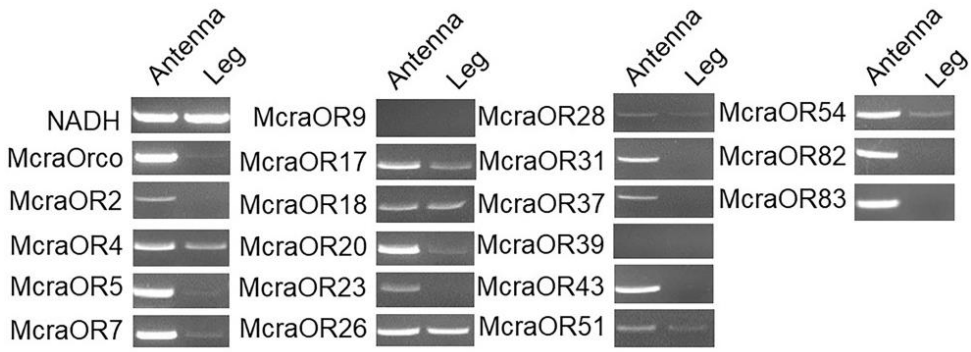


Figure 15. Tissue-specific expression levels of odorant receptor genes in *Megoura crassicauda* (Mcra).

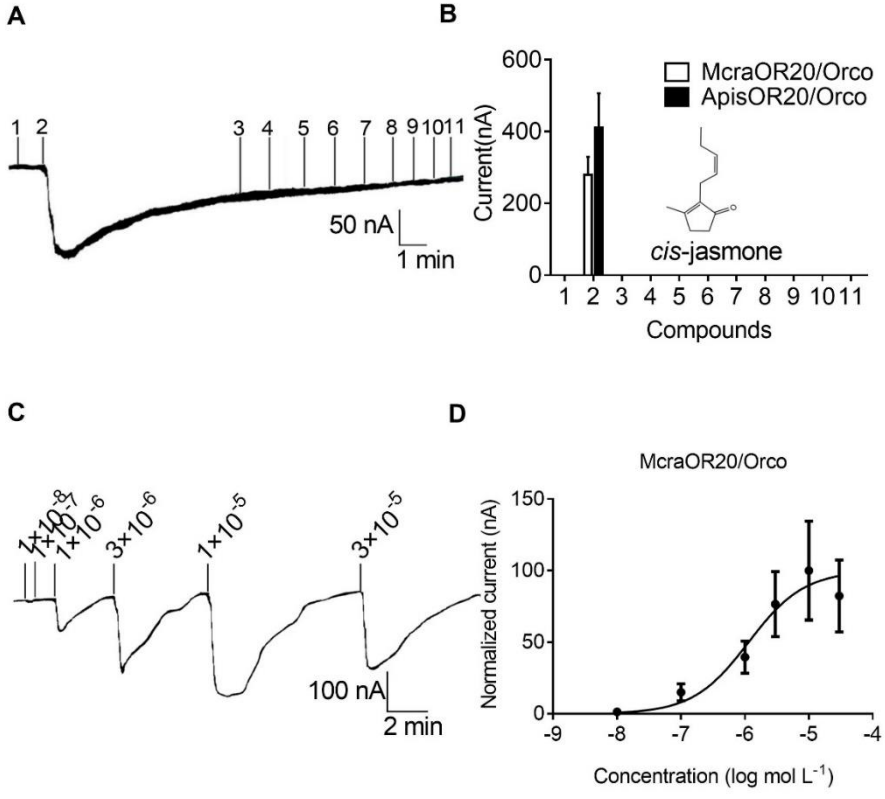


Figure 16. Functional characterization of McraOR20/Orco and ApisOR20/Orco co-expressed in *Xenopus* oocytes to 11 herbivore-induced plant volatile (HIPV) compounds. A, inward current responses of McraOR20/Orco to the tested HIPVs (1×10^{-4} mol L^{-1}). B, response profiles of McraOR20/Orco and ApisOR20/Orco co-expressed in *Xenopus* oocytes in response to the tested compounds. C, inward current dose-responses of McraOR20/Orco activated by *cis*-jasmone. D, dose-response curve of McraOR20/Orco to *cis*-jasmone. The tested HIPV compounds are as follows: 1, (-)-trans-caryophyllene; 2, *cis*-jasmone; 3, trans-2-hexenal; 4, ocimene; 5, linalool; 6, *cis*-3-hexen-1-ol; 7, trans-2-hexen-1-ol; 8, α-pinene; 9, (*E*)-4,8-dimethyl-1,3,7-nonatriene; 10, methyl salicylate; 11, (*E*, *E*)-4,8,12-trimethyl-1,3,7,11-tridecatetraene. Error bars indicate mean±SE (n=6).

4. Discussion

In the present study, we identified a total of 33 putative chemosensory receptor genes, including 20 ORs, four GRs, and nine IRs, in the antennal transcriptomes of *M. crassicauda*. However, a previous study showed that more chemosensory receptors were annotated from the genomes of *A. glycines* and *A. pisum* (Robertson et al., 2019). There are two possible reasons for the reduced number of chemosensory receptor genes obtained from the transcriptomes used in this study. First, transcripts with low expression levels are often only partially assembled, or even lost from the assembly. Studies on the sensilla morphology of aphids have presented a limited number of sensilla, also indicating that the types of sensilla in aphids are simple (Bruno et al., 2018; De Biasio et al., 2015), suggesting a relatively low abundance of these receptor genes in the aphid antennae. Second, chemosensory receptor genes exhibit tissue expression specificity. The paucity of GRs in antennal transcriptomes may be attributed to the specific expression of GRs in mouthparts (Chen, He, Li, Zhang, & He, 2016).

Phylogenetic analysis of the OR family shows that McraOR5 is ortholog to ApisOR5, which is reportedly responsible for alarm pheromone detection in *A. pisum* (Zhang et al., 2017). Amino acid sequence alignment also revealed a relatively high sequence identity (86.65%) between these two ORs, indicating that McraOR5 may also respond to EBF. We also found that McraOR4 clusters phylogenetically with another functionally investigated OR, ApisOR4 (Zhang et al., 2019), suggesting that they may share a similar odorant detection profile. GRs are responsible for detecting non-volatile compounds, such as sugars and caffeine (Dus, Min, Keene, Lee, & Suh, 2011; Moon, Kottgen, Jiao, Xu, & Montell, 2006). Two GRs (McraGR1 and McraGR5) clustered in the putative sugar receptor lineage inferred from a large-scale phylogenetic analysis of different insects (Robertson, 2015), which suggests that GR1 and GR5 may be responsible for sugar detection in *M. crassicauda*. Several studies in *Drosophila* showed that IR25a, IR40a, and IR93a are involved in the perception of humidity (Knecht et al., 2016), and the IRs in the IR75 clade are reportedly responsible for the detection of acids (Prieto-

Godino et al., 2017). Hence, the functions of aphid IRs can be inferred from their homologies with *Drosophila* IRs.

Twelve OBPs and nine CSPs were identified in the present transcriptomic analysis, which is comparable to the numbers of OBPs and CSPs found in other aphids, such as *A. glycines*, *A. gossypii* and *S. avenae* (Gu et al., 2013; Xue et al., 2016). The number of OBPs in *M. crassicauda* was less than that of *A. pisum* (Robertson et al. 2019). Full-length assemblies were obtained for all OBP and CSP transcripts, suggesting they may be abundantly expressed in the antennae. OBPs have been reported to be essential for olfactory reception of the alarm pheromone (Paolo Pelosi, Immacolata Iovinella, Jiao Zhu, Guirong Wang, & Francesca R. Dani, 2018). A series of studies also revealed that OBP3, OBP7, and OBP9 possess a high affinity to EBF (Fan et al., 2017; Northey et al., 2016; Qiao et al., 2009; Qin et al., 2020; Zhong et al., 2012). Phylogenetic analysis revealed that the OBP3, OBP7, and OBP9 clades are considerably conserved among six aphid species. Therefore, we hypothesize that McraOBP3, McraOBP7, and McraOBP9 may also be capable of binding EBF, consequently playing vital roles in EBF detection by *M. crassicauda*. Aside from the OBPs clustered in conserved clades, we also identified two OBPs, McraOBP14 and McraOBP15, which did not show ortholog relationships with other OBPs, suggesting they may be involved in the binding of volatiles that specifically exist in the niche of *M. crassicauda*.

cis-Jasmone is a natural volatile compound that is released when the plants are attacked by herbivores (Loughrin, Manukian, Heath, & Tumlinson, 1995; Rose & Tumlinson, 2004). It not only elicits plant defense responses (Matthes et al., 2010), but also serves as an attractant for predators of herbivorous insects (Powell & Pickett, 2003), therefore it plays an important role in the interactions between host plants, aphids and their natural enemies. Previous studies have shown that cis-jasmone could be detected by an olfactory cell located on the fifth antennal segment of *Nasonovia ribisnigri* (Birkett et al., 2000). Behavioral studies revealed that *cis*-jasmone correlated strongly with the repellence of numerous aphid species, including *M. euphorbiae* (Sobhy et al., 2017), *M. persicae* (Dewhirst et al., 2012), *N. ribisnigri* and *Phorodon humuli* (Birkett et al., 2000), and *S. avenae* (Bruce et al., 2003). These results suggest that at least one OR is involved in the detection of *cis*-jasmone, and this chemoreception process is important for producing the repellent response. In this study, we identified McraOR20 from the antennal transcriptomes of *M. crassicauda* and screened for its best ligands among 11 HIPVs which were reported to be released by plants challenged with aphid infestation. The functional characterization results indicate that *cis*-jasmone is the best ligand linked to McraOR20, showing that McraOR20 is vital for *cis*-jasmone detection. Additionally, we found a conserved clade (with 1:1:1 orthologs in three aphid species) of OR20 in the phylogenetic analysis. ORs in conserved clades may

undertake irreplaceable functions (Guo et al., 2020), and they may also share the same odorant binding pattern (Zhang et al. 2017). Our functional studies confirmed that McraOR20 and its ortholog in *A. pisum*, ApisOR20, are tuned to the same volatile, suggesting a conserved function for the OR20 clade in different aphid species. Therefore, we hypothesize that the homolog s of McraOR20 in other aphid species may also be responsible for detecting *cis*-jasmone.

In the present study, only McraOR20 was identified as the specific OR of *cis*-jasmone. However, we cannot exclude the possibility that other ORs may also be involved in *cis*-jasmone detection, and they may affect aphid behavior through certain combinatorial coding. Future work should investigate the *in vivo* functional characterization of McraOR20 using RNA interference (RNAi). Moreover, the specific HIPVs induced by *M. crassicauda* should be identified, allowing us to characterize the functions of species-specific ORs in *M. crassicauda*. We should also focus on studying the peripheral coding maps of aphids to *cis*-jasmone as well as other HIPVs, and combined this with behavioral experiments to reveal the mechanisms of HIPV olfactory recognition in aphids.

5. Conclusions

The putative chemosensory genes identified from the antennal transcriptomes of *M. crassicauda* will offer useful resources for future functional studies of McraORs, as well as genome and/or transcriptome annotation in other aphids and closely related species. Expression level analysis was performed by both bioinformatics and PCR-based methods, and helped in identifying some McraORs with antenna-biased expression in *M. crassicauda*. Furthermore, in vitro expressions of McraOR20 and ApisOR20, as well as McraOR43, in *Xenopus* oocytes were measured with the two-electrode voltage clamp technique revealing that McraOR20 and its ortholog ApisOR20 specifically respond to *cis*-jasmone, while McraOR43 did not show any response to the tested compounds. Our study not only lays the foundation for further functional studies of ORs in *M. crassicauda*, but also sheds light on the molecular mechanisms of *cis*-jasmone detection in aphids.

6. Acknowledgements

This work was funded by the National Natural Science Foundation of China (31572072 and 31801994), the Shenzhen Science and Technology Program, China (KQTD20180411143628272), the Natural Science Foundation of Tianjin, China (18JCYBJC96100), and the Tianjin Normal University Foundation, China (135305JF79).

7. Conflicts of Interest

The authors declare no conflicts of interest.

Chapter V

Chromosome-level genome assembly of the spotted alfalfa aphid, *Therioaphis trifolii*

Adapted from the following reference: Huang TY, He Kang, Liu Y, Francis F, Wang GR, Wang B. Chromosome-level genome assembly of the spotted alfalfa aphid, *Therioaphis trifolii*. (*Accepted.*)

Introduction to chapter V

Omics data, including genomic and transcriptomic data, are crucial resources for identifying evolutionarily conserved ORs, which are essential for functional studies in the context of chemosensory mechanisms. In the previous two chapters, we characterized two conserved ORs that tune to host plant volatiles and HIPVs, respectively. However, these ORs were selected based on the phylogenetic comparisons between Aphidinae species, which may have missed evolutionary information from other aphid subfamilies. Moreover, many non-Aphidinae species have distinct host ranges and alarm pheromone components compared to Aphidinae species, making them valuable target for studying the evolutionary process of ORs during the evolution of aphids. Currently chemosensory-related gene families have only been annotated from genomes of Aphidinae species, which greatly limits our understanding of the olfactory mechanisms of non-Aphidinae species. In this chapter, we report the genome assembly of *Therioaphis trifolii*, presenting the first genome assembly of the subfamily Calaphidinae.

Abstract: The spotted alfalfa aphid (SAA, *Therioaphis trifolii*) (Hemiptera: Aphididae) is a destructive pest of cultivated alfalfa (*Medicago sativa* L.) that leads to large financial losses in the livestock industry around the world. Here, we present a chromosome-scale genome assembly of *T. trifolii*, the first genome assembly for the aphid subfamily Calaphidinae. Using PacBio long-read sequencing, Illumina sequencing, and Hi-C scaffolding techniques, a 541.26 Mb genome was generated, with 90.01% of the assembly anchored into eight scaffolds, and the contig and scaffold N50 are 2.54 Mb and 44.77 Mb, respectively. BUSCO assessment showed a completeness score of 96.6%. A total of 13,684 protein-coding genes were predicted. The high-quality genome assembly of *T. trifolii* not only provides a genomic resource for the more complete analysis of aphid evolution, but also provides insights into the ecological adaptation and insecticide resistance of *T. trifolii*.

Keywords: *Therioaphis trifolii*, chromosome-level genome, comparative genomics, phylogenetics, genome synteny.

1. Introduction

Alfalfa (*Medicago sativa* L.), also called lucerne, is one of the world's most important cultivated fodder plants. It is cultivated in at least 80 countries, and because it is an abundant and stable source of nutrients, it has become the backbone of the global livestock industry (Bai et al., 2018; Frank et al., 2016; Radović, Sokolović, & Marković, 2009). The spotted alfalfa aphid (SAA), *Therioaphis trifolii*, is one of the most serious insect pests of legumes, mainly causing the wide-scale destruction of alfalfa crops (Blackman & Eastop, 2000). *T. trifolii* was first recorded in New Mexico in the United States of America (Dickson, Laird, & Pesho, 1955), and it also occurs in many regions of Australia, China, Europe, India, the Middle East, and the Mediterranean (Dickson et al., 1955; Lake, 1989; Wang et al., 2020). SAA damages its host plants by extracting nutrients from the leaves and phloem, and also by transmitting plant-pathogenic viruses, such as alfalfa mosaic virus and bean yellow mosaic virus (Jones, 2004), thereby severely restricting the growth of plants and causing devastating losses in alfalfa production (C. G. He & Zhang, 2006; Irwin, Lloyd, & Lowe, 2001).

The intensive use of chemical insecticides is the primary means of controlling aphids on many crops; however, this approach has become more challenging because aphids possess a great capacity to overcome multiple insecticides through the evolution of resistance (Bass et al., 2014; A. Chen, Zhang, Shan, Shi, & Gao, 2020; Lokeshwari, Krishna Kumar, & Manjunatha, 2016). Detoxifying enzymes contribute considerably to the development of insecticide resistance in aphids. For example, the peach potato aphid *Myzus persicae* is able to generate resistance to sulfoxaflor via overexpression of many detoxification-related enzymes, including UDP-glucuronosyltransferase (UGT) and cytochrome P450 (P450) enzymes (Pym et al., 2022), and *Aphis gossypii* overcomes sulfoxaflor through the up-regulation of ATP-binding cassette (ABC) transporter expression (Wang et al., 2021). One practical and sustainable strategy to reduce insecticide applications is the cultivation of aphid-resistant plants (Smith & Chuang, 2014). Many studies have mined for specific genes that can generate durable genetic resistance to *T. trifolii* in plants (Jacques et al., 2020; Kamphuis et al., 2013; Zhao et al., 2019); however, the molecular mechanisms by which *T. trifolii* responds to aphid-resistant alfalfa plants still remain unclear. Studies on other aphids have revealed that many digestive proteases may be involved in overcoming the defenses of aphid-resistant plants. For example, significant changes in the expression levels of various digestion-related genes, such as serine proteases (SPs) and carboxypeptidases (CPs), have been detected in *Aphis glycines* after feeding on resistant soybean (Bansal, Mian, Mittapalli, & Michel, 2014). The availability of a high-quality genome sequence will be considerably beneficial for gaining an improved understanding of the molecular mechanisms underlying SAA resistance to pesticides and aphid-resistant alfalfa.

Taking advantage of the feasibility of inexpensive sequencing, researchers have sequenced the genomes of many aphids (Chen et al., 2019; International Aphid Genomics, 2010; Jiang et al., 2019; Li et al., 2019; Mathers et al., 2021; Wenger et al., 2020), but the number of available genomes is still limited compared with the number of recorded aphids (more than 5000 species) (Emden & Harrington, 2017). In addition, most of the sequenced aphids belong to the subfamily Aphidinae, one large group

consisting of various important pests, and only a few efforts have focused on other subfamilies (Biello et al., 2021; Julca et al., 2020). The lack of genome sequences for other subfamilies has greatly limited our understanding of the genomic diversity and evolution of aphids. Calaphidinae is the second largest subfamily within the family Aphididae (Favret, 2013); it consists of nearly 400 valid species, some of which are notorious pests damaging a distinctive range of host plants (Herbert, Mizell, & McAuslane, 2009; Nebreda et al., 2004). However, despite its importance, no reference genome is yet available for this group.

Here, we present a high-quality chromosome-level genome assembly of *T. trifolii*, generated using a combination of PacBio, Illumina, and chromatin conformation capture (Hi-C) techniques. Phylogenetic analysis was performed to determine the relationship of SAA with other members of the superfamily Aphidoidea. Moreover, annotation and comparative analyses of digestion- and detoxification-related gene families and genome synteny analyses were carried out between *T. trifolii* and other representative aphid species. Our study provides the first genome assembly for a Calaphidinae aphid, which will facilitate studies on the genome evolution of aphids and also significantly benefit efforts to control this important alfalfa pest.

2. Methods and results

2.1. Sample preparation and genomic sequencing

An asexual, parthenogenetic *T. trifolii* colony was collected from the alfalfa fields at the Langfang Experimental Station of the Chinese Academy of Agricultural Sciences and reared on alfalfa (*Medicago sativa*) in natural light in a greenhouse maintained at 20±2℃ and relative humidity of 75%. To reduce the heterozygosity of the sequenced genome, one single parthenogenetic female was selected from the reared population to establish another colony, and one of its offspring was selected for generating the next colony. The selection was performed until we obtained the fifth generation of the aphid colony, which was used as the sample for all the genome sequencing experiments.

For PacBio sequencing, DNA was extracted from about 200 individuals, consisting of wingless parthenogenetic female adults and nymphs. Two single-end 20-kb libraries were constructed with the PacBio SMRT (Single-Molecule Sequencing in Real Time) system (Pacific Biosciences). Raw reads were generated from one cell sequenced on the PacBio Sequel II platform. After quality control filtering, 118.55 Gb (~220× coverage) of SMRT PacBio sequences were obtained, with a mean read length of 14.40 kb (N50 = 21.04 kb). For Illumina sequencing, about 200 wingless parthenogenetic female adults and nymphs were used for DNA extraction, and the library (400-bp inserts) was constructed using standard Illumina protocols and sequenced on the Illumina HiSeq X Ten platform, generating 33.73 Gb of data with 150bp paired-end reads. To further assemble the contigs into chromosomes, we generated a Hi-C library using protocols described in a previous study (Yang Liu et al., 2020). Fresh tissues from about 150 individual samples

(including adults and nymphs) were crosslinked with paraformaldehyde to obtain the interacting DNA segments. The cross-linked sample was digested with DpnII, and biotinylated nucleotides were used to label the ends of the restriction fragments. The library was quantified and sequenced on the Illumina Novaseq/MGI-2000 platform, and ~49.21 Gb of data with 150bp paired-end sequencing raw reads were generated.,

2.2. RNA sequencing

Total RNA was extracted from 100 adult parthenogenetic females using TRIzol reagent (Invitrogen, Carlsbad, CA, USA) (Huang et al., 2022) and dissolved in RNase-free water. The integrity of the RNA was assessed by 1% agarose gel electrophoresis. RNA purity and concentration were assessed using a Nanodrop ND-2000 spectrophotometer (ThermoFisher, USA). The qualified RNA was used for constructing cDNA libraries. Raw sequencing data were generated using an Illumina NovaSeq 6000 platform (Illumina, San Diego, CA, USA) with the 200 bp paired-end strategy.

2.3. Genome assembly

The quality control of raw Illumina reads was carried out using FASTP v0.20.0 (S. Chen, Zhou, Chen, & Gu, 2018). Clean reads were used to construct a 17-mer frequency distribution map using JELLYFISH v2.3.0 (Marçais & Kingsford, 2011). The genome size of *T. trifolii* was estimated to be 542.4 Mb based on k-mer analysis.

For contig assembly, we first used FALCON v1.8.7 (reads_cutoff: 1k, seed_cutoff: 33k) (Chin et al., 2016) for the error correction of PacBio reads. The corrected reads were assembled into the preliminary genome assembly using SMARTDENOVO v1.0 with parameters -J 3000 and -k 19 (H. Liu, Wu, Li, & Ruan, 2021). To correct errors generated during the assembly process, PacBio reads were mapped to the genome using BLASTR v5.1 (Chaisson & Tesler, 2012), and ARROW v2.2.2 was used for one round of genome polishing with default parameters. Illumina reads were also mapped to the assembly using BWA v0.7.12 (Li & Durbin, 2009), and then four iterations of contig polishing were carried out using NEXTPOLISH v1.0.5 with default parameters (Hu, Fan, Sun, & Liu, 2020). A contig-level assembly with a total length of 541.26 Mb was generated, which is comparable to the estimated genome size, and the contig N50 length was 2.54 Mb (Table 4).

2.4. Hi-C scaffolding

To further assemble the contigs into chromosomes, we generated a Hi-C library using protocols described in a previous study (Yang Liu et al., 2020). Fresh tissues from about 150 individual samples (including adults and nymphs) were crosslinked with paraformaldehyde to obtain the interacting DNA segments. The cross-linked sample was digested with *Dpn*II, and biotinylated nucleotides were used to label the ends of the restriction fragments. The library was quantified and sequenced on the Illumina Novaseq/MGI-2000 platform, and ~49.21 Gb of data with 150bp

paired-end sequencing raw reads were generated. Low-quality raw reads (quality score <20 and shorter than 30 bp) and adaptors

Table 4. Major indicators of the *Therioaphis trifolii* genome

Features	Statistics
Estimated genome size (bp)	542,395,090
Assembly size (bp)	541,263,359
Contigs N50 (bp)	2,544,558
Scaffolds number	575
Scaffolds N50 (bp)	44,770,504
BUSCO genes	C: 96.6% [S: 93.5%, D: 3.1%], F: 0.7%
Number of protein-coding genes	13,684

were removed using FASTP v0.20.0, then the clean reads were mapped to the contig assembly using BOWTIE2 v2.3.2 (-end-to-end --very-sensitive -L 30) (Langmead & Salzberg, 2012). HI-C PRO v2.8.1 (Servant et al., 2015) was used to identify valid interaction paired reads and to filter out reads with multiple hits and singleton reads. LACHESIS (Burton et al., 2013) was used to cluster, order, and orient the contigs with parameters CLUSTER MIN RE SITES = 100; CLUSTER MAX LINK DENSITY = 2.5; CLUSTER NONINFORMATIVE RATIO = 1.4; ORDER MIN N RES IN TRUNK = 60; ORDER MIN N RES IN SHREDS = 60.

As a result, Hi-C data were combined with the contig-level assembly to generate a chromosome-level assembly comprising eight large scaffolds (Figure 17a), which corresponds to the previously reported haploid chromosome number for this species (Sunnucks et al., 1997). Around 90.07% of the contigs were anchored onto

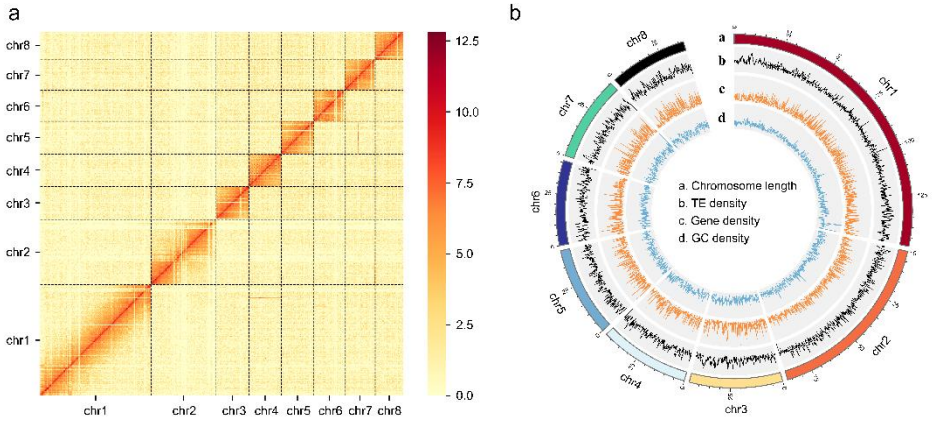


Figure 17. Heatmap of genome-wide Hi-C data and circular representation of the chromosomes of *Therioaphis trifolii*. (a) The heatmap of chromosome interactions in *T. trifolii*. The frequency of Hi-C interaction links is represented by colours, which ranges from white (low) to red (high). (b) Circos plot of distribution of the genomic elements in *T. trifolii*. The tracks indicate (moving inwards): **a** length of the chromosome, **b** distribution of transposable element (TE) density, **c** gene density, and **d** GC density; the densities of TEs, genes, and GC were calculated in 100 kb windows.

chromosomes, resulting in a scaffold N50 length of 44.77 Mb (Table 3). The longest chromosome was 149.16 Mb while the shortest was 37.54 Mb (Figure 17b).

2.5. Repeat annotation

TANDEM REPEAT FINDER v4.07b (parameters: 2 7 7 80 10 50 500 -f -d -h -r) (Benson, 1999) was used to identify all tandem repeat elements. Transposable elements (TEs) were identified using a combination of two methods. First, a *de novo* repeat library was generated using REPEATMODELER v1.0.11 and MITE-hunter (Y. Han & Wessler, 2010) with default parameters. This library was searched against the Repbase (Bao, Kojima, & Kohany, 2015) to classify repeat families using REPEATMASKER v1.331, and then merged with Repbase to generate the final repeat sequence library. Next, REPEATMASKER v1.331 was used to predict TEs based on the final TE library. The result showed that repeat sequences make up 36.86% of the genome, most of which are TEs (33.31%) (Table 5).

Table 5. Statistics of the repeat elements in *Therioaphis trifolii* genome

Repeat types	Number	Length occupied (bp)	Percentage of sequence
SINE	5,290	506,439	0.09%
LINE	135,921	27,613,234	5.10%
LTR	73,841	14,343,500	2.65%
MITE	53,954	13,738,880	2.54%
DNA	716,470	121,720,828	22.49%
Unknown	45,373	7,137,015	1.32%
Total base masked	1,236,028	199,531,663	36.86%

2.6. Protein coding gene prediction and functional annotation

Gene model prediction from the TE soft-masked *T. trifolii* genome was performed using multiple approaches, namely transcriptome-based prediction, ab initio prediction, and homology-based gene prediction. For transcriptome-based analysis, clean reads were aligned to the genome assembly using STAR v2.7.3a (Dobin et al., 2013) with the default parameters. Next, STRINGTIE v1.3.4d (Kovaka et al., 2019) was used to obtain transcript locations, and open reading frames of the transcripts were predicted using PASA v2.3.3 (Haas et al., 2008). For *de novo* gene model prediction, the transcript set generated by PASA was utilized by GENEMARK-ST v5.1 (Tang, Lomsadze, & Borodovsky, 2015) for self-training. The training set was applied to AUGUSTUS v3.3.1 (Stanke, Diekhans, Baertsch, & Haussler, 2008) for gene model prediction. For the homology-based gene modeling process, protein sets of several aphids with high-quality genome assemblies were aligned to the genome assembly using GEMOMA v1.6.1 (Keilwagen et al., 2016). Finally, we combined the results from the three gene prediction approaches to create a consensus gene model set using EVIDENCEMODELER v1.1.1 (--segmentSize 1000000 --overlapSize 100000) (Haas et al., 2008). As a result, 13,684 protein-coding gene models were generated, with an average gene length of 15 kb, average coding sequence length of 1.5 kb, and average exon number of 7.1.

For gene functional annotation, protein sequences encoded by the predicted gene models were aligned to the non-redundant (nr), SWISS-PROT, Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa & Goto, 2000), and eukaryotic orthologous groups (KOG) databases (Galperin, Makarova, Wolf, & Koonin, 2015) using BLASTP v2.7.1 with a cutoff of 1e-5. We also used INTERPROSCAN v5.32-71.0 (Zdobnov & Apweiler, 2001) to obtain gene ontology (GO) annotations for the proteins.

2.7. Phylogenetic and comparative genomic analyses

The longest predicted protein sequences of 12 aphid genomes, namely *Aphis glycines* (Mathers, 2020), *Acyrthosiphon pisum* Mathers et al., 2021), *Cinara cedri* (Julca et al., 2020), *Diuraphis noxia* (Nicholson et al., 2015), *Eriosoma lanigerum* (Biello et al., 2021), *Myzus cerasi* (Thorpe, Escudero-Martinez, Cock, Eves-van den Akker, & Bos, 2018), *M. perisicae* (Mathers et al., 2021), *Pentalonia nigronervosa* (Mathers, Mugford, Hogenhout, & Tripathi, 2020), *Rhopalosiphum maidis* (Chen et al., 2019), *R. padi* (Thorpe et al., 2018), *Sitobion miscanthi* (Jiang et al., 2019), and *T. trifolii*, and the greenhouse whitefly *Trialeurodes vaporariorum* (Xie, He, Fei, & Zhang, 2020), which was used as an outgroup, were utilized for identifying orthologous groups among aphids using ORTHOFINDER v2.4.0 61. A total of 2758 single-copy orthogroups were identified and used to generate a concatenated alignment for inferring phylogenetic relationships. The species tree of the 12 aphids was also inferred using ORTHOFINDER (Emms & Kelly, 2019) and rooted by STRIDE (Emms & Kelly, 2017). Divergence times among aphids were calculated by R8S (Sanderson, 2003) based on divergence information extracted from TimeTree (http://www.timetree.org/): *A. pisum* vs *M. persicae* 42.5–

48.0 million years ago (MYA) (Figure 18). We also used CAFE v4.2.1 (De Bie, Cristianini, Demuth, & Hahn, 2006) to analyze the expansion and contraction of gene families in all 12 tested aphid lineages. The results from the phylogenetic tree with divergence times were used as inputs (Figure 18).

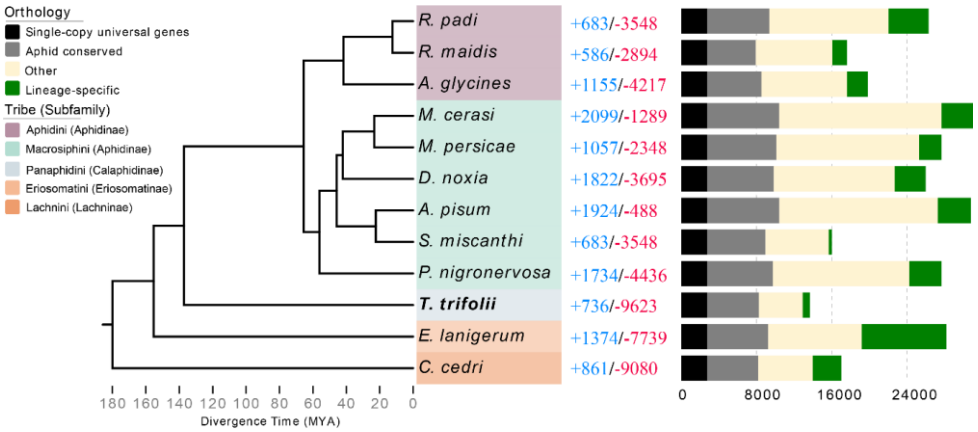


Figure 18. Phylogeny and orthology analyses between *Therioaphis trifolii* and other aphid species. The phylogenetic tree was constructed based on 2,758 single-copy orthogroups obtained from the genomes of all tested aphids. The greenhouse whitefly *Trialeurodes vaporariorum* (not shown) was selected as the outgroup. Aphid species are colored according to their tribe. Numbers of expanded (blue) and contracted (red) gene families are presented next to the species and nodes. Comparison of orthologs between 12 aphids. 'Single-copy universal' indicates a single-copy ortholog is present in all the aphids. 'Aphid conserved' indicates genes that can be detected in at least 11 aphid genomes. 'Lineage-specific' indicates genes without an ortholog in any other aphid. 'Other' indicates orthologs found in some of the aphids (e.g., in 1 to 10 aphids).

2.8. Synteny analysis

The synteny analysis were carried out between the chromosome-level genome assemblies of *T. trifolii*, *A. pisum* (JIC1 v1), and *E. lanigerum*. To obtain syntenic blocks, we uploaded the official gene sets to the ORTHOVENN2 server (Xu et al., 2019). The 1:1 single-copy ortholog pairs from each comparison (*T. trifolii* vs *A. pisum* and *T. trifolii* vs *E. lanigerum*) were identified using the parameters e-value = 1e-5 and inflation value = 1.5. These gene pairs were selected for genome synteny analyses using MCSCANX v1.1 (Y. Wang et al., 2012) with default parameters. SYNVISIO (https://synvisio.github.io) was used to visualize genome synteny (Figure 19).

3. Data records

The genome sequencing, RNA sequencing reads data has been updated to the National Center for Biotechnology Information (NCBI) as a BioProject no.

PRJNA804007. Pacbio, Hi-C, Illumina and transcriptome sequencing reads have been deposited in the Sequence Read Archive (SRA) databases with the accession number of SRP35901569. Genome assembly has been deposited at the NCBI, under the accession number of JALBXZ00000000070, and can be download from National Genomic Data Center (NGDC) under accession number GWHBQDZ00000000. The annotated detoxification and digestion related genes among aphids have been uploaded to the NGDC under accession number OMIX002672. Gene sequences predicted from the genome assembly are also publicly available in NGDC, under the accession number OMIX002673. All data in NGDC were related to the BioProject PRJCA014018.

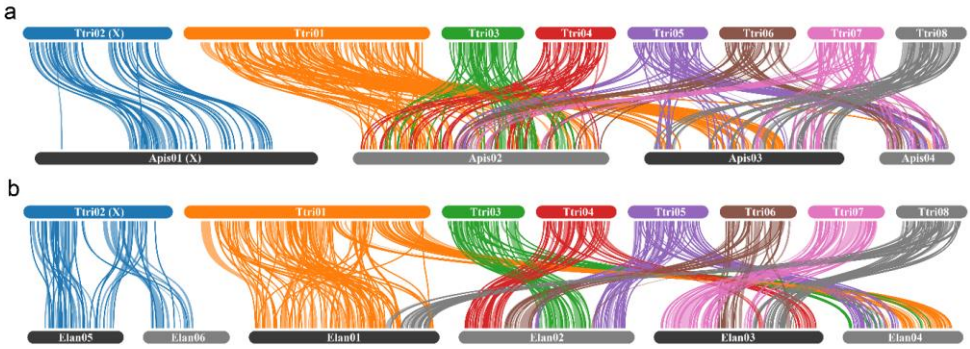


Figure 19. Genome synteny between (a) *Therioaphis trifolii* (Calaphidinae) and *Acythosiphon pisum* (Aphidinae) and (b) *T. trifolii* and *Eriosoma lanigerum* (Eriosomatinae). Links indicate the edges of syntenic blocks of gene pairs identified by synteny analyses and are shown in the same color as that of the chromosome ID of *T. trifolii*. Ttri indicates *T. trifolii*, Apis indicates *A. pisum*, and Elan indicates *E. lanigerum*.

4. Technical validation

The accuracy and completeness of the contig assembly were validated using three methods. First, clean Illumina reads were mapped to the contigs assembled by BWA v0.7.12, and the total mapped reads and mapping rate were calculated using SAMTOOLS v1.4 (Li et al., 2009), resulting in a mapping rate of 99.40%. Second, Benchmarking Universal Single-Copy Orthologs (BUSCO) v4.0.5 (Simão, Waterhouse, Ioannidis, Kriventseva, & Zdobnov, 2015) was employed to assess the completeness of the genome assembly based on the insecta_odb10 database (-l insecta_odb10 -g genome), the BUSCO analysis indicated that 97.3% of gene orthologs were identified in *T. trifolii*, including complete and fragment scores of 96.6% and 0.7%, respectively. Finally, CEGMA v2 (Parra, Bradnam, & Korf, 2007) with default parameters was used to validate the integrity of the core genes in the assembly, 242 core eukaryotic genes were assembled, among which 94.76% were complete. Among the predicted gene models, 12,995 (94.96%) possessed significant homology to proteins from at least one of the following databases: nr, SWISS-PROT, GO, KOG, and KEGG.

5. Code availability

All software and pipelines used for data processing were executed according to the manuals and protocols of the bioinformatics software cited above, and the parameters are clearly described in the Methods section. If no detailed parameters are mentioned for a software, the default parameters were used. The version of the software has been described in Methods.

6. Acknowledgements

This study was supported by a grant from Shenzhen Science and Technology Program (Grant No. KQTD20180411143628272) and the National Natural Science Foundation of China (31872039), and the Science and Technology Innovation Program of the Chinese Academy of Agricultural Sciences.

7. Author contributions

B.W., G.W. and F.F. conceived this study. T.H. prepared DNA and RNA for sequencing. T.H. performed the experiments and analyzed the data. T.H., K.H., Y.L., B.W., F.F., and G.W. wrote the manuscript. All authors reviewed the manuscript.

Chapter VI

Evolution and functional characterization of aphid ORs in respond to EBF

Adapted from the following reference: Huang TY, Yang LL, Wang B, Liu WB, Bai PH, Francis F, Wang GR, Wang B. Evolution and characterization of aphid ORs in respond to EBF. (*In prep*)

Introduction to chapter VI

The most intriguing aspect of aphids is their avoidance activity in response to the alarm pheromone released by conspecies. This behavior increases the survival rate of aphid populations and is a crucial adaption against predators. Despite extensive research on the interactions between alarm pheromone and aphids, the underlying mechanisms of how aphid sense this vital chemical cue remain poorly understood. In aphids, the complex chemosensory system is responsible for the reception of EBF, the primary component of alarm pheromone in various aphids. Although odorant-binding proteins (OBPs) responsible for transporting EBF in aphid antennae have been characterized, the ORs responsible for detecting and mediating avoidance behavior towards EBF have received little attention. Furthermore, previous studies have suggested the potential divergence of alarm pheromone among different aphid species, suggesting that EBF receptors may have evolved in distinct patterns. Additionally, previous research on EBF receptors has mainly focused on Aphidinae species, which limits our understanding of how EBF receptors have evolved in other aphid subfamilies. In this chapter, we present a systematic annotation of the OR gene family among a broad range of aphids from different subfamilies, ultilizing the first genome assembly of the Calaphidinae aphid and other publicly available genome assemblies. We also attempted to characterize the potential EBF receptors using various functional approaches.

Abstract: Alarm pheromone-inducing avoidance behavior is an essential adaptation for insects to reduce their predator risk. Perception of alarm pheromone is accomplished by insect chemosensory system and variation in the alarm cue composition may lead to evolutionary changes in olfaction mechanisms. Aphid is ideal model for addressing this question because alarm pheromone composition varies among different aphids. For example, EBF is the main or only component in a specific aphid subfamily, Aphidinae, which suggests that Aphidinae aphids may possess a unique chemoreception mechanism for detecting EBF. In this study, we show that perception and mediation of repellent behavior towards EBF are involved in two ORs specifically evolved among Aphidinae species. We annotated and compared the OR repertoires from genome assemblies of 13 aphid species (8 from Aphidinae subfamily and 5 from non-Aphidinae subfamilies) followed by the identification of 8 single-copy ORs that were conserved in amino acid sequences and under strong purifying selection among 8 Aphidinae aphids. Using a combination of two-electrode voltage-clamp (TEVC) technique, RNA interference, and behavioral experiments, we characterized two ORs (OR5 and OR43) as the EBF receptors, which are essential for mediating repellent behavior in aphids.

Keywords: (*E*)-β-farnesene, odorant receptor, evolution, Aphidinae, alarm pheromone.

1. Introduction

Alarm signal is crucial required communication for adaptation and fitness in many organisms (Wyatt, 2003). In insects, predator risk is one of the major factors that threaten their survival. To protect themselves from predation, insects evolve a great diversity of alarm pheromones to warn conspecific individuals from coming danger (Basu et al., 2021; Nouvian, Reinhard, & Giurfa, 2016).The vital essential role of the alarm pheromone in terms of adaptation evolution and chemical ecology has received initial attention (Harraca, Ryne, & Ignell, 2010; Norman, Butterfield, Drijfhout, Tasman, & Hughes, 2017). Most alarm pheromones appear to function as repellents, which are released into the air by alarmed workers (Fox & Adams, 2022). Besides, more attention has been given to aphids, which are sap-sucking Hemipteran insects (Blackman & Eastop, 2000; Dixon, 1998; Vandermoten et al., 2012). When attacked by natural enemies, aphids secret waxy droplets from the cornicle, the specialized organ located on their abdominal segment (Michaud, 2022). These secretions contain the alarm pheromone that alerts nearby aphids to walk away or drop off the host plants, consequently enhancing the survival rate of the colony.

The first alarm pheromone to be identified in aphids is the sesquiterpene (*E*)-β-farnesene (EBF) (Bowers et al., 1972) and it has been characterized later as the sole alarm component in numerous aphids (Bayendi Loudit, Boullis, Verheggen, & Francis, 2018; Francis et al., 2005; Pickett & Griffiths, 1980), many of which are world-widely destructive pests, such as the pea aphid *A. pisum* (Bowers, et al. 1972), the green peach aphid *M. persicae* (de Vos, Cheng, Summers, Raguso, & Jander, 2010; Edwards et al., 1973), the cotton aphid, *Aphis gossypii* (Bowers, et al. 1972). Also, few repellent chemicals have also been identified as alarm compounds in some aphid species. For example, sesquiterpene germacrene A was found as an alarm pheromone component in the genus *Therioaphis* (Bowers et al., 1977; Nishino et al., 1977). Monoterpenes, such as α-pinene and β-pinene, are the major active alarm components in the vetch aphid, *Megoura viciae* (Pickett & Griffiths, 1980; Song et al., 2021). These observations suggested a divergence of alarm pheromone components in aphids. Although the biological and ecological effects of common component EBF on aphids has been extensively studied (Beale et al., 2006; Wang et al., 2019), the evolutionary relationship of alarm pheromone divergence remains unrevealed and received a little attention. Understanding of the molecular mechanism of coevolution between the alarm pheromones and their detections in aphid interspecies is still limited.

The perception of odorants in insect mainly relies on odorant receptors (ORs), expressed on the membrane of peripheral olfactory sensory neuron, which is capable of translating sensory stimuli into electrical signals, thereby initiating a rich repertoire of behavioral responses (Benton, 2006; Brand et al., 2018; Fleischer et al., 2018). The evolution of the OR gene family follows a birth-and-death process over a long-term period (Almeida, Sánchez-Gracia, Campos, & Rozas, 2014; Nozawa & Nei, 2007). Both duplication (birth) and pseudogenization (death) of ORs could facilitate the adaptation or acquisition of new phenotypes in insects. For

instance, the functional loss of ORs resulting from deletion or pseudogenization is associated with host switch in the herbivorous fly *Scaptomyza flava* (Goldman-Huertas et al., 2015). Whereas gene family expansion of OR, such as the extensively amplified '9-exon' OR subfamily in eusocial insects, is usually under positive selection pressure. Indeed, function of '9-exon' ORs are involved in the detection of cuticular hydrocarbons, which is extremely critical essential for mediating their social behaviors (Pask et al., 2017; Slone et al., 2017).

Many ORs are highly conserved between closely related species in evolutionary process (Keesey et al., 2022; Roberts et al., 2022; Wang et al., 2017). Functional studies have revealed that conserved ORs are usually tuned to volatile chemicals with ecological importance to insects. For instance, HarmOR42 orthologs of Lepidoptera families in insects are essential for sensing floral scents (Guo et al., 2020). Besides, our previous study also revealed that orthologs OR5 across several aphid species, such as *A. pisum* and *A. gossypii*, are responsible for the detection of EBF (Zhang et al., 2017). Given that EBF is the major or only alarm component in numerous aphids, we hypothesized that EBF may be detected by evolutionarily and functionally conserved ORs, which may not limit to OR5. However, the knowledge about the evolutionary mechanism of EBF-receptors in aphid species is very limited. Recent efforts to sequence aphid genomes have provided abundant genomic resources, but mostly converge on the subfamily Aphididnae, which is significantly restricted the study of how OR genes evolved in aphid families.

In this study, we aimed to answer the following questions. (1) Which ORs are tuned to EBF, and are these ORs functionally conserved in aphids? (2) Is there any special mechanism for Aphidinae aphids to sensory EBF? (3) What is the possibly origin of EBF receptors? To address these questions, we identified the alarm pheromones of three aphids from distinct taxonomic ancient lineages and performed *in vivo* functional study assays to test the antennal and neuronal responses to EBF in these three aphid species. Further, we carried out manual gene model annotations in a comprehensive aphid genomic data set, which allow us to build the phylogeny and reveal the evolutionary dynamic of aphid OR gene family, and subsequently identified a set of OR clades with conserved sequences and under strong purify selection. Combining two-electrode voltage-clamp (TEVC) technique and RNA interference (RNAi) experiments technique, we characterized two odorant receptors that are both essential for mediating behavioral repellence to EBF. We also functionally characterized an OR5 ortholog in the non-Aphidinae aphid *Cinara cedri* (Lachninae), this species has provided a possible model of the origination of EBF receptor.

2. Material and methods

2.1. Extraction of aphid alarm pheromones and GC-MS analysis

Ten adults of *C. cedri*, and twenty adults of *E. lanigerum* or *T. trifolii*, respectively,

were submerged into a brown sample bottle containing 45 mL hexane and 5 mL of a 0.75 ng/mL solution of heptyl acetate as the standard. The bottle was immediately swirled for 30 s and kept at room temperature for 30 min. The extracts were stored at -20 ℃ with all aphids removed. The odor sample was injected in an Rtx-5MS column (30 m × 0.25 mm × 0.25 μm, Shimadzu, Tokyo, Japan), fitted in a GC-MS QP2020 (Shimadzu, Tokyo, Japan). The GC oven temperature was held at 50 ℃ for 3 min, increased gradually to 190 ℃ at 15 ℃/min, and then 15 °C/min to 240 °C. The inlet was held at 250°C and applied a helium flow rate of 1mL/min. The mass spectra was manipulated at 70 eV, and mass scanning was operated ranged from 35 to 500 amu at 0.2 scans/sec. The structures of the alarm pheromones were confirmed by that of authentic compounds measured under the same conditions.

2.2. Single sensillum recordings

The body and antennae of adult aphids were stuck to a coverslip with double-face adhesive tape. The electrophysiological recordings were performed on the large and small placoid sensilla of adult aphids. The extracellular signals produced by the odorant receptor neurons (ORNs) were recorded through inserting a tungsten wire electrode into the base of a sensillum, and a reference electrode was inserted in the eye. The recorded signals were amplified 10 × by a preamplifier (IDAC-4 USB System, Syntech, Kirchzarten, Ger many), and filtered with a 500 Hz low cutoff and a 3 kHz high cutoff. The signals were processed by the software package Autospike 32 (Syntech) and visualized on a computer screen. Neuronal responses were recorded for 10 s, starting 1 s before a stimulation period of 0.3 s. All odorants were stocked at the concentration of 100 mg/mL in a paraffin oil solution, and presented in a humidified continuous airflow generated by a stimulus controller (CS-55, Syntech, Kirchzarten, Germany) at 1.4 L/min. Each stimulus pulse was maintained for 300 ms.

2.3. Genome-wide OR gene annotation

OR gene models were annotated from the genome assemblies of 13 aphids: *A. gossypii* (Zhang et al., 2022), *A. glycines* (Mathers, 2020), *R. padi* (Thorpe et al., 2018), *R. maidis* (Chen et al., 2019), *S. miscanthi* (Jiang et al., 2019), *A. pisum*, *M. persicae* (Mathers et al., 2021), *P. nigronervosa* (Mathers et al., 2020), *S. flava* (NCBI database), *T. trifolii* (this study), *C. cedri* (Julca et al., 2020), *E. lanigerum* (Biello et al., 2021), *H. cornu* (Korgaonkar et al., 2021). OR gene models of *A. pisum* and *A. glycines* were updated based on recently published genome assemblies, and gene models newly annotated from this study were added to the former OR sets. OR5 and OR43 homologs were annotated from the genome assemblies of *Chaitophorus viminalis*, *Stegophylla sp.*, *Geopemphigus sp.*, *Pemphigus obesinymphae* (T. E. Smith, Li, Perreau, & Moran, 2022). OR sequences of *A. pisum* were selected as queries because these genes were finely corrected with the support of RNAseq data (Robertson et al., 2019). Amino acid sequences of ApisORs were searched against the aphid genomes using TBLASTN with an e-value cutoff of 10^{-5}. The matched regions extracted from high-scoring

segments were used for inferring possible OR coding regions. The putative coding regions were chained according to the similar criteria applied by (Legan, Jernigan, Miller, Fuchs, & Sheehan, 2021): closely organized regions were considered as a whole if the query of the upstream region was the N-terminal to that of the downstream region. The boundaries between exon and intron were manually checked based on the evidence from CDS sequences generated by automated approach and TBLASTN homology. Newly annotated ORs were added into the query set and used for next round TBLASTN until no novel OR gene model was found. All ORs annotated from this pipeline were named after their order in the genome, except that of *A. pisum* and *A. glycines*, in which newly identified ORs were named following the existing ORs. OR gene was considered as a pseudogene if a frame-shift mutation or a premature stop codon was detected in its coding region, and was considered as a partial model if the gene model was incomplete but without any frame-shift mutation or premature stop codon. Transmembrane domains of OR gene models were predicted by combining a consensus method-based software TOPCONS (v.2.0) (Tsirigos, Peters, Shu, Käll, & Elofsson, 2015), and a deep learning protein language model-based algorithm DeepTMHMM (v.1.0.13) (Hallgren et al., 2022). OR genes and pseudogenes were identified to organize in a tandem array if they were within 10kb of each other.

2.4. Phylogenetic and sequence analysis

For construction of OR phylogenetic tree, amino acid sequences were aligned using MAFFT (v.7.305) (Katoh & Standley, 2013) with default parameters. Poorly matched regions in the alignments were trimmed by trimAl (v.1.2) (Capella-Gutierrez et al., 2009). Phylogenetic trees were built using IQ-TREE (v.2.1.3) (Nguyen, Schmidt, von Haeseler, & Minh, 2015). Branch supports of the phylogenetic trees were estimated using UFBoot2 (Hoang, Chernomor, von Haeseler, Minh, & Vinh, 2018) implemented in the IQ-TREE package with 5,000 replicates. For construction of aphid species tree, the longest predicted protein sequences of the aphid genomes, and that of *A. lucorum* (Yang Liu et al., 2020), which was used as an outgroup, were used for constructing aphid species phylogeny by OrthoFinder (v.2.5.2) (Emms & Kelly, 2019). Divergence times were inferred using r8s (v1.81) (Sanderson, 2003) based on divergence information extracted from TimeTree (http://www.timetree.org/): *A. pisum* vs *M. persicae* 42.5–48.0 million years ago (mya). The phylogeny and gene orthology results were visualized in EVOLVIEW v2 (He et al., 2016). Gene gain and loss events were estimated by reconciling intact OR gene tree and species tree of aphids using NOTUNG (v.3.0 BETA) (K. Chen, Durand, & Farach-Colton, 2000). The nonsynonymous (dN) to synonymous (dS) substitution rate (ω) of each conserved branch was estimated using FitMG94.bf model implemented in HyPhy software (v.2.5.25) (Kosakovsky Pond et al., 2020). The pseudogenization of SflaOR1P was verified by mapping Illumina reads onto the genome assembly using HISAT2 (v.2.2.1) (D. Kim, Paggi, Park, Bennett, & Salzberg, 2019) and visualized by Integrative Genomics Viewer (Thorvaldsdóttir, Robinson, & Mesirov, 2013).

2.5. Vector construction, cRNA synthesis, and two-electrode voltage clamp recording

Full-length coding sequences of ApisOR2, 3, 4, 5, 20, 39, 40, 43, and Orco identified from our annotation analyses were cloned according to (Huang et al., 2022). Coding sequences of these ORs were amplified by PCR from *A. pisum* antennae cDNA. The expression vectors pT7Ts containing the open reading frames of the eight ORs and *ApisOrco* were used for subclone. Primers for gene cloning and expression vector construction are listed. cRNAs produced by linearized expression vectors were synthesized using the mMESSAGE mMACHINE T7 Ultra Kit (Ambion) following the manufacturer's protocol. A mixture containing 27.6 ng of ApisOR cRNA and 27.6 ng of ApisOrco was injected (Nanoliter 2010, WPI Inc., Sarasota, FL) into mature healthy oocytes, then incubated at 18 ℃ for 4-5 days in a nutrient solution. A two-electrode voltage clamp (OC-725C oocyte clamp, Warner Instruments, Hamden, CT) was used to test the response curves of the injected oocytes to multiple odorants. The resulting data were obtained and analyzed using Digidata 1440 A and pCLAMP 10.2 software (Axon Instruments Inc., Union City, CA), respectively.

Stock solutions of each odorant were prepared at 1 mol/L concentration using dimethyl sulfoxide (DMSO) as a solvent, and these solutions were then diluted in 1× Ringer's buffer (96 mM NaCl, 2 mM KCl, 5 mM MgCl2, 0.8 mM CaCl2, 5 mM HEPES, pH 7.6 adjusted with NaOH) to the indicated concentrations (10-4 M) for electrophysiological recording. For dose-response curves recording, dilutions were prepared at a concentration gradient: 1×10^{-7}, 3×10^{-7}, 1×10^{-6}, 2×10^{-6}, 3×10^{-6}, 1×10^{-5}, 1×10^{-4}, 3×10^{-4}, and 1×10^{-3} mol/L, with 1×Ringer's buffer containing 0.1% DMSO as the negative control.

2.6. Synthesis of dsApisOR5, OR43, and RNA interference using nanocarrier/dsApisOR complex

The primers for dsRNA synthesis and quantitative real-time PCR were listed in Table S12. dsApisORs were synthesized according to (Zheng et al., 2019). Briefly, amplified sequences were purified by gel extraction and 1 μg was introduced into pMD19T-Vector (Takara, Japan) and then transfected to DH5a™ competent cells (Invitrogen, USA). The extracted plasmid was used as the template for dsApisOR5 and dsApisOR43 synthesis using a T7 RiboMAX expression (Promega, USA). The dsGFP synthesized by the same method was applied as a negative control. A mixture of 4 μL dsApisOR and 1 μL nanocarrier was dispersed in ddH2O at room temperature. The nanocarrier/dsApisOR complex was acquired by mixing detergent and the nanocarrier/dsApisOR solution in a 1:10 ratio. The nanocarrier/dsGFP complex prepared in the same method was used as a negative control. A droplet containing 10-50 nL nanocarrier/dsApisOR complex was produced by a microinjector, and then applied to the abdominal of an adult aphid. Aphids treated with nanocarrier/dsApisOR or nanocarrier/dsGFP complex for 1 h were used as the sample for gene expression level analysis and behavioral assays.

2.7. RT-qPCR

Total RNA of dsApisOR-infiltrated aphids was extracted using TRIzol reagent (Invitrogen, Carlsbad, CA, USA) and then used for synthesizing cRNA utilizing RevertAid First Strand cDNA Synthesis Kit (Thermo Scientific, Waltham, MA, USA). The PCR mixture was prepared based on the protocol of GoTaq qPCR Master Mix (Promega, Madison, USA). Applied Biosystems 7500 Fast Real-Time PCR System (ABI, Carlsbad, CA, USA) was applied to perform the RT-qPCR reactions. The relative mRNA levels of ApisORs were normalized to that of ribosomal protein L7 gene (ApisRPL7) using the $2^{-\Delta\Delta CT}$ method (Livak & Schmittgen, 2001). Primer sequences for RT-qPCR are listed in Table S12.

2.8. ElectroAntennoGraphy recording

EAG responses were detected from antennae of the RNAi-treated aphids stimulated with EBF and GA. Antennae of the aphids were cut from the base and also removed the tip, then inserted between two glass electrodes containing 0.1 mol/L KCl solution. For each test, a 10 mL test solution (containing odorant or solvent) was dropped on a filter paper (0.5 cm×5 cm) which was then inserted into a Pasteur pipette. Paraffin oil was used to stock (in 1 mol/L) and dilute compounds for the test. A humidified continuous airflow was delivered by a stimulus controller (CS-55, Syntech, Kirchzarten, Germany) at 30 mL/s. Odor stimulations were controlled by 0.2 s pulses at 10 mL/s airflows and with a time span of 30 s. EAG signals were amplified using a 10×AC/DC headstage preamplifier (Syntech) and further received with an Intelligent Data Acquisition Controller (IDAC-4-USB, Syntech). The signals were recorded, monitored, and analyzed by Syntech EAG-software (Syntech, Germany). The EAG response values obtained with paraffin oil were subtracted.

2.9. Behavioral experiments

Behavioral assays were carried out using a glass Y-tube olfactometer with a 2.8 cm uniform diameter, 21 cm trunk length and 17.3 cm branch length. To facilitate the movement of the aphids, a Y-shaped copper wire was put in the center of the Y tube. A humidified continuous airflow filtered by activated granular carbon was delivered at 0.5 L/min. Aphids were placed at the end of the copper wire, while two strips of filter paper loaded with EBF and hexane with a concentration of 0.5% (V/V) were placed at the end of each arm randomly. The number of aphids moving into 3 cm of the branch arm and staying there for at least 30 s was discounted. 96 to 132 replications (one apterous aphid each replication) for each group were carried out.

2.10. Statistics analysis

Differences in the number of aphids in each arm of the Y-tube olfactometer were analyzed by a two-sided binominal test (Kappers et al., 2005). Multiple comparisons more than three groups of data were evaluated by the one-way analysis of variance (ANOVA) following Duncan's multiple range test. Two-sample

analysis was performed using Student's t-test. GraphPad Prism software (v.6.0) (GraphPad Software Inc., San Diego, CA, USA) was used to visualize the data. All statistical analyses were assayed using SPSS Statistics (v.22.0) (SPSS Inc., Chicago, IL, USA). The GCMS solution software (v.4.20) was used to output the chromatographic (TIC) data, and the graphics were carried out by the OriginPro software (v.9.5.1.195).

3. Results

3.1. Divergence in alarm pheromone components and in electrophysiological responses to EBF among aphids

To explore how alarm pheromone evolve within Aphididae, we performed gas chromatography/mass spectrometry (GC/MS) analysis to identify the alarm pheromone components of three species from distinct taxa groups, *Therioaphis trifolii* (Calaphidinae), *Eriosoma lanigerum* (Eriosomatinae), and *Cinara cedri* (Lachninae). β-germacrene A and β-elemene were identified in *T. trifolii*, which was consistent with previous reports (Bowers et al., 1977; Nishino et al., 1977). α-pinene and β-pinene were detected as the potential alarm pheromones in *C. cedri*, while three unknown chemicals were detected in *E. lanigerum*. Furthermore, combining the previously published alarm pheromone component data and the pheromone compounds of the three aphids reported in this study, we summarized the composition of alarm pheromone among 42 aphids from 6 subfamilies. A clear divergence in the alarm pheromone component between different subfamilies was found. Notably, EBF is exclusively identified in Aphidinae and Chaitophorinae species and it is identified as the main, or only component specifically in Aphidinae aphids (Figure S6).

3.2. Manual gene annotation result in 532 OR genes from 13 aphid genome assemblies

The divergence of alarm pheromone within aphids indicates the corresponding changes in the olfaction system. Comparation between OR gene families of Aphidinae and non-Aphidinae species may provide valuable information about how ORs in Aphidinae species evolved for detecting EBF. However, the aphid OR gene family has not been widely annotated, especially in the genomes of non-Aphidinae species. We downloaded 15 aphid genomes from NCBI or Aphidbase, and the completeness of the 15 genomes was evaluated using contig N50 and BUSCO. High-quality genomes were selected if it matched one of the following screening criteria: 1) > 90% of the BUSCO genes were characterized as complete and single-copy, or 2) contig N50 was higher than 1000 bp. Two genome assemblies of Aphidinae aphids, *Diuraphis noxia* and *Myzus cerasi*, were abandoned (Figure 20A). The remaining 13 genome assemblies were finally selected based on this threshold. These species covered a taxonomically diverse group belonging to six Aphid subfamilies, including Lachninae, Erioasomatinae, Hormaphidinae,

Chaitophorinae, Calaphidinae and Aphidinae, presenting 25% (6/24) of the aphid subfamilies, of which containing more than 70% of the species currently recognized in Aphididae (Blackman & Eastop, 2020).

In total, 633 OR genes were identified from the 13 aphid genomes through manual gene annotation, among which 83 ORs were pseudogenes, 19 were partial gene models, and 532 ORs were intact gene models. The number of intact OR genes vary considerably among aphids, with the highest number of 62 in *E. lanigerum*, and the smallest number of 17 in *Sipha flava* (Table 6). The intact ORs encoded protein sequences with an average length of 409-419 amino acids, which is

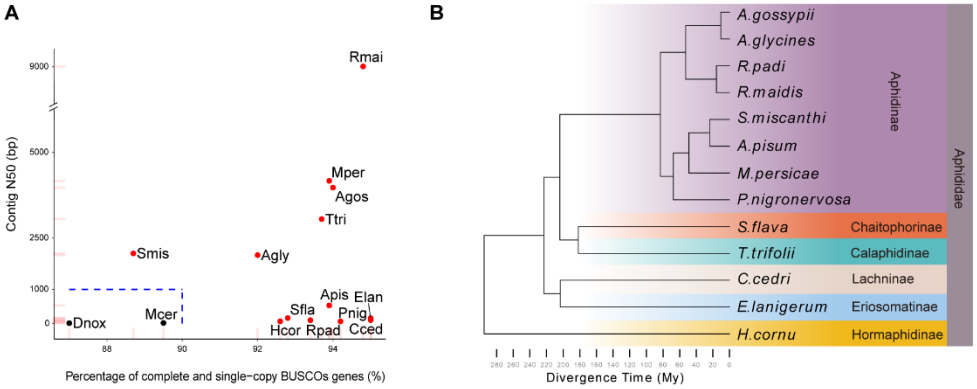


Figure 20. Genome selection and species phylogeny of 13 aphid species. (A) Quality screening of genome assemblies based on percentage of complete and single-copy BUSCOs genes and length of contig N50; Genome is discarded from the following analyses if none of the two conditions is matched: 1) > 90% of the BUSCO genes are characterized as complete and single-copy, or 2) contig N50 is higher than 1000 bp. The name of each species is abbreviated using the first letter of the genus combined with the first three letters of the species epithet. (B) Phylogenetic tree was constructed from the concatenated 1583 orthologous groups from the 13 aphid species, and *Apolygus lucorum* as outgroup (not shown).

consistent with the typical length of insect OR genes (Robertson, 2019). To verify the completeness of intact OR genes, we predicted the transmembrane domains (TMDs) of OR protein sequences. The result showed that an average of 6.26 ± 0.08 (mean±SD) TMDs was predicted by TOPCONS, and 6.04 ± 0.03 (mean±SD) TMDs by DeepTMHMM, which is comparable to the predicted number of TMDs in *Drosophila melanogaster*, suggesting the high accuracy and completeness of the annotated ORs of aphids (Table S8).

The number of OR pseudogenes also varied widely among aphid species, with the pseudogene rate ranging from 2.44% of *A. gossypii* to 26.74% of *A. pisum* (Table 6). There was no positive correlation between the number of pseudogenes and the number of intact genes (R2 = 0.195, P = 0.073) (Figure S7A). High proportions of tandemly located intact OR genes were observed in tested aphid

genomes, such as 63.27% in *Rhopalosiphum maidis*, 60.71% in *T. trifolii*, and 58.90% in *E. lanigerum* (Table 6). The correlation between the number of tandemly located ORs and the number of intact ORs is significantly positive (R2 = 0.830, $P< 0.001$) (Figure S7B), suggesting the contribution of tandem duplication to the expansion of aphid OR gene family.

3.3. Molecular evolution of OR gene family in 13 aphid species

With the highly complete OR gene models, we were able to investigate the molecular evolution of the OR repertoire. Aphididae-wide phylogenetic analysis were conducted using sequences of 532 putatively functional (with intact gene model) ORs from 13 aphid species. The aphid ORs were classified into 16 subfamilies based on high bootstrap support (>95%) and named from A to P (Figure 21A, S8). These subfamilies displayed variable evolutionary patterns. For example, an extremely strict one-to-one orthologous relationship was observed in the subfamily A, which consisted of Orco genes from all selected aphid species. The subfamily C was overall specific to Aphidinae species, whereas one OR from the non-Aphidinae aphid *C. cedri* was found. The subfamily P presenting the largest OR subfamily in the aphid OR phylogeny, in which ORs of both Aphidinae and non-Aphidinae species have undergone massive gene duplication events.

To gain insight into the evolutionary dynamics of the size of aphid OR repertoire, gene gain and loss events for each taxonomic lineage were inferred during the evolution of the aphids. Intensive OR gene turnover was found in each taxonomic lineage (Figure 21B). Several major gain and loss events were observed. For example, the gain of 13 OR genes in *R. padi* and of 14 in *A. pisum* when compared to their most recent common ancestor (MRCA). These changes in the gene numbers were mainly due to the specific expansion in several species-specific OR subfamilies, including the expansions in the subfamily C and P of *A. pisum*, and massive gene duplications found in the subfamily M of *T. trifolii* (Figure 21C). We also found that expansions have occurred even within a short evolutionary time scale, as the OR numbers of subfamily L in *R. padi* have significantly expanded compared to its closely related species *R. maidis* (diverged ~22 million years ago) (Figure 21B and C). Although the aphid OR family has experienced frequent gene gain and loss events, several OR clades were found particularly conserved. To identify ORs that putatively tune to EBF, we focused on searching OR clades that are especially conserved among Aphidinae aphids. As a result, 8 OR clades were identified, and each of them consisted of single-copy ORs from 8 Aphidinae aphids, with no duplication or loss events occurred. These clades were named after the corresponding OR ortholog of *A. pisum*, including ApisOR2, OR3, OR4, OR5, OR20, OR39, OR40, and OR43 (Figure 21A). To evaluate the selective constraints on these OR clades, we calculated the nonsynonymous substitutions per nonsynonymous site (dN) and synonymous substitutions per synonymous site (dS). The ω ratio (ω=dN/dS) for the sequences in each clade was consistently lower than

Table 6. Summary of manual annotation of OR genes in 13 aphid genome assemblies

Species	Intact ORs	Mean Length[a]	Mean TMD TOPCONS[b]	Mean TMD DeepTMHMM[c]	PSE[d]	Partial Models	Tandem[e]	All
Cinara cedri	32	414 ± 15	6.25 ± 0.43	6.03 ± 0.17	7(17.50%)	1	17(42.50%)	40
Eriosoma lanigerum	46	411 ± 15	6.22 ± 0.41	6.02 ± 0.15	13(22.03%)	0	29(49.15%)	59
Hormaphis cornu	23	411 ± 16	6.43 ± 0.50	6.04 ± 0.20	5(16.67%)	2	11(36.67%)	30
Sipha flava	17	417 ± 23	6.24 ± 0.55	6.12 ± 0.33	4(19.05%)	2	4(19.05%)	21
Therioaphis trifolii	44	415 ± 16	6.11 ± 0.32	6.02 ± 0.15	6(11.76%)	1	29(56.86%)	51
Acyrthosiphon pisum	60	414 ± 24	6.25 ± 0.43	6.02 ± 0.13	23(26.74%)	3	24(27.91%)	86
Sitobion miscanthi	40	419 ± 18	6.28 ± 0.45	6.05 ± 0.22	10(20.00%)	0	18(36.00%)	50
Myzus persicae	39	413 ± 20	6.23 ± 0.42	6.03 ± 0.16	7(14.89%)	1	17(36.17%)	47
Pentalonia nigronervosa	25	417 ± 16	6.20 ± 0.49	6.04 ± 0.20	4(12.12%)	4	7(21.21%)	33
Aphis glycines	39	412 ± 19	6.36 ± 0.48	6.03 ± 0.16	3(6.67%)	3	20(44.44%)	45
Aphis gossypii	40	411 ± 19	6.25 ± 0.43	6.03 ± 0.16	1(2.44%)	0	19(46.34%)	41
Rhopalosiphum maidis	46	411 ± 23	6.39 ± 0.49	6.04 ± 0.20	3(6.12%)	0	31(63.27%)	49
Rhopalosiphum padi	56	413 ± 20	6.20 ± 0.40	6.02 ± 0.13	2(3.33%)	2	31(51.67%)	60

[a]Mean length of amino acid sequences of intact ORs (± SD).
[b]Mean transmembrane domains predicted by TOPCONS.
[c]Mean transmembrane domains predicted by DeepTMHMM.
[d]Number of pseudogenes and the rate of pseudogenes.
[e]Number of OR genes in tandem arrays and the rate of OR genes in tandem arrays.

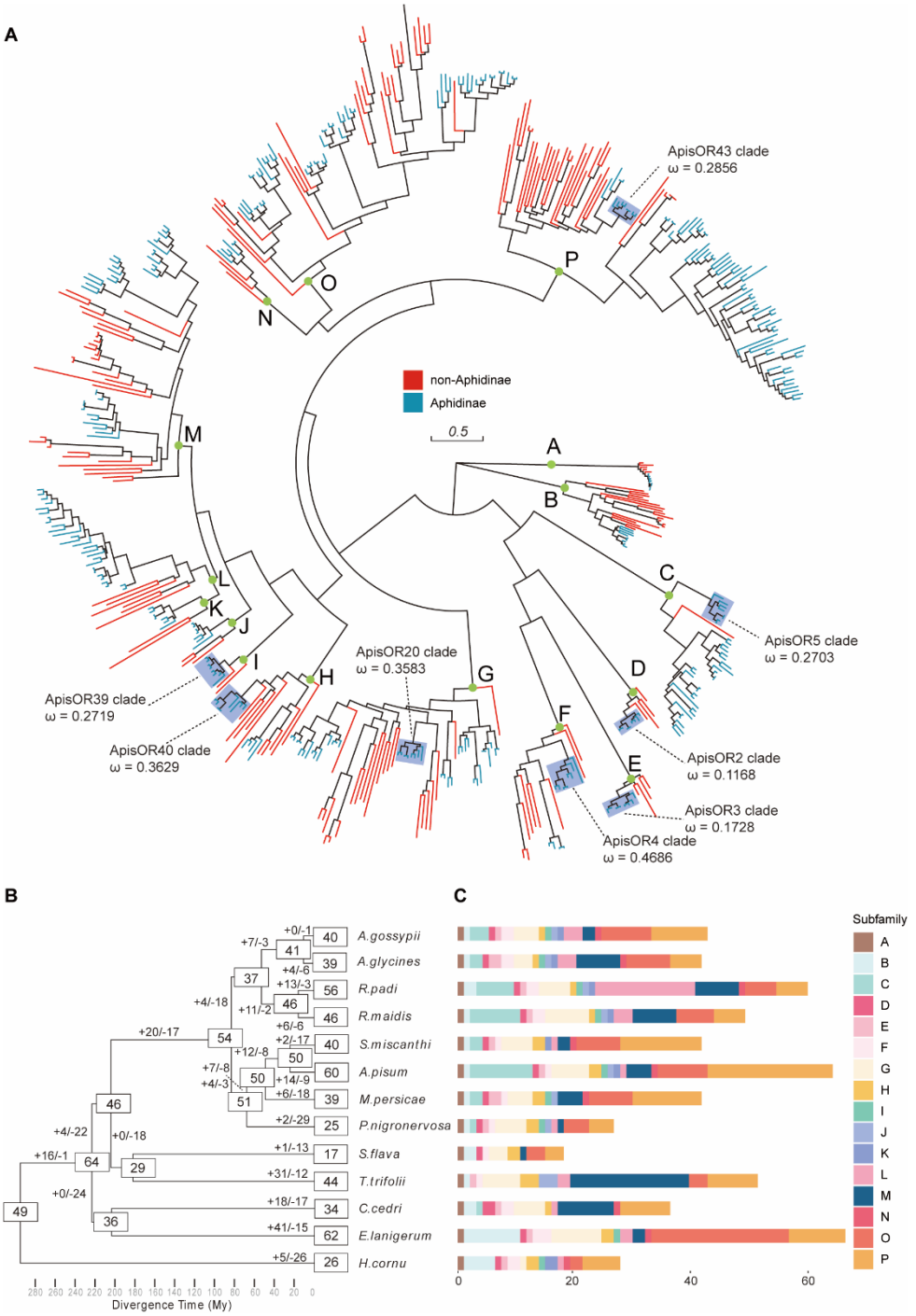

A
ApisOR43 clade
ω = 0.2856
non-Aphidinae
Aphidinae
0.5
ApisOR5 clade
ω = 0.2703
ApisOR20 clade
ω = 0.3583
ApisOR39 clade
ω = 0.2719
ApisOR40 clade
ω = 0.3629
ApisOR2 clade
ω = 0.1168
ApisOR3 clade
ω = 0.1728
ApisOR4 clade
ω = 0.4686
B
A.gossypii
A.glycines
R.padi
R.maidis
S.miscanthi
A.pisum
M.persicae
P.nigronervosa
S.flava
T.trifolii
C.cedri
E.lanigerum
H.cornu
Divergence Time (My)
C
Subfamily
A
B
C
D
E
F
G
H
I
J
K
L
M
N
O
P

Figure 21. Phylogenetic analysis of intact ORs and evolutionary dynamic of OR repertoires among 13 aphids. (A) Phylogenetic tree constructed by protein sequences of 532 intact aphid ORs based on the maximum likelihood algorithm. Branches that indicate ORs of Aphidinae species are colored in blue, while those indicate ORs of non-Aphidinae species are colored in red. Branches of OR subfamilies were identified on the base of high bootstrap support (>80%) and labeled with light blue dots and black letters (A-P). Eight conserved OR clades were indicated in purple shadow, and each of the clades was inferred under functional constraint ($\omega < 0.5$). (B) Changes in the number of OR genes during the evolution of 13 aphids inferred by intact ORs. (C) The number of intact OR genes identified from each subfamily across 13 aphid species.

0.5, which indicated that these ORs have undergone strong purifying selection (Figure 21A).

3.4. Sequence similarity and genome organization of the 8 Aphidinae-specific OR Clades

Highly similar amino acid sequences shared by different ORs may underlie their conserve functions. Therefore, to evaluate the possible functional conservation of the 8 Aphidinae-specific OR clades, pairwise amino acid sequence alignments were carried out to investigate their sequence similarities. The results showed considerably high sequence similarities shared by each ortholog pair, especially in close-related species. For example, pairwise sequences with more than 95% similarity were found between OR orthologs of *A. glycine* and *A. gossypii*, as well as between *R. padi* and *R. maidis* (Figure 22A). The distribution of pairwise sequence similarities with median and mean value of each clade have been shown (Figure 22B). The mean value of each clade was higher than 80%, indicating significant sequence conservation of these OR subgroups.

Some of the ORs in these clades have been functionally characterized. For example, two members of the ApisOR5 clade, ApisOR5 and AgosOR5, have been reported to be involved in EBF detection and mediated repellency behavior of aphids to EBF (Zhang et al., 2017), Indeed, pairwise sequence similarities of OR5 orthologs ranged from 76.82% to 100.00% with an average sequence similarity of 84.85%, suggesting ApisOR5 orthologs in Aphidinae species may be functionally conserved for detection of EBF. The pairwise sequence similarities in ApisOR20 clade ranged from 77.73% to 97.86%, and had an average value of 84.85%. Functional characterization of OR20 clade showed that ApisOR20 was tuned to the herbivore-induced plant volatile *cis*-jasmone (Bo Wang et al., 2022), which was repellency to several aphid species (Birkett et al., 2000; Dewhirst et al., 2012; Sobhy et al., 2017). Moreover, the ApisOR4 clade, which had the smallest average sequence similarity (80.01%) in all tested clades, and pairwise similarities of this clade varied from 65.90% to 98.4%. In this clade, ApisOR4 was characterized to detect 8 host plant volatiles (Zhang et al., 2019), suggesting the important role of this OR in host location.

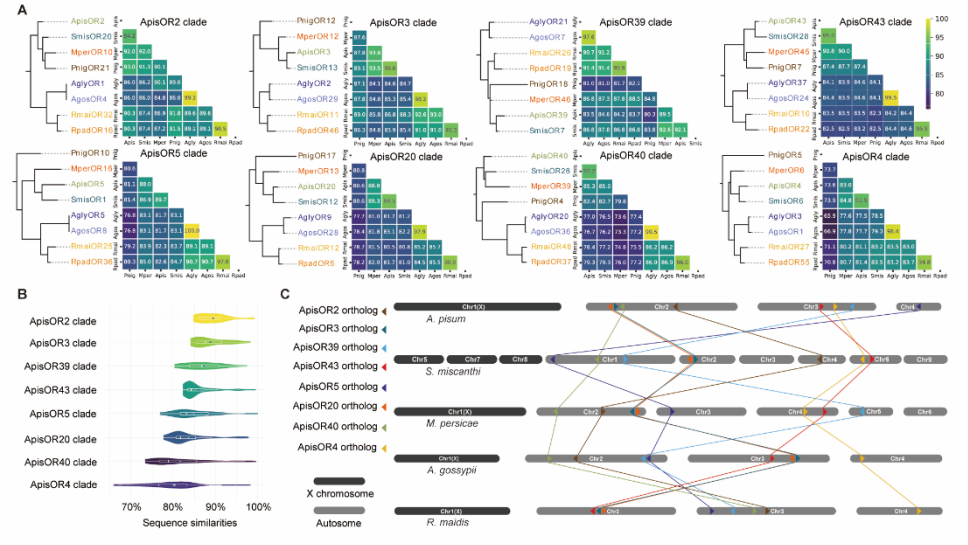


Figure 22. Amino acid sequence similarity analyses and genomic position of the 8 conserved OR clades in aphids. (A) Phylogenetic tree and pairwise amino acid sequence similarities amongst ORs of 8 conserved OR clades. Color codes in the heat map indicate the pairwise amino acid sequence similarities, which range from yellow (100% identity) to dark blue (65.9% identity). (B) The distribution of pairwise amino acid sequence similarities in each clade. Boxplots overlapped by violin plots showing the median values (white bar) of pairwise sequence identities in each clade. Dark blue spots show the distribution and mean values of the sequence identities of OR2, 3, 39, and 43 clades, while light blue spots show the median values of the sequence identities of OR5, 20, 40, and 4 clades. (C) Genomic position and synteny of ORs from 8 conserved OR clades in five chromosome-level genome assemblies. Orthologs to ApisORs are shown on each genome, and orthologs of the same ApisOR in each genome are linked to synteny.

We next investigated the genome organization of the orthologs from five chromosome-level aphid genome assemblies. Although tandem duplications have greatly contributed to the expansion of OR gene family, we did not observe tandem duplicated OR cluster in these orthologs, suggesting the independent evolution of these ORs (Figure 22C).

3.5. OR5 and OR43 in EBF detection in aphids

To explore whether and which of the 8 Aphidinae-specific ORs are EBF receptors, we measured the responses of ORs co-expressed with Orco to EBF using the *Xenopus* oocytes system under two-electrode voltage-clamp perfusion in the model species *A. pisum*. We did not find any EBF response in the oocytes co-expressing ApisOR2, OR3, OR4, OR20, OR39, and OR40, consistent with our previous results (Zhang et al. 2017). Surprisingly, in addition to the previously characterized EBF receptor, ApisOR5, oocytes co-expressing ApisOR43/Orco significantly responded

to EBF, with a very weak response to geranyl acetate (GA), a structurally related chemical to EBF (Figure 23A, B). The responses of ApisOR43/Orco to EBF were significantly higher than that of ApisOR5/Orco, and the opposite results were activated by GA (Figure 23 C, D). Moreover, a dose-dependent curve of ApisOR43 to EBF showed a value of EC50 of 2.91×10^{-6} mol/L (Figure 23E, F), indicating the sensitive response of ApisOR43/Orco to EBF. These results show that ApisOR43 is also tune to EBF, and may be more specific to EBF than ApisOR5.

To test whether ApisOR43 is essential for mediating EBF-induced repellent behavior in *A. pisum*, *ApisOR5* and *ApisOR43* were knocked down using RNA interference (RNAi) technique (Figure 23G, H). The repellent behavior to EBF was lost in both dsApisOR5-infiltrated, dsApisOR43-infiltrated and dsApisOR5&43-infiltrated aphids compared to WT, and no significant difference between these three groups (Figure 23I). Next, EAG recording was performed on the antennae of RNAi-treated aphids to compare the *in vivo* function of ApisOR5 and ApisOR43. We observed that EAG responses to EBF significantly reduced when *ApisOR5* and *ApisOR43* were knocked down. Moreover, when ApisOR43 and ApisOR5 were simultaneously interfered, the EAG responses were significantly lower than that of when ApisOR43 and ApisOR5 are interfered separately (Figure 23J). The above results suggest that these two ORs may equally functioned in the olfaction of EBF.

3.6. Evolutionary history of EBF receptors in aphids

Phylogenetic analysis shows that the OR clades of ApisOR5 and ApisOR43 are highly conserved in Aphidinae species, but they are absent from the five non-Aphidinae species (Figure 2B and 3A). We are curious about when is EBF receptors originated and how they evolved during the evolution of aphids. We further analyzed the duplication and pseudogenization of OR5 and OR43 genes in 17 genomes of aphid species, including the 13 genomes used in previous analyses and 4 recently published genomes: *Chaitophorus viminalis* (Chaitophorinae), *Stegophylla sp.* (Phyllaphidinae), *Geopemphigus sp.* (Eriosomatinae) and *Pemphigus obesinymphae* (Eriosomatinae) (Smith et al., 2022).

As a result, the orthologs of ApisOR5 and ApisOR43 were found in all tested Aphidinae species. Interestingly, orthologs of ApisOR5 were also found in the genomes of 2 non-Aphidinae aphids, including *C. viminalis* and *S. flava*. *C. viminalis* and *S. flava* belong to Chaitophorinae, which is a close subfamily to Aphidinae. In order to verify whether the orthologs of ApisOR43 in Chaitophorinae species are also involved in sensing EBF, the responses of oocytes co-expressing CvimOR5/Orco to EBF were measured. CvimOR5 is found to be also tuned to EBF, suggesting the same function as orthologs in Aphidinae species.

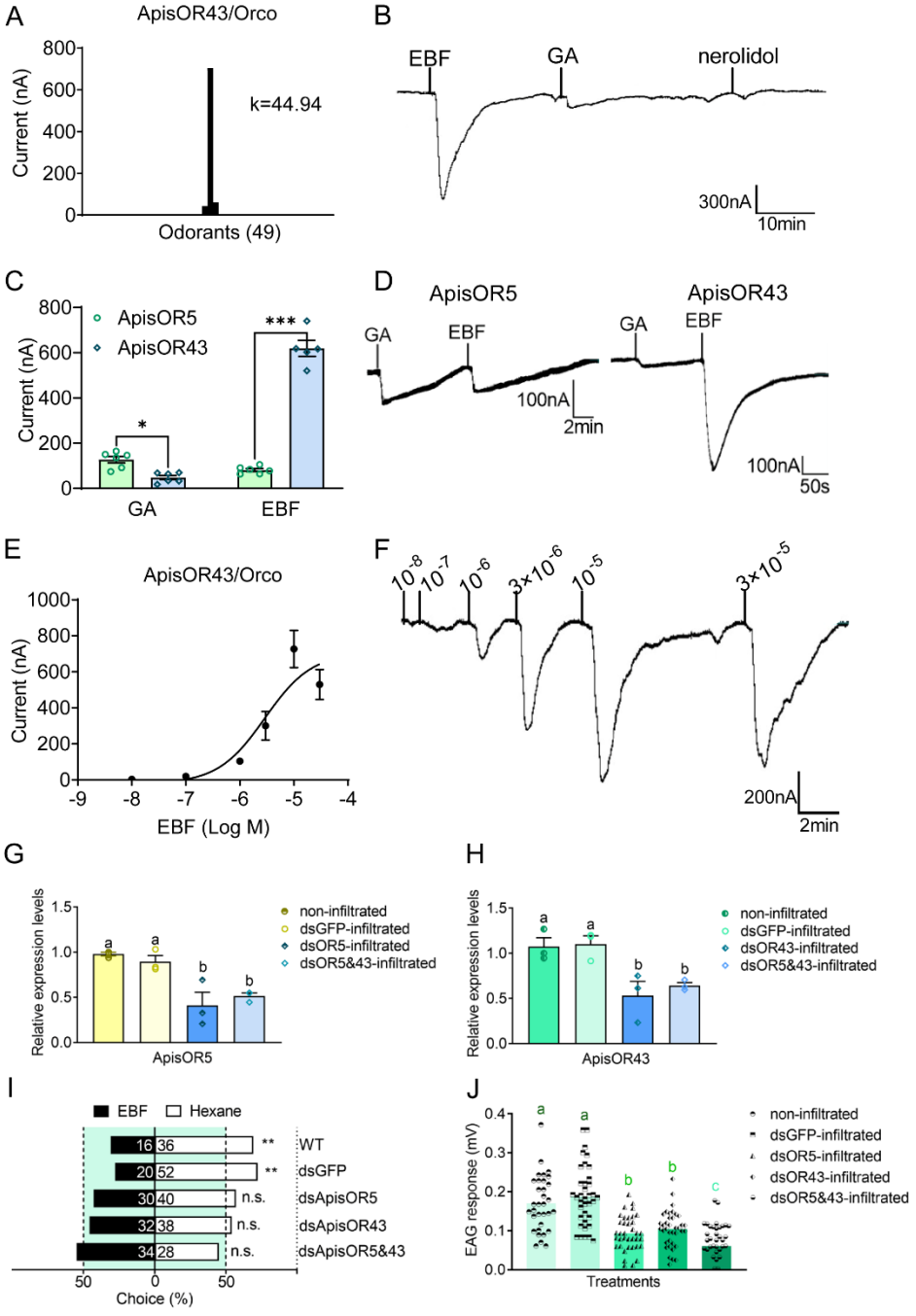

A
ApisOR43/Orco
Current (nA)
k=44.94
Odorants (49)
B
EBF
GA
nerolidol
300nA
10min
C
ApisOR5
ApisOR43
Current (nA)
GA
EBF
D
ApisOR5
ApisOR43
GA
EBF
100nA
2min
100nA
50s
E
ApisOR43/Orco
Current (nA)
EBF (Log M)
F
200nA
2min
G
Relative expression levels
non-infiltrated
dsGFP-infiltrated
dsOR5-infiltrated
dsOR5&43-infiltrated
ApisOR5
H
Relative expression levels
non-infiltrated
dsGFP-infiltrated
dsOR43-infiltrated
dsOR5&43-infiltrated
ApisOR43
I
EBF
Hexane
WT
dsGFP
dsApisOR5
dsApisOR43
dsApisOR5&43
n.s.
Choice (%)
J
EAG response (mV)
non-infiltrated
dsGFP-infiltrated
dsOR5-infiltrated
dsOR43-infiltrated
dsOR5&43-infiltrated
Treatments

Figure 5. *In vitro* and *in vivo* functional characterization of ApisOR5 and ApisOR43 from *Acyrthosiphon pisum*. (A) Tuning curve of ApisOR43/Orco to 49 tested chemicals. (B) Representative inward current response of ApisOR43/Orco to (*E*)-β-farnesene (EBF), geranyl acetate (GA), and nerolidol at the concentration of 10^{-4} mol/L. (C) Inward current values of ApisOR5/Orco and ApisOR43/Orco to GA, and EBF. Significant differences were observed between the responses of ApisOR5/Orco and ApisOR43/Orco to GA (*$p < 0.05$) and to EBF (***$p < 0.001$). Data are plotted as mean ± SEM (n = 6 for ApisOR5; n = 5 for ApisOR43). (D) Representative inward current response of ApisOR5/Orco and ApisOR43/Orco to EBF and GA. (E) Dose-response curves of ApisOR43/Orco to EBF. EC_{50} value of ApisOR43/Orco is 2.91×10^{-6} mol/L (n = 6). Data are reported as mean ± SEM. (F) Dose-response trace of ApisOR43/Orco expressed in *Xenopus* oocytes to EBF. (G, H) Relative expression levels of *ApisOR5* and *ApisOR43* transcripts after dsRNA infiltrated in *A. pisum*, respectively (mean ± SEM, n = 3, GLM following by Duncan's multiple range test). Bars labeled with different letters are significantly different. (I) Behavior of wild-type and dsRNA-infiltrated *A. pisum* to EBF (10 μg/μL) in a Y tube olfactometer. EBF is significantly repellent to wide-type and dsGFP-infiltrated aphids, while no significant differences of behavioral choice are found in dsApisOR5, dsApisOR43 and dsApisOR5&43 infiltrated aphids. The white bar represents hexane treatment (control) and the black bar represents EBF treatment (n.s., $p > 0.05$, no significant difference; **$p < 0.01$; unpaired two-tailed Student's *t* test, n=52-72). (J) EAG responses of dsRNA-infiltrated pea aphids to EBF (n = 33-41, GLM following by Duncan's multiple range test). Bars labeled with different letters are significantly different. Plotted data are mean ± SEM.

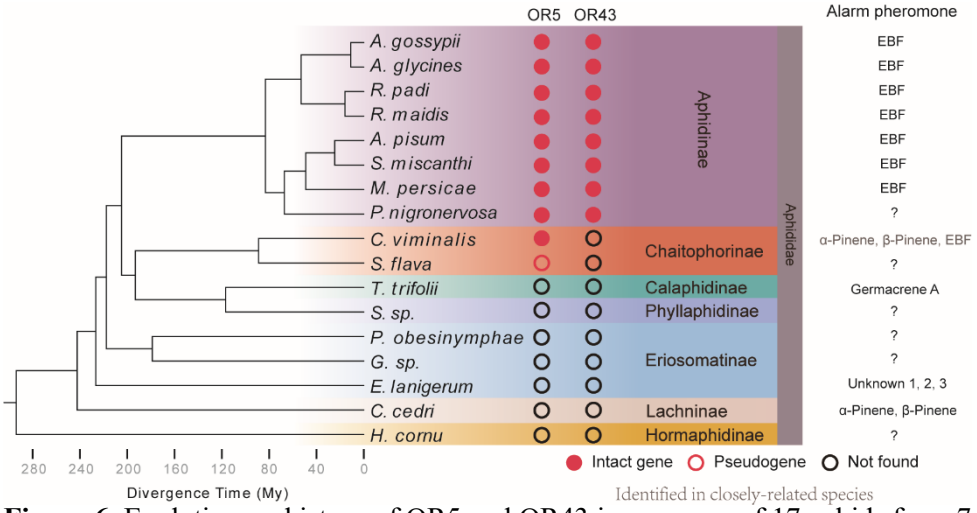


Figure 6. Evolutionary history of OR5 and OR43 in genomes of 17 aphids from 7 subfamilies. Phylogeny of 17 aphids inferred from the concatenated 1074 orthologous groups using *Apolygus lucorum* as an outgroup (not shown). Orthologous OR5s are only identified in the genomes of Aphidinae and Chaitophorinae aphids, while orthologous OR43s are constructively identified in

the genomes of Aphidinae aphids. None of OR5 and OR43 orthologues was identified in the genomes of the 7 relatively ancient aphids. Correspondingly, the alarm pheromone components of these species were identified as monoterpenoids (α-pinene and β-pinene) or sesquiterpenes (Germacrene A), but not EBF, though alarm pheromone components of some species were unknown or have not been identified due to these species were not available. Notably, α-pinene, β-pinene, and tiny EBF was identified as alarm pheromone components of *Chaitophorus populet* in Chaitophorinae species (Francis, et al. 2005), which is a closely related species to *C. viminalis*, suggesting that EBF maybe also used by *C. viminalis* as alarm pheromone components. Besides, we identified an orthologous OR5 in the genomes of *C. viminalis*, and a pseudogene SflaOR1 from S. flava, both belonged to the Chaitophorinae subfamily. CvimOR5 maybe co-evolved with EBF when it became alarm pheromone.

EBF has been identified as a minor alarm pheromone component in a species from the genus *Chaitophorus*, *C. populet*, suggesting the possible role of CvimOR5 in mediating EBF repellent behavior in this genus. We also observed that SflaOR1, the ApisOR5 ortholog in *S. flava*, was pseudogenized due to a premature stop codon, suggesting the selection on this OR may has relaxed (Figure S10). Moreover, phylogenetic trees showed that no orthologs of ApisOR43 were identified out of Aphidinae species due to the low bootstrap support (55%) observed in OR subfamily P (Figure S11B). This result was further supported by ortholog group classification using OrthoFinder algorithm (Emms & Kelly, 2019), indicating that OR43 has evolved exclusively in the Aphidinae aphids (Table S13)

4. Discussion

The rapid advances of genome-sequencing techniques have led to a remarkable increase in the number of aphid genome assemblies (Mei, et al. 2021), which provides a great opportunity for investigating the molecular evolution of the OR gene family in aphids. While previous identification of aphids OR repertoires have been mostly focused on species of Aphidinae subfamily (Robertson et al., 2019), little is known about the molecular evolution among a larger species group. In the present work, we manually annotated a substantial set of OR gene from 13 aphids across six aphid subfamilies. This set for the first time, obtained OR genes of aphids out of Aphidinae, which provide valuable information for investigating the potential evolutionary pattern of non-Aphidinae aphids. Indeed, some patterns would be difficult to detect if only study on the lineage of Aphidinae, including the massive gene loss of non-Aphidinae species in the OR subfamily C as well as the various OR lineages of non-Aphidinae aphids that have been specifically expanded. However, the species selected in present analysis are still limited, as we may not know whether there are other species- or subfamily-specific OR lineages (like the OR subfamily C). Future studies should consider to perform genome sequencing on more aphids to enrich genetic resources of the whole aphid group.

EBF has long been considered the most representative aphid alarm pheromone, and consequently received considerable attention over the last few decades (Vandermoten et al., 2012). However, studies have found that, instead of EBF, monoterpenoids or sesquiterpenes are also acted as alarm pheromone in some species (Francis et al., 2005; Nishino et al., 1977). These results have implied a clear divergence of aphid alarm pheromone composition. However, due to the lack of a systematic summary of the pheromone composition of different species and the serious shortage of records about non-EBF pheromones, the evolutionary pattern of this ecologically important pheromone within aphids is largely unknow. Through statistics of the previously published data and the results of pheromone components of *E. lanigerum* (Eriosomatinae) and *C. cedri* (Lachninae) in this work, we show a possible evolutionary path (though not fully conclusive) of aphid alarm pheromone: monoterpenoids and sesquiterpenes are widely existing in aphids until the emergence of Chaitophorinae and Aphidinae. EBF was mostly become the major or only component of alarm pheromone in Aphidinae species. Future efforts should consider identified alarm pheromone from the species of more aphid subfamilies.

Although the composition of alarm pheromone among aphids may have differentiated, we would like to know if aphids without EBF can detect the latter. We found that OSNs housed in LP5 and LP6 were not activated by EBF in both aphids *T. trifolii* and *E. lanigerum*, but neuronal responses to EBF were unexpectedly recorded in LP5 and SP5 of *C. cedri*. Our results have confirmed that the alarm pheromone of *C. cedri* consisted of monoterpenes instead of EBF. We thereby suggest that the neuronal responses of the fifth segment of antenna in *C. cedri* may be activated by plant-derived EBF. This chemical can be isolated from cedar tree, the host plant of *C. cedri* and most Lachninae species (Boudarene, Baaliouamer, Meklati, & Scharff, 2004) and also identified from aphid attacked plants (Kivimäenpää, Babalola, Joutsensaari, & Holopainen, 2020). Subsequently behavioral assay didn't show any repellent behaviors of *C. cedri* to EBF, suggesting that EBF may be active as host plant volatile instead of the alarm signal to *C. cedri*. Moreover, genomic annotation has shown that orthologs of OR5 and OR43 are absent from the *C. cedri* genome, which further indicates the EBF receptor of *C. cedri* is divergent from Aphidinae species, and the mechanism underlying EBF detection in *C. cedri* may be different.

5. Acknowledgements

This study was supported by a grant from Shenzhen Science and Technology Program (Grant No. KQTD20180411143628272) and the National Natural Science Foundation of China (31872039), and the Science and Technology Innovation Program of the Chinese Academy of Agricultural Sciences.

Chapter VII

General discussion, conclusions and perspectives

Aphid are among the most serious of agricultural and horticultural pests in both Europe and China. Around 450 aphids feed on crop plants, causing significant economic damage (Blackman & Eastop, 2000). The ability to sense chemosensory cues is crucial for survival of insects, as it has widely participated in various vital behaviors, including host seeking, response to plant defense and avoidance of dangerous. Therefore, understanding the molecular basis of olfaction can be valuable in developing biological control methods and products, which able to mediate behaviors of pests. While the olfaction system and molecular mechanisms of chemical recognition has been extensively studied in model species such as the fruit fly, *D. melanogaster* (Anholt, 2020), *Anopheles* mosquito (Konopka et al., 2021) and some destructive pests, including the cotton bollworm, *Helicoverpa armigera* (Guo et al., 2020) and beetles (Roberts et al., 2022), less is known how aphids perceive various ecologically important chemical cues in their environment. This thesis characterizes odorant receptors (ORs) involved in the detection of plant-source volatiles and aphid alarm pheromones, which significantly improves our knowledge of the molecular mechanisms of chemical perception in aphids.

1. Molecular basis of host plant location may conserve among species with distinctive host range

Insect OR genes evolve rapidly, resulting extensive differentiation and birth and death of gene copy number (Eyun et al., 2017; McBride, 2007; Mitchell, Schneider, Schwartz, Andersson, & McKenna, 2020). Such evolutionary pattern has been detected in the OR repertoire of the pea aphid *A. pisum* (Smadja et al., 2009). These authors present significantly expansion of OR and gustatory receptor (GR) genes in the genome of *A. pisum*, suggesting the potential functional innovation in OR gene family. Although this work shed light on the OR evolutionary mechanism in aphids, the information provided is relatively limited due to the fact that only one species was included. *A. pisum* is able to feed on multiple legume plants, including more than 20 legume genera, such as pea, clover, alfalfa, and broad bean (Ferrari, Via, & Godfray, 2008; Peccoud, Ollivier, Plantegenest, & Simon, 2009). The considerably wide host range of *A.pisum* suggests a diversity of OR functions in perceiving various host plant volatiles. Thus, the OR expansions in the genome of *A. pisum* is questionable as a special case.

In the chapter III, we presented an OR phylogenetic tree to compare the OR evolutionary pattern of aphids with distinctive host ranges. Besides *A. pisum*, two aphid species are included. *A. glycines*, a legume specialist with a narrower host range than *A. pisum*, mainly feeding on soybean during the spring and summer and *Rhamnus* trees in winter (Crossley & Hogg, 2015). Additionally, we add the cotton aphid, *A. gossypii*, a generalist species that infests nearly 100 crop species worldwide, including cotton, cucurbits, citrus, coffee, cocoa, eggplant, peppers, potato, okra and has highly differentiated host species compared to *A. pisum* and *A. glycines* (Hulle, Chaubet, Turpeau, & Simon, 2020). The smaller OR gene number and less species-specific OR gene expansion detected in the genome of *A. glycines*

and *A. gossypii* suggesting the different evolutionary pattern of OR gene family between the three species.

While the general OR gene evolution in these aphids may differ, certain OR orthologs were conserved. We showed the sequence conservation of ApisOR23 orthologs among eight aphid species, suggesting the crucial function of this receptor for the survival of aphid species. Indeed, the functional characterization of ApisOR23 indicates it may plays an important role in detecting green leaf volatiles (GLVs) including *cis*-2-hexen-1-ol, hexyl acetate and *trans*-2-hexen-1-al) (Sarang, Rudziński, & Szmigielski, 2021). Additionally, ApisOR23 is likely to be involved in regulating behavioral responses of aphids to *trans*-2-hexen-1-al, as the attractiveness of this chemical to the black bean aphid *A. fabae* has been reported (Webster et al., 2008). Future work should consider to carry out behavioral experiments upon aphids with ApisOR23-interfered, consequently to show how this OR influences the responses of aphids to various host plant volatiles. Moreover, it is possible that other ORs also mediate the related behaviors of this chemical. Future efforts should be considered to characterize more ORs using the same approach that we performed in this study.

Notably, *trans*-2-hexen-1-al is not the major chemical component released by cotton plants, whether it is used as a chemical cue for host location in *A. gossypii* is still unknown. Future studies analyzing behavior could investigate the attractiveness of *trans*-2-hexen-1-al to *A. gosspyii*. The fact that *trans*-2-hexen-1-al is the best ligand for ApisOR23 suggests the vital role of this OR in host seeking behaviorof two legume specialists (*A. pisum* and *A. glycines*). While the function of ApisOR23 maybe conserved among multiple species with distinct host ranges, this OR may still be involved in different aspects of aphid life cycle.

2. cis-Jasmone perception in legume specialist aphids

Herbivory-induced plant volatiles (HIPVs) are a class of volatile organic compounds (VOCs) emitted by plants in response to herbivory, or the feeding of plant-eating organisms such as insects or mammals. These VOCs are known to play an important role in the ecological interactions between plants and herbivores, as they serve as important chemical signals for both plant defense and communication (Takabayashi, 2022; Takabayashi & Shiojiri, 2019).

For aphids, however, HIPVs can be detrimental to their survival as these chemical cues are used by aphid parasitoids to locate aphids (van Poecke & Dicke, 2004). Therefore, the ability to detect HIPVs and generate avoidance behavior in response to them are crucial for aphid. In chapter IV, we used *M. crassicauda* as a target species for investigating the molecular basis underpinning HIPV perception in aphids. *M. crassicauda* is a legume specialist which mainly feed on the genera *Vicia* (Hales, Gillespie, Wade, & Dominiak, 2017). The host range of *M. crassicauda* is relatively narrow compared to that of *A. pisum* and *A. glycines*, we therefore suggest their chemosensory-related gene set maybe contracted, with fewer redundant functions remaining in their chemosensory system. We did identify a limited number of OR genes in the antennae transcriptome of *M. crassicauda*, but we

cannot exclude the possibility that the small number of OR transcripts is due to the low expression level of some OR genes in the antennae. The identical number of OBPs and CSPs between the three aphids may support our hypothesis, as OBPs and CSPs are typically highly expressed and readily assembled into complete structures.

It is important to note that the completeness of a transcript assembly is largely influenced by the quantity and coverage of Illumina reads that are mapped onto it. Therefore, if the expression level of OR genes in the antennae is particularly low, assembling the whole transcript (or even part of it) may pose challenges. A feasible approach to obtaining a complete OR gene repertoire would involve utilizing genome resources for gene annotation. However, the genome assembly for *M. crassicauda* is currently unavailable. Given that this species serves as an ideal model for comparative analysis between legume specialists with differing host ranges, the genome assembly of *M. crassicauda* will provide an invaluable genomic resource for studying host specialization in aphids.

The OR responds to *cis*-jasmone has conservatively evolved in *A. pisum* and *M. crassicauda*, possibly due to the critical ecological role *cis*-jasmone plays in the lives of aphids. *cis*-Jasmone is a naturally occurring volatile compound that plants emit when under attack by herbivores (Loughrin, Potter, & Hamilton-Kemp, 1995; Rose & Tumlinson, 2004). This compound not only triggers plant defense responses (Matthes et al., 2010), but also acts as an attractant for predators of herbivorous insects. Therefore, *cis*-jasmone likely serves as a chemical signal that alerts aphids to potential dangers, both from predators and plant defense responses. Although this compound is crucial for aphid survival, the corresponding chemoreception mechanisms remain poorly understood. The characterization of McraOR20 and ApisOR20, for the first time, sheds light on the molecular mechanism of *cis*-jasmone perception in aphids. Given the limited number of ORs characterized in this study, other ORs may also respond to HIPVs. Moreover, OR20 may not be the only OR responsible for detecting *cis*-jasmone; other ORs attuned to this chemical may contribute to a multi-receptor mechanism for *cis*-jasmone detection. In addition, the functional validation of McraOBP3/7/9 should be considered, as the OBPs in *M. crassicauda* closely resemble those in other aphids, suggesting these three OBPs may also play a role in transporting EBF to the neurons. In general, the understanding of the chemoreception of *cis*-jasmone, as well as other HIPVs in aphids, remains limited, and future efforts on characterization more HIPV-tuned ORs are needed.

3. Evolution of EBF receptors in aphids

Aphids, insects with limited mobility that typically live in colonies, are easy targets for predators. When under attacked, they release alarm pheromones from their cornicles. These chemicals induce various behavioral adaptations, such as walking away or dropping off the host plant, which ultimately enhance their survival rate against predators or parasitoids. Extensive research has been conducted to investigate the ecological role of alarm pheromones in the interactions between plants, aphids and their natural enemies (Bushra & Tariq, 2014; Powell &

Pickett, 2003; Shih, Sugio, & Simon, 2023). EBF is one of the most notable alarm pheromones in aphids, as it is the major alarm compound of many economically destructive aphid species, such as the pea aphid, *A. pisum* (Bowers et al., 1972), the green peach aphid, *M. persicae* (de Vos et al., 2010), and the cotton aphid, *A. gosspyii* (Bowers et al., 1972). In chapter VI, we summarized the published data on the composition of alarm pheromones in multiple species and, for the first time, characterize the alarm pheromones of species from the Eriosomatinae (*E. lanigerum*) and Lachninae (*C. cedri*). We discovered that EBF is the main or sole alarm pheromone component in most Aphidinae species, suggesting a unique evolutionary trajectory of EBF within the Aphidinae species.

The subfamily Chaitophorinae is likely pioneered the synthesize and use of EBF as a behavior-mediating compound among the Aphididae. Francis et al. (2005) identified the components of alarm pheromone in *Chaitophorus populeti*, including α-pinene, β-pinene, isobornyl acetate, camphene, EBF, limonene. α-pinene was found to be the most abundant chemical, with EBF present at relatively lower levels. To date, Chaitophorinae is the first non-Aphidinae subfamily in which EBF has been identified, and repellent behavior in respond to EBF has been observed in the *C. viminalis* (Nault, Montgomery, & Bowers, 1976). Phylogenetic analysis based on the single-copy orthologs from multiple aphid species indicate that Chaitophorinae is closer to other non-Aphidinae subfamilies (such as Calaphidinae) than to Aphidinae. Given that Chaitophorinae evolved 100 million years earlier than Aphidinae, EBF may have originated in Chaitophorinae. Moreover, the chemoreception system responding to EBF is simpler in Chaitophorinae compare to Aphidinae, with only OR5 detected in their genome assemblies. This suggests a possible evolutionary trajectory: EBF detection may have initially been accomplished by a single OR in Chaitophorinae and subsequently evolved into a two-receptor system in Aphidinae. However, we cannot exclude the possibility that other Chaitophorinae-specific ORs may have been involved in this process.

In another Chaitophrinae aphid, *S. flava*, the ortholog of ApisOR5 is annotated as a pseudogene, suggesting a possible relaxed selection on this OR. This could be attributed to EBF not serving as the primary alarm signal in *S. flava*. Even though the exact alarm pheromone compounds in *S. flava* remain unknown, we hypothesize that its alarm pheromone composition is identical to that of *C. viminalis* due to their close phylogenetic relationship. However, the role of EBF in the alarm signal transduction in *S. flava* may have diminished, with other compounds now functioning as the major pheromone components. This finding underscores *S. flava* as an important species for understanding EBF evolution among Chaitophrinae aphids. It is therefore necessary to identify the alarm pheromone components of additional species closely related to Aphidinae. Moreover, future work should consider to sequence the genomes and identify the alarm pheromone components of more species close to Aphidinae, as these data are important for revealing the origination of EBF in aphids.

The characterization of EBF receptors specific to Aphidinae species provides insight into the unique molecular mechanisms underpinning EBF detection in these

aphid species. Notably, in addition to the ApisOR5 orthologs, the orthologs of ApisOR43 identified in Aphidinae aphids are also implicated in perceiving and mediating repellant responses to EBF. Interestingly, the substantial reduction in EAG responses in aphids with OR5&OR43 interfered, compared to those with OR5/OR43 interfered, suggests a crutial role for OR43 in amplifying EBF signals within aphid antennae. This amplification of the EBF signal is particularly significant in the Aphidinae species where EBF has evolved to become the main or even sole compound in their alarm pheromone. The evolution of a novel EBF receptor is likely an adaption to the transition of alarm pheromone composition (from monoterpenes or sesquiterpenes in non-Aphidinae species to EBF in Aphidinae species).

The behavioral experiments yielded no significant differences, whether ApisOR5 and ApisOR43 were subjected to RNAi simultaneously or indenpendently. This result suggests that both ApisOR5 and ApisOR43 play equally critical roles in mediating aphid repellent behaviors. Analyzing the projections of neurons expressing ApisOR5 and ApisOR43 could provide further insight into the neuron system that controls repellent behaviors. Two models are speculated to explain this process. In one model, the neurons expressing ApisOR5 and ApisOR43 are projected to different glomerulus, while in the other model, these two ORs co-expressed in the same neuron. The latter situation has indeed been observed with both the highly divergent ORs (Goldman, Van der Goes van Naters, Lessing, Warr, & Carlson, 2005) and closely-related paralogs (Auer, Álvarez-Ocaña, Cruchet, Benton, & Arguello, 2022) in flies, and the co-expression of ORs appears surprisingly widespread in *Drosophila* species (Task et al., 2022). Due to the lack of available genetic tools, we cannot explore how these two EBF receptors code for repellent behaviors in the central neuron system. As such, it is of great importance to develop an efficient method for generating genome-edited aphid lineages for further studies at the neuronal level.

APPENDICES

Scientific publications and supplementary materials

1. Scientific publications

Huang, T., Liu, Y., He, K., Francis, F., Wang, G., and Wang, B. (2023). Chromosome-level genome assembly of the spotted alfalfa aphid *Therioaphis trifolii*. *Scientific Data*, 10：274.

Huang, T.*, Zhang, R.*, Yang, L., Francis, F., Wang, B., and Wang, G. (2022). Identification and functional characterization of a novel odorant receptor in pea aphid *Acyrthosiphon pisum*. *Journal of Integrative Agriculture*, 21(5): 1414-1423. (Cover article)

Wang, B.*, **Huang, T.***, Yao, Y., Francis, F., Yan, C., Wang, G., and Wang, B. (2021). Identification and characterization of a conserved odorant receptor from the aphid *Megoura crassicauda* that specifically responds to *cis*-jasmone. *Journal of Integrative Agriculture*, 21(7): 2042-2054.

Wang, C., Liu, Lei., **Huang, T.**, Zhang, Yu., Liu, Yang., Wang, Guirong. (2023). Characterization of the pheromone receptors in *Mythimna loreyi* reveals the differentiation of sex pheromone recognition in *Mythimna* species. *Insect Science*. (Accept)

Yang, L.*, Liu, H.*, Wang, H.*, **Huang, T**.*, Liu, B., Yang, B et al. (2020). *Apolygus lucorum* genome provides insights into omnivorousness and mesophyll feeding. *Molecular Ecology Resources*. 21:287-300.

Cao, S., **Huang, T**., Shen, J., Wang, GR. (2020). An Orphan Pheromone Receptor Affects the Mating Behavior of *Helicoverpa armigera*. *Frontiers in Physiology*. 11：413.

Tian, K., Liu, W., Feng, L., **Huang, T.**, Wang, G., Lin, K. 2020. Functional characterization of pheromone receptor candidates in codling moth *Cydia pomonella* (Lepidoptera: Tortricidae). *Insect Sciences*. 28(2):445-456.

Huang, T.*, Yang, L.*, Francis, F., Wang, B., Wang, G. 2023. Genome-wide analysis reveals evolutionary mechanisms of EBF-receptors in aphids. *Molecular Biology and Evolution*, (In preparation)

Supplementary materials

>ApisOr23

**MNLNDEQNYIVNLKLMKITGFYHLISPRAPKYFGFNVYKVTAAIEVMTGIFSIIMLFLSS
YYYLDNTNELMSHFMLVVAIFFSTLKIFWVSRNSETIWNNMDMTCINFLLYTGHKKEI
LKKARAKSISTTILFVILWSSVTVAWSISPFFVKDVYLNIKFKDETRRFRYNSLNYVYPIS
EEFYNEHFLYFYVVEMLSVVFWGHGTVAYDTFVISICITIAFQLKTIAVSYISLNDKKGDI
KNLKDNDLEAMFNLKLLIQDQQNMFKKIKEIYKIFEPVTFVQLAAQSMLIILQAYMIFI
NHYNGFSLLSVPIIKLIVTVAPNIIHLFITCYLYTNININHQQDSMNFALYSSDWTAMSINYK
KMLLFTMRMNDAEKLKLKISLRKIVNLEMFASVMHLTYSIISVLAKSYGNTNTK**

>AglyOr14

**MNPNDENYIINLKLMKITGFYQLINPHTSKYLGFNVYKVGAAFEVMFGIISMLLLCLSS
YYYLDNTNELMSHFMLIVAIFFSIFKISWASKKSEMIWNNLDMTSINFLSYTGHKQEIL
QTARAKSISTTIIFVILWSSVTVAWSISPFFIKDVYLNVKFNDEIRRFRYNSLNYVYPITEE
SYNENFLYFYVVEMLQVIFWGHGTVAYDTFVISICISIAFQLKTIAVSYTSLNDIKGDIKN
LKHNDLEAILNLKLVIQDQQKMFKKIKEIYKIFQPVTFVQLVAQSMLIILQAYMIFINYY
NGFSLLSVPIIKLIVTVAPNIIHLFITCYLYSNINDQKDSMNFALYSGDWTAMSIKYKNML
LFAMRMNDAEKLKLKISLRKIVNLEMFANVMHLTYRIISVLSKSYGNTKSK**

>AgosOr23

**MNPNDENYIINLKLMKITGFYQLINPHTSKYLGFNVYKVGAGLEVMFGIISMLLLFLSS
YYYLDNTNELMSHFMLIVAIFFSTFKISWVSKKSEMIWNNLDMTSINFLSYTGHKQEIL
QTARAKSISTTIIFVILWSSVTVAWSISPFFIKDVYLNVKFNDEIRRFRYNSLNYVYPITEE
SYNENFLYFYVVEMLQVIFWGHGTVAYDTFVISICISIAFQLKTIAVSYTSLNDIKGDIKN
LKHNDLEAILNLKLVIQDQQKMFKKIKEIYKIFQPVTFVQLAAQSMLIILQAYMIFINH
YNGFSLLSVPIIKLIVTVAPNIIHLFITCYLYSNINDQKDSMNFALYSCDWTAMSIKYKNM
LLFAMRMNDAEKLKLKISLRKIVNLEMFASVMHLTYSIISVLSKSYGNTKSK**

>DnoxOr23

**MILNDEQNYLINLKLMKITGFYQLIHPRTTKYFGSNAYNAVAAIEVMAGVFSISLLFLSS
YYYLDNTNELMNHFMLIVAIFFSTFKIFWVSKNSKRIWNNMDITSINFLTYTGHKREIL
HNGRAKSIFTTILFVILWSSVTVAWSISPFLIKDVYLNVKFKDGIRRLRYNSLNYVYPISE
EFYNGHFLYFYVIELLQVILWGHGTVAYDTFVISICISIAFQLKTIAVSYISLNDRKCDEK
NFKDDDLEAMFNLKLLIQDQQNMFKKIKKIYQIFQPVTYVQLAAQSMLIIFQAYMIFIN
YYNGFSLISVPILKLVVTVAPNIMHLFTTCYLYSNINYQKDSMNFALYSSDWTAMSINYK
KMLLFAMRMNDAEKLKQKISLRKIVNLEMFASVMHLTYSIISVLAKSYGKTNTK**

>McerOr23

**MDLNDEQNYIINLKLMKITGFYQLINPCTPKHFGFNVFKVAAAIEVMTGVISVSLLFSSS
YYYLNNTNELMSHFMLVVAIFFSTFKIFWVSRNSETIWNNMDMTSINFLSYTGHKKEIL
QNARAKSISTTILFIILWSSVTVAWCISPFFVKDVYLNVKFKDDEIRQFRYNSLNYVYPIS
GEFYNEHFLYFYVVEMLQVVFWGHGTVAYDTFVISICISIAFQLKTIAVSYISLNDRKCD
IKHFEDNDLEAMFNLKLLIQDQQNIKIKEIYKIFQPVTYVQLAAQSMLIILQAYMIFINH
YNGFSLLSVPIIKLVVTVAPNIIHLFITCYLYSDINYQKDSMNFALYSSDWTAMSISYKKM
LLFTMRMNDAEKLKLKISLRKIVNLEMFASVMHLTYSIISVLAKSYGNTNTK**

>MperOr23

**MNFNDEKNYIFNLRLMKITGFYQLIYPSAPKCFGFNAYKVAAAIEVMTGVLSVSFLFSS
SYYYLDNTNELMSHFMLVVAIFFSTFKIFWVSRNSKTIWNNLDMTSINFLSYTGHKKEI
LQNARAKSISTTILFVILWSSVTVAWCISPFFVKDVYLNVKFKDDEIRRFRYNSLNYVYPI
SGESYNEHFLYFYVVEMLQVVFWGHGTVAYDTFVISICISIAFQLKTIAVSYISLNDRKG
DVKNLKDNDLEAIFNLKLLIQDQQNMFKKIKEIYKIFQPVTYVQLAAQSMLIILQAYMI
FINHYNGFSLLSVPIIKLVVTVAPNIIHLFITCYLYSDINYQKDSMNFALYSSDWTAMSISY
KKMLLFTMRMNDAEKLKLKISLRKIVNLEMFASVMHLTYSIISVLAKSYGNTNTK**

>RmaiOr23

**MTLDDEQNYIVNLKLMKITGFYQLINPGTPKYFGFNVYKIGAAIEVMSGIISILLLCSSS
YYYLDNTNELMSHFMLVVAIFFSTFKISWVSRNSETIWNNLDMTSVDFLSYTGHKKET
LQIARAKSISTTILFVILWSSVTVAWSISPFFVRDVYLNVKFNDEIRRFRYNSLNYVYPINE
ESYNKHFLYFYVVEMLQVVFWGHGTVAYDTFVISICISIAFQLKTIAVSYTSLNDIKRDV
KNLTHNDLEAIFNLKLLIQDQQNMFKKIKETYKIFQPVTFVQLAAQSMLIILQAYMIFI
NHFNGFSLLSVPIIKLVVTVAPNIIHLFITCYLYHNINDQKESMNFALYSSDWTAMSIKYK
NMLLFIMRMNDAEKLKLKISLRKIVNLEMFASVMHLTYSIISVLAKSYGNIKSK**

>SmisOr23

**MNLNDEQNYIVNLKLMKITGFYHLINPRAPKYFGFNVYKVTAALEVMTGLFSITMLFL
SSYYYLDNTNELTSHFMLVVAIFFSTLKIFWVSRNSETIWNNMDMTCINFLSYTGHKKE
ILKKARAKSMSTTILFVILWSSVTVAWCISPFFVKDVYLDVKFKDETRRFRYNSLNYVY
PISEEFYNQHFLYFYVIEMLSVVFWGHGTVAYDTFVISICITIAFQLKTIAVSYISLNGRK
GDIKNFKDNDLEAMFNLKLLIQDQQNMFKKIKEIYKIFEPVTFVQLAAQSMLIILQAY
MIFINHYNGYSLLSVPIIKLVVTVAPNIIHLFITCYLYTNINHQQDSMNFALYSGDWTAM
NINYKKMLLFTMRMNDAEKLKLKISLRKIVNLEMFASVMHLTYSIISVLAKSYGNTNT
K**

Figure S1 Ten motifs predicted from eight Or23 clade in aphids

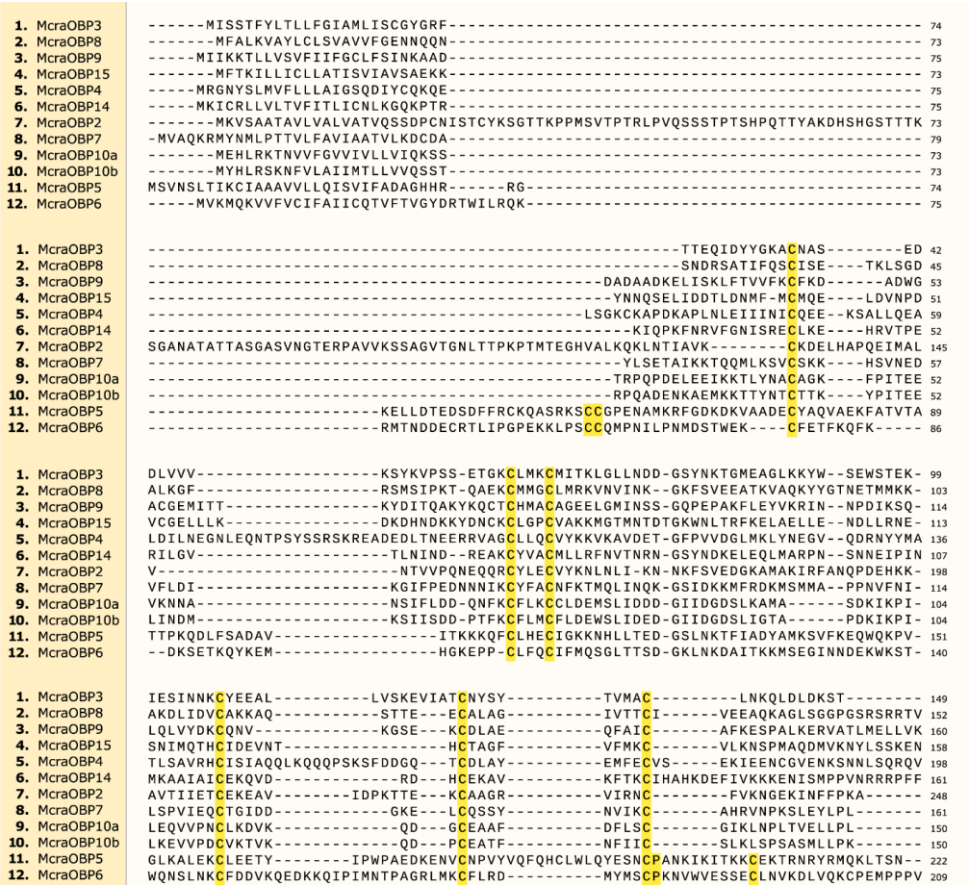


Figure S2. Sequence alignment of odorant binding proteins. The full-length amino acid sequences were aligned by DNAMAN v8 and then edited using SnapGene v4.3.6. The conserved amino acid site was marked by yellow shadow.

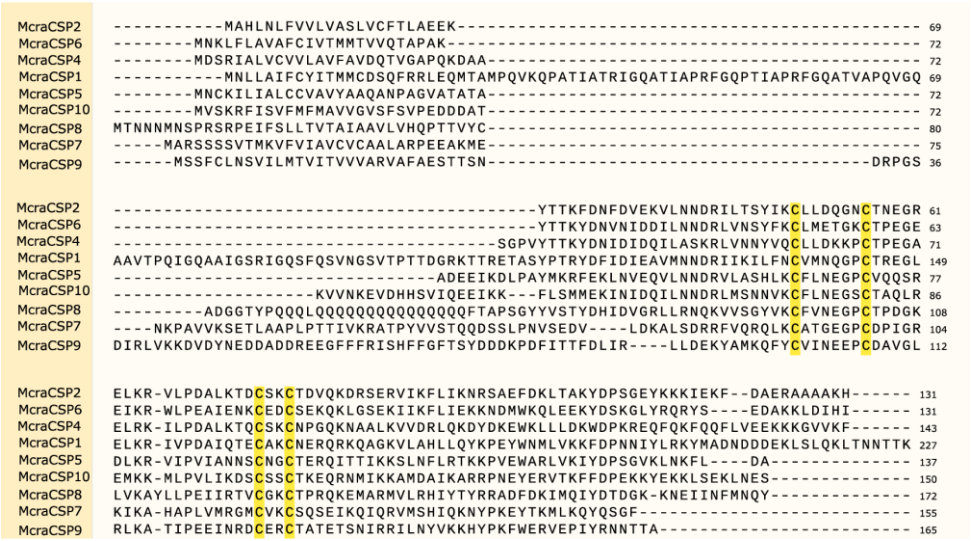
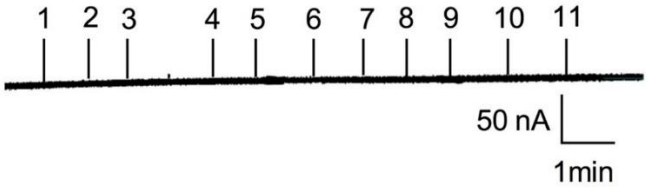


Figure S3. Sequence alignment of chemosensory proteins. The full-length amino acid sequences were aligned by DNAMAN v8 and then edited using SnapGene v4.3.6. The conserved amino acid site was marked by yellow shadow.

Figure 3. Expression profiles of ORs in *M. crassicauda.*



1. (-)-*trans*-caryophyllene
2. *cis*-jasmone
3. *trans*-2-hexenal
4. ocimene
5. linalool
6. *cis*-3-hexen-1-ol
7. *trans*-2-hexen-1-ol
8. α-pinene
9. DMNT
10. methyl salicylate
11. TMTT

Figure S5. Functional characterization of McraOR43/Orco in *Xenopus* oocytes. Inward current responses of McraOR43/Orco to tested HIPVs (1×10^{-4} mol/L).

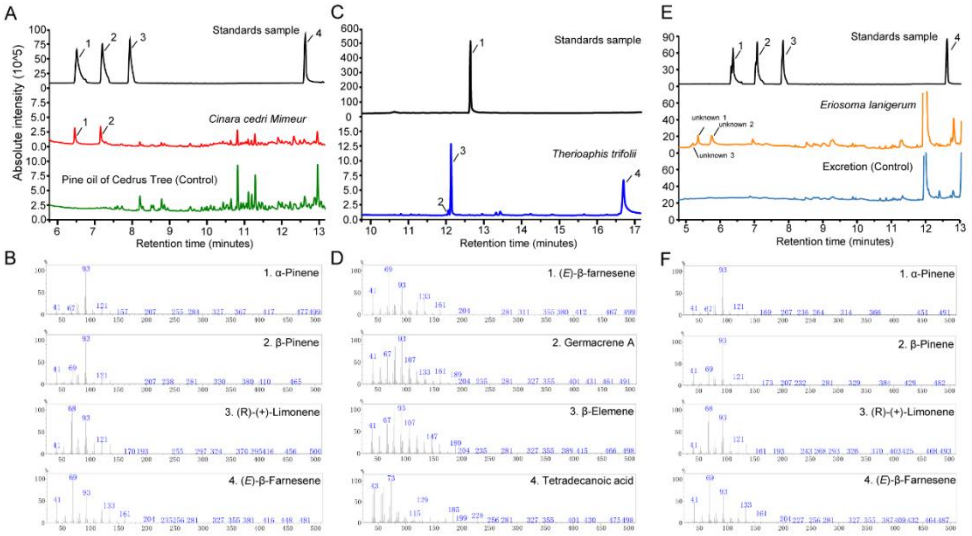


Figure S6. Gas chromatography - mass spectrometry analysis of the alarm pheromone components of *Therioaphis trifolii*, *Eriosoma lanigerum* and *Cinara cedri*. (A) Black: authentic sample of α-pinene, β-pinene, (*R*)-(+)-limonene, and (*E*)-β-farnesene (EBF); red: gas chromatogram trace from an extract of *C. cedri*; green: gas chromatogram trace from an extract of pine oil of cedrus tree (as control). (B) Mass spectra of α-pinene, β-pinene, (*R*)-(+)-limonene, and EBF. (C) Black: authentic sample of EBF; blue: gas chromatogram trace from an extract of *T. trifolii*. (D) Mass spectra of EBF, germacrene A, β-elemene, and

tetradecanoic acid. (E) Black: authentic sample of α-pinene, β-pinene, (*R*)-(+)-limonene, and EBF; orange: gas chromatogram trace from an extract of *E. lanigerum*; dark blue: gas chromatogram trace from an extract of the excretion on the body surface of *E. lanigerum* (as control). (F) Mass spectra of α-pinene, β-pinene, (*R*)-(+)-limonene, and EBF.

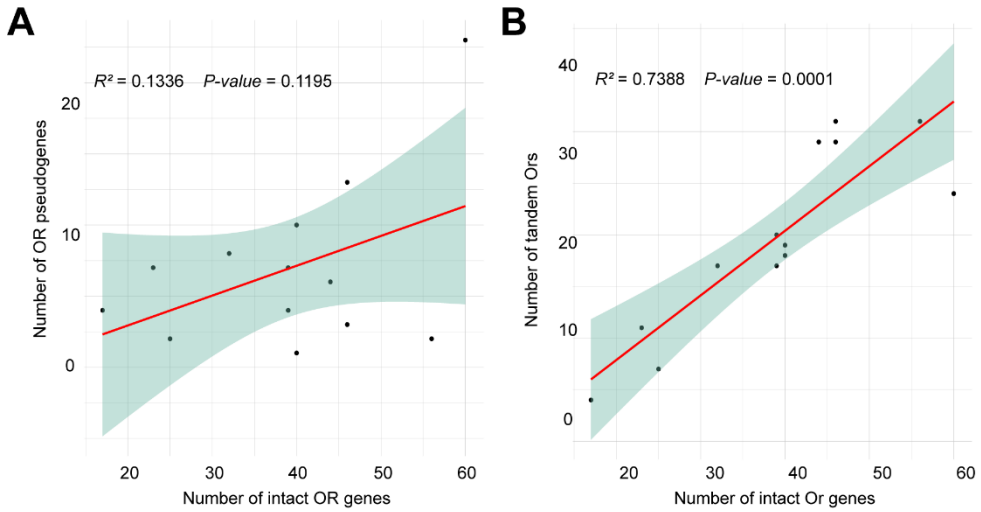


Figure S7. Correlation between the number of different types of OR genes in 13 aphids. (A) Number of intact OR genes vs number of OR pseudogenes. (B) Number of intact OR genes vs number of tandem OR genes.

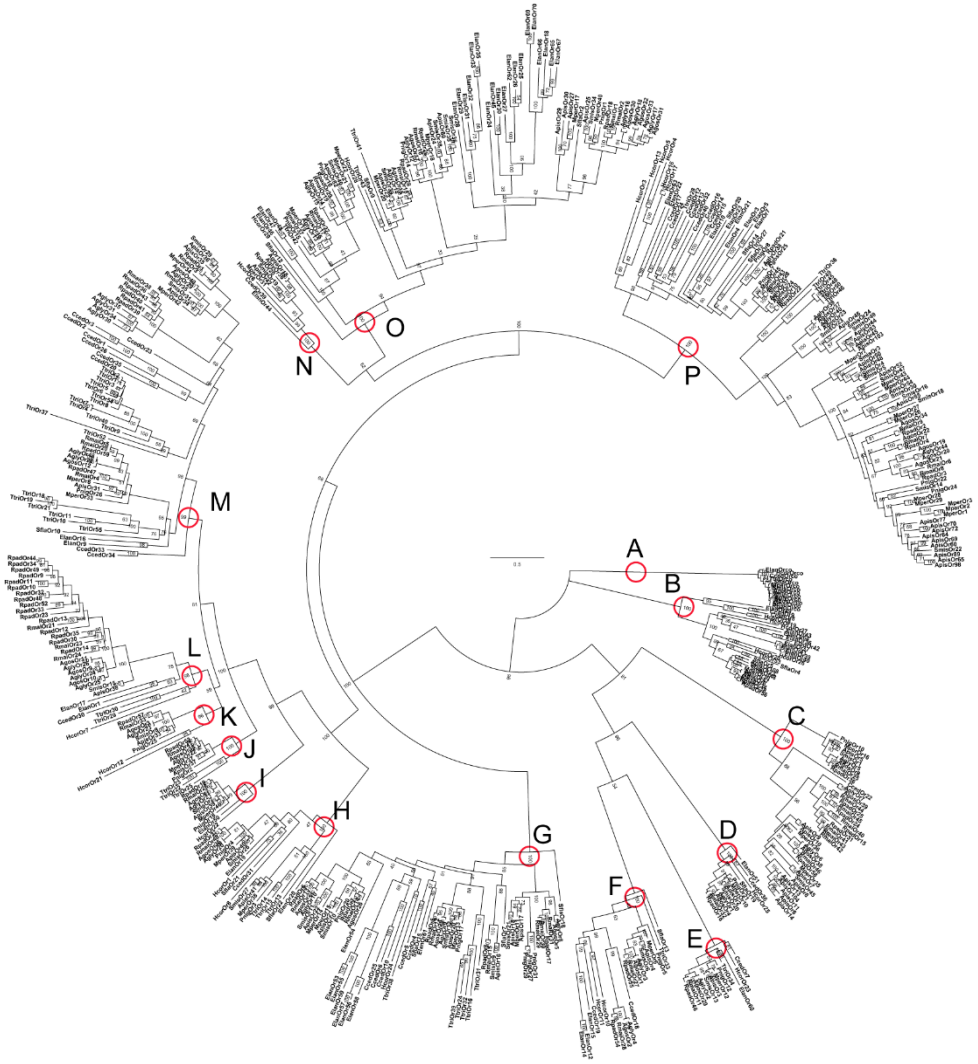


Figure S8. Phylogenetic tree of 532 intact ORs. All the ORs were classified into 16 subfamilies named from A to P, based on high bootstrap support (>80%).

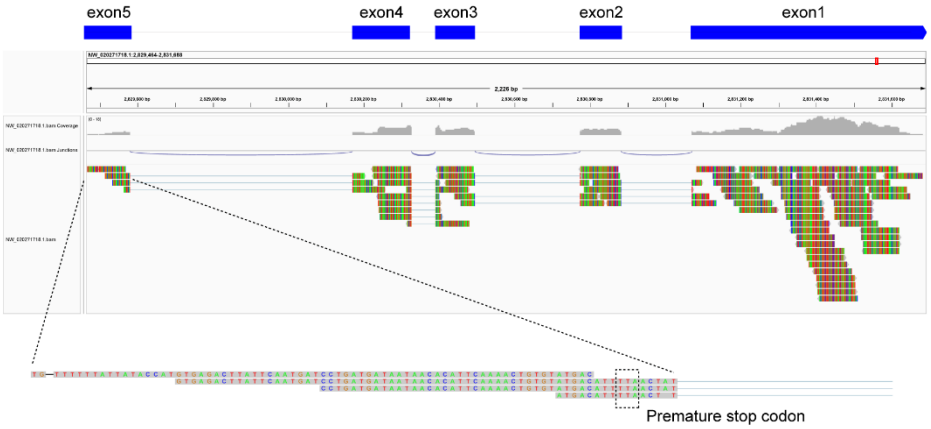


Figure S10. Verifying pseudogenization of SflaOR1 using transcriptomic data. Transcriptomic reads were mapped onto the gene model of SflaOR1. A premature stop code (TAA) was detected on the fifth exon of SflaOR1. The reverse complementary sequence of this premature stop code is shown in a dashed box.

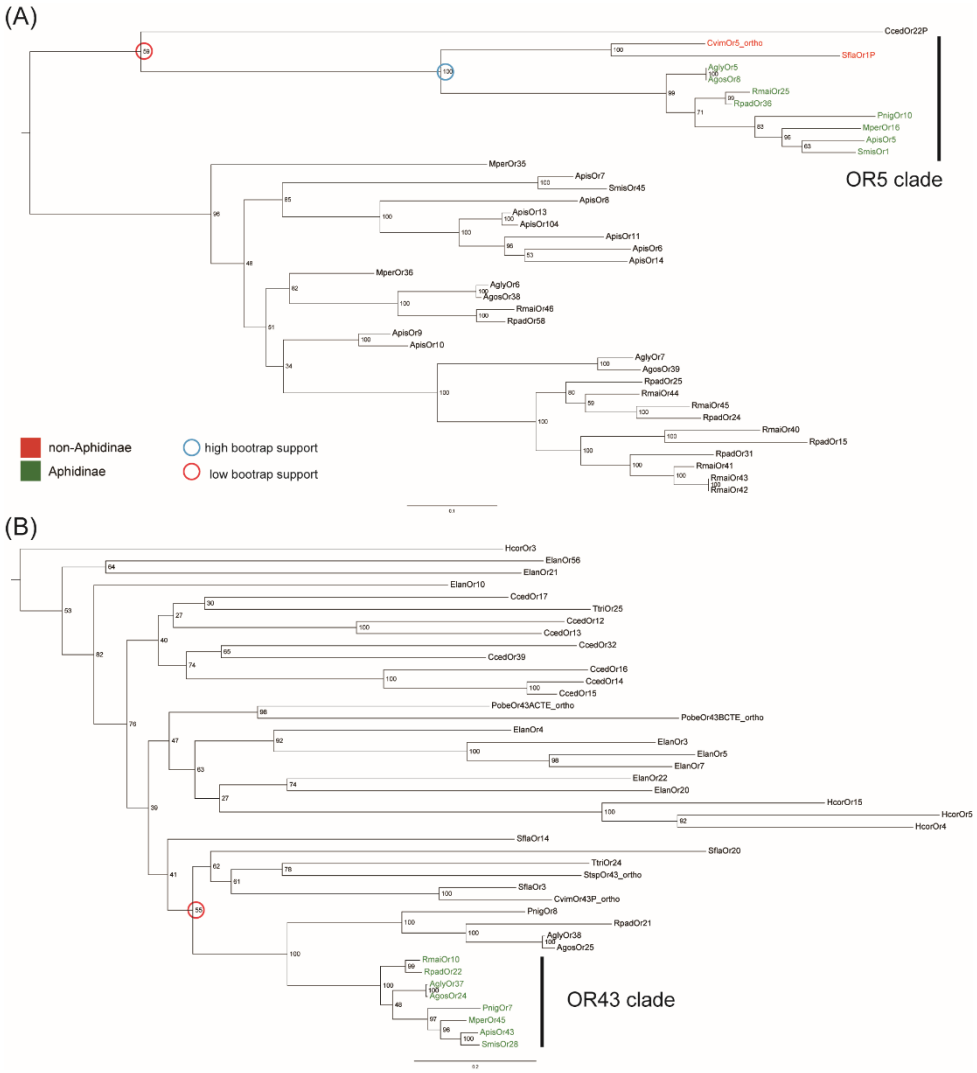


Figure S11. Classification of OR5 and OR43 clade based on bootstrap support. (A) Phylogenetic tree of OR subfamily F. Members of OR5 clade are identified in Aphidinae and Chaitophorinae species based on high bootstrap support (100%). (B) Phylogenetic tree of OR subfamily P (in part). Members of OR43 clade are only identified in Aphidinae species based on high bootstrap support (100%). Low bootstrap support (<60%) are indicated in light blue circle, high bootstrap value support (100%) is indicated in red circle.

Table S1. Primers used in this study

Primer name	Sequence (5´- 3´)
Gene cloning	
ApisOr23-F	ATGAATCTCAATGATGAGCAAAAC
ApisOr23-R	TTATTTAGTATTAGTATTTCCATAACTTTTTG
Expression vector construction	
ApisOr23-F	TCA**ACTAGT**GCCACCATGAATCTCAATGATGAGCAAA (**SpeI**)
ApisOr23-R	TCA**CTCGAG**TGAACTTATCAGGTGATAGAAACC (**XhoI**)
RT-PCR	
ApisOr23-F	AAATTACTGGTTTCTATCACCTGATAAGTCC
ApisOr23-R	GGCGTGTTTCATCTTTGAACTTAAT
ApisSDHBRT-F	CAGAAACTCCCGAAGTGAAGC
ApisSDHBRT-R	TAATCCAACGATACGCCTGC

Note: The restriction enzyme sites are marked with bold fronts; Kozak sequences are marked with underline.

Table S2. All 57 chemicals used for functional characterization of ApisOr23

Number	chemical name	CAS
1	2-Phenylethanol	60-12-8
2	cis-3-Hexen-1-ol	928-96-1
3	β-Citronellol	106-22-9
4	Geraniol	106-24-1
5	(1S)-(−)-Verbenone	1196-01-6
6	1-Hexanol	111-27-3
7	(S)-cis-Verbenol	18881-04-4
8	trans-3-Hexen-1-ol	928-97-2
9	3,7-Dimethyl-3-octanol	78-69-3
10	(−)-Borneol	464-45-9
11	(+)-Borneol	464-43-7

12	(1R)-(−)-Myrtenol	19894-97-4
13	(−)-trans-Pinocarveol	547-61-5
14	(−)-Linalool	126-91-0
15	Linalool	78-70-6
16	Methyl benzoate	93-58-3
17	Myrcene	123-35-3
18	(R)-(+)-Limonene	5989-27-5
19	α-Pinene	80-56-8
20	(−)-β-Pinene	18172-67-3
21	Camphene	79-92-5
22	α-Humulene	6753-98-6
23	(S)-(−)-Limonene	5989-54-8
24	α-Terpinene	99-86-5
25	(−)-trans-Caryophyllene	87-44-5
26	Farnesene, mixture of isomers	
27	cis-2-Hexen-1-ol	928-94-9
28	1-Heptanol	111-70-6
29	4′-Ethylacetophenone	937-30-4
30	Hexyl acetate	142-92-7
31	trans-2-Hexen-1-al	6728-26-3
32	4-Ethylbenzaldehyde	4748-78-1
33	3-Vinylbenzaldehyde	19955-99-8
34	(1R)-(−)-Myrtenal	18486-69-6
35	Benzaldehyde	100-52-7
36	Heptanal	111-71-7
37	cis-3-Hexenyl acetate	3681-71-8
38	trans-2-Hexenyl acetate	2497-18-9
39	cis-3-Hexenyl acetate	3681-71-8
40	1,4-Diethylbenzene	105-05-5
41	Tetradecane	629-59-4

42	Nonyl acetate	143-13-5
43	Ocimene	13877-91-3
44	Tridecane	629-50-5
45	Methyl salicylate	119-36-8
46	(±)-Camphor	76-22-2
47	Nerolidol	7212-44-4
48	2,6-Di-tert-butylphenol	128-39-2
49	2-Pentadecanone	2345-28-0
50	Acetophenone	98-86-2
51	(+)-Cedrol	77-53-2
52	Ethyl butyrate	105-54-4
53	Ethyl hexanoate	123-66-0
54	Salicylaldehyde	90-02-8
55	Methyl 2-methoxybenzoate	606-45-1
56	Eugenol	97-53-0
57	Methyl phenylacetate	101-41-7

Table S3. Primers used in this study

Primer name	Sequence (5´- 3´)
Specific primers for gene cloning	
McraOrco-F	ATGGCTTATAAGAAAGACGGTCTTATAA
McraOrco-R	TTATTTAAGCTGCACCAAAACCAT
McraOR43-F	ATGGATTTAAAACAAGAAAAACAATACATT
McraOR43-R	TTATTTATTTTTATTATAATTAATTAACATTGAAA
McraOR20-F	ATGCGGTCATCGTCAGCGAC
McraOR20-R	TTAAGATCTTGAATTTAGTAACACCGAAACT
ApisOR20-F	ATGCGGTCATCGTCAGCGACA
ApisOR20-R	TTAAGATCTTGAATTTAGTAACACGGAAACTGT
ApisOrco-F	ATGGGTTATAAGAAAGATGGTCTTAT
ApisOrco-R	TTATTTAAGCTGCACCAAAACC

McraOrco-SF	**GCAGATCTGATATCACTAGT***GGGCCC*GCCACCATG GCTTATAAGAAAGACGGTCTTATAA
McraOrco-SR	**TAACCAGATCCTAGTCAGTC***GCGGCCGC*TTATTTAA GCTGCACCAAAACCAT
McraOR20-SF	**GCAGATCTGATATCACTAGT***GGGCCC*GCCACCATG CGGTCATCGTCAGCGAC
McraOR20-SR	**TAACCAGATCCTAGTCAGTC***GCGGCCGC*TTAAGAT CTTGAATTTAGTAACACCGAAACT
McraOR43-SF	**GCAGATCTGATATCACTAGT***GGGCCC*GCCACCATG GATTTAAAACAAGAAAAACAATACATT
McraOR43-SR	**TAACCAGATCCTAGTCAGTC***GCGGCCGC*TTATTTAT TTTTATTATAATTAATTAACATTGAAA
ApisOR20-DF	TCA*ACTAGT*GCCACCATGCGGTCATCGTCAGCGACAG TAGTG
ApisOR20-DR	TCA*CTCGAG*TTAAGATCTTGAATTTAGTAACACGGAA ACTGT
ApisOrco-DF	TCA*ACTAGT*GCCACCATGGGTTATAAGAAAGATGGTC
ApisOrco-DR	TCA*CTCGAG*TTATTTAAGCTGCACCAAAA

Primers for RT-PCR

McraOrco-RT-F	CACGGGTACCAAGTCGTTCT
McraOrco-RT-R	CGGGTTCAAAATCTCCTTCA
McraOR2-RT-F	GTTCGTTTTCGTTCTTCGTTAGC
McraOR2-RT-R	TGTAACGGCGTCTGTGTCCTG
McraOR4-RT-F	TTGCATCCAAATGTGGCTTA
McraOR4-RT-R	TCAATAATTGTTTGATGGTGACG
McraOR5-RT-F	TGGGAATGTTTGTCGGTCAC
McraOR5-RT-R	AAAAGGTAAAACATCGTCAAAGC
McraOR7-RT-F	TGTACGGGATGGAATGACAA
McraOR7-RT-R	TCGGAAACACATCCACGTAA
McraOR9-RT-F	TTAGTTGTTTGGCGGGTTTC
McraO9-RT-R	CAAGCCGTTTCTTTAAATTTGC
McraOR17-RT-F	CTCGGAAGAGCTAACCATCG
McraOR17-RT-R	TGTGATCGGCATTCATCACT
McraOR18-RT-F	TTTTGAAATTATTCAACCTGTCG
McraOR18-RT-R	TCATTTCTGGACTCGTGAATTT
McraOR20-RT-F	TTCGTTCCAGGTGCCTTATC
McraOR20-RT-R	GCGACCATATGAACACGTTG
McraOR23-RT-F	TCTTCTGTGACCGTTGCTTG
McraOR23-RT-R	TTGAGCAGCTAGTTGAACAAATG
McraOR26-RT-F	TCTGGGCTCATGCTACATTG
McraOR26-RT-R	TAGTTGCGAACGTTGTCTGC
McraOR28-RT-F	TGCCTGGACGACATAAATGA
McraOR28-RT-R	CGCTGTACATTCGATGGTGT

McraOR31-RT-F	CTGGCTGTAGCGTTTGAGAA
McraOR31-RT-R	TTTTGCCGTACAGTTCATCG
McraOR37-RT-F	TGATCGTTCCGTTGGTGTTA
McraOR37-RT-R	ATTCCGCAGACGAACAAATC
McraOR39-RT-F	TATTAGACCCGAAGGGACCG
McraOR39-RT-R	TATTAGACCCGAAGGGACCG
McraOR43-RT-F	TGGAAATGCATTGACGTTACA
McraOR43-RT-R	ATTTCGTGGCCAAGTGATTT
McraOR51-RT-F	TGGGAATGTTTGTCGGTCAC
McraOR51-RT-R	AAAAGGTAAAACATCGTCAAAGC
McraOR54-RT-F	TTAACATTGCATTGGCGAAA
McraOR54-RT-R	AACCGTCGTGATTTTTGACC
McraOR82-RT-F	GCAAAGCAACGTCAAATCAA
McraOR82-RT-R	TGGGGGTAGAACCAATAACG
McraOR83-RT-F	CCGTAAAATATCGGGGAAAAA
McraOR83-RT-R	CCCTACTGAACCGTCGTGAT
McraNADH-RT-F	CGAGGGCTGGTAATACAGTTTTG
McraNADH-RT-R	GTTGTCCGACGACTTCCATCA

Note: The KOZAK sequence is highlighted in underline, the homologous arms of the expression vector are highlighted in bold, the restriction enzyme of each forward primer is ApaI, the reverse is NotI and the cutting sites are in italics. F: Forward primer; R: Reverse primer. SF: The forward primer for seamless cloning; SR: The reverse primer for seamless cloning. DF: The forward primer for double digestion; DR: The reverse primer for double digestion. RT-F: The forward primer for RT-PCR; RT-R: The reverse primer for RT-PCR.

Table S4 Herbivore-induced plant volatiles used in the functional study

Number	Name	CAS	Manufacturer	purity
1	(-)-*trans*-caryophyllene	87-44-5	Sigma-Aldrich	90%
2	*cis*-jasmone	488-10-8	Sigma-Aldrich	≥95%
3	*trans*-2-hexenal	6728-26-3	Sigma-Aldrich	98%
4	ocimene	13877-91-3	Sigma-Aldrich	≥90%
5	linalool	78-70-6	Sigma-Aldrich	97%
6	*cis*-3-hexen-1-ol	928-94-9	Sigma-Aldrich	≥95%
7	*trans*-2-hexen-1-ol	928-95-0	Sigma-Aldrich	96%
8	α-pinene	80-56-8	Sigma-Aldrich	98%

9	DMNT	19945-61-0	J&K SCIENTIFIC LTD	≥95%
10	methyl salicylate	119-36-8	Sigma-Aldrich	≥99%
11	TMTT	62235-06-7	J&K SCIENTIFIC LTD	≥98%

Table S5. Assembly summary of antennal transcriptome from *M. crassicauda*

	Transcripts Number	Total Length (nt)	Mean Length (nt)	N50 (nt)	N70 (nt)	N90 (nt)	GC (%)
Mcra_Al_1	15,566	38,402,199	2,467	3,145	2,300	1,418	34.34
Mcra_Al_2	16,765	46,666,454	2,783	3,546	2,603	1,591	34.04
Mcra_Al_3	18,047	44,951,453	2,490	3,232	2,370	1,404	34.44
All-Unigene	15,984	47,860,781	2,994	3,711	2,746	1,720	34.44

Table S6. Unigenes of candidate odorant receptors in *M. crassicauda*

Unigene reference	Gene name	Length (nt)	ORF (aa)	Blastx best hit (Reference/Name/Species)	Evalue	Identity	Full length
Unigene8980	McraOrco	2974	463	XP_022162891.1| odorant receptor coreceptor isoform X1 [Myzus persicae]	0	97%	Yes
Unigene5450	McraOR2	648	194	XP_003245950.2| odorant receptor 22c-like [Acyrthosiphon pisum]	1.00E-133	97%	No
CL2070.Contig1	McraOR4	1275	260	AQS60743.1| olfactory receptor 4 protein [Acyrthosiphon pisum]	3.00E-126	84%	No
Unigene7191	McraOR5	1510	369	XP_022163466.1| odorant receptor 43b-like [Myzus persicae]	0	86%	Yes
CL773.Contig2	McraOR7	3323	369	XP_022174880.1| odorant receptor 43b-like [Myzus persicae]	2.00E-134	56%	Yes
Unigene5223	McraOR9	626	137	XP_016660447.1| odorant receptor 43b [Acyrthosiphon pisum]	4e-59	75%	No
Unigene6603	McraOR17	2767	304	XP_016662512.2| odorant receptor 2a-like [Acyrthosiphon pisum]	0	85%	No
Unigene5512	McraOR18	942	140	AQS60746.1| olfactory receptor 17 protein [Acyrthosiphon pisum]	1e-82	89%	No

Unigene710	McraOR20	1893	420	XP_016657950.2| odorant receptor 2a-like [Acyrthosiphon pisum]	0	89%	Yes
Unigene8085	McraOR23	1490	313	AQS60748.1| olfactory receptor 23 protein [Acyrthosiphon pisum]	0	90%	No
CL2087.Contig2	McraOR26	4096	390	XP_025413669.1| odorant receptor 46a-like isoform X3 [Sipha flava]	2.00E-67	40%	No
Unigene14982	McraOR28	1411	305	XP_025413669.1 | odorant receptor 46a-like isoform X3 [Sipha flava]	3.00E-85	42%	No
Unigene5096	McraOR31	1181	214	AQS60750.1| olfactory receptor 31 protein [Acyrthosiphon pisum]	1.00E-92	77%	No
Unigene10281	McraOR37	1423	416	AQS60751.1| olfactory receptor 37 protein [Acyrthosiphon pisum]	0	89%	Yes
Unigene7723	McraOR39	969	268	KAF0748431.1| odorant receptor 67a [Aphis craccivora]	1.00E-163	83%	No
Unigene636	McraOR43	1654	420	XP_025195448.1| odorant receptor 4-like [Melanaphis sacchari]	0	85%	Yes
Unigene13595	McraOR51	4157	432	XP_027840127.1| odorant receptor 46a-like [Aphis gossypii]	5.00E-81	35%	Yes
Unigene7172	McraOR54	2791	252	XP_026806824.1| odorant receptor 46a-like [Rhopalosiphum maidis]	5.00E-33	33%	No
Unigene8086	McraOR82	1879	275	XP_025412107.1| odorant receptor 49b-like [Sipha flava]	2.00E-48	38%	No
Unigene5231	McraOR83	1551	428	XP_027840127.1| odorant receptor 46a-like [Aphis gossypii]	3.00E-79	38%	Yes

Table S7. Unigenes of candidate gustatory receptors in *M. crassicauda*

Unigene reference	Gene name	Length (nt)	ORF (aa)	Blastx best hit (Reference/Name/Species)	Evalue	Identity	Full length
Unigene3161	McraGR1	2854	419	XP_022180783.1| gustatory receptor for sugar taste 64f-like isoform X1 [Myzus persicae]	0	96%	Yes
Unigene13552	McraGR5	736	202	XP_022161088.1| gustatory receptor for sugar taste 64f-like isoform X3 [Myzus persicae]	6.00E-131	94%	No
Unigene9894	McraGR20	669	223	KAF0773181.1| Gustatory receptor [Aphis craccivora]	2.00E-110	75%	No
CL4354.Contig3	McraGR21	1436	444	KAF0762692.1| Gustatory receptor, partial [Aphis craccivora]	0	74%	No

Table S8. Unigenes of candidate ionotropic receptors in *M. crassicauda*

Unigene reference	Gene name	Length (nt)	ORF (aa)	Blastx best hit (Reference/Name/Species)	Evalue	Identity	Full length
CL2432.Contig7	McraIR8a	5427	272	XP_029345195.1| ionotropic receptor 25a [Acyrthosiphon pisum]	0	97%	No
Unigene849	McraIR25a	4351	939	XP_022161203.1| ionotropic receptor 25a [Myzus persicae]	0	99%	Yes
CL4903.Contig1	McraIR40a	2589	635	XP_022182311.1| ionotropic receptor 40a isoform X2 [Myzus persicae]	0	98%	Yes
Unigene16952	McraIR75d1	2248	525	XP_029347273.1| glutamate receptor ionotropic, delta-1-like [Acyrthosiphon pisum]	0	93%	No
Unigene20241	McraIR75d3	1045	314	XP_016663076.2 | glutamate receptor ionotropic, delta-1-like [Acyrthosiphon pisum]	0	89%	No
Unigene9857	Mcra75j	1732	405	AKI28987.1| glutamate receptor ionotropic, delta-2 [Aphis craccivora]	0	89%	No
Unigene8539	McraIR93a	1752	444	XP_029346904.1| ionotropic receptor 93a [Acyrthosiphon pisum]	0	96%	No
Unigene2162	McraIR323	605	311	XP_022176917.1 | probable glutamate receptor [Myzus persicae]	3.00E-101	90%	No
Unigene11517	McraIR325	1312	335	KAF0758656.1| ionotropic receptor 68a, isoform B [Drosophila melanogaster]	2.00E-28	50%	No

Table S9. Unigenes of candidate odorant binding proteins in *M. crassicauda*

Unigene reference	Gene name	Length (nt)	ORF (aa)	Blastx best hit (Reference/Name/Species)	Evalue	Identity	Signal peptide	Full length
Unigene11074	McraOBP2	1895	223	CAR85651.1| odorant-binding protein 2, partial [Megoura viciae]	2.00E-126	100%	19	Yes
Unigene3758	McraOBP3	1027	141	AXE72019.1| OBP3 [Megoura viciae]	1.00E-96	99%	23	Yes
Unigene8284	McraOBP4	1025	199	AXE72024.1| OBP4 [Megoura viciae]	4.00E-141	100%	22	Yes

Unigene3875	McraOBP5	1363	221	AXE72025.1| OBP5 [Megoura viciae]	6.00E-17	99%	25	Yes
CL3364.Contig 2	McraOBP6	1292	218	AXE72020.1| OBP6 [Megoura viciae]	3.00E-154	99%	-	Yes
Unigene566	McraOBP7	894	155	AXE72021.1| OBP7 [Megoura viciae]	3.00E-108	100%	30	Yes
Unigene11483	McraOBP8	765	157	AXE72022.1| OBP8 [Megoura viciae]	5.00E-103	100%	18	Yes
CL4330.Contig 2	McraOBP9	834	166	AXE72026.1| OBP9 [Megoura viciae]	2.00E-106	99%	24	Yes
CL879.Contig3	McraOBP10	747	144	AXE72027.1| OBP10 [Megoura viciae]	3.00E-77	100%	25	Yes
CL3873.Contig 5	McraOBP12	947	144	NP_001153525.1| odorant-binding protein 10 precursor [Acyrthosiphon pisum]	2.00E-46	66%	24	Yes
Unigene8175	McraOBP13	1555	335	XP_029342020.1| general odorant-binding protein 71 isoform X1 [Acyrthosiphon pisum]	7.00E-72	97%	30	Yes
CL1874.Contig 2	McraOBP14	678	175	APB03436.1| odorant-binding protein 15 [Sitobion avenae]	5.00E-105	86%	21	Yes
Unigene11512	McraOBP15	749	167	APB03435.1| odorant-binding protein 14 [Sitobion avenae]	6.00E-87	85%	21	Yes

Table S10. Unigenes of candidate chemosensory proteins in *M. crassicauda*

Unigene reference	Gene name	Length (nt)	ORF (aa)	Blastx best hit (Reference/Name/Species)	Evalue	Identity	Signal peptide	Full length
Unigene3562	McraCSP1	1549	227	APB03441.1| chemosensory protein 5 [Sitobion avenae]	2.00E-149	90%	-	Yes
CL1331.Contig3	McraCSP2	674	131	NP_001119651.1| chemosensory protein-like precursor [Acyrthosiphon pisum]	2.00E-83	95%	20	Yes
CL53.Contig3	McraCSP4	1139	143	NP_001119652.1| chemosensory protein-like precursor [Acyrthosiphon pisum]	2.00E-69	90%	22	Yes
Unigene635	McraCSP5	843	137	APB03439.1| chemosensory protein 3 [Sitobion avenae]	5.00E-76	94%	19	Yes
CL3361.Contig4	McraCSP6	790	131	ACJ64047.1| putative chemosensory protein CSP1 [Myzus persicae]	4.00E-60	89%	21	Yes
Unigene14722	McraCSP7	865	155	NP_001156200.1| chemosensory protein 1-like precursor [Acyrthosiphon pisum]	1.00E-103	96%	24	Yes

CL2068.Contig3	McraCSP8	1227	172	XP_022183495.1| putative odorant-binding protein A10 [Myzus persicae]	2.00E-89	79%	37	Yes
CL1331.Contig2	McraCSP9	1433	165	XP_016661164.1| uncharacterized protein LOC100162203	8.00E-81	77%	20	Yes
Unigene11647	McraCSP10	850	150	XP_022172561.1| putative odorant-binding protein A10 [Myzus persicae]	3.00E-69	77%	21	Yes

Table S11. Expression levels of candidate odorant receptors in *M. crassicauda*

Transcript Name	FPKM		Log_{10}(FPKM+1)	
	Antennae	Leg	Antennae	Leg
McraOrco	1411.22	3.51	3.15	0.65
McraOR2	27.08	0.00	1.45	0.00
McraOR4	213.38	37.28	2.33	1.58
McraOR5	39.58	1.05	1.61	0.31
McraOR7	3407.51	21.20	3.53	1.34
McraOR9	56.68	0.00	1.76	0.00
McraOR17	27.31	7.18	1.45	0.91
McraOR18	75.29	19.91	1.88	1.32
McraOR20	142.18	7.38	2.16	0.92
McraOR23	25.78	0.00	1.42	0.00
McraOR26	228.25	33.54	2.36	1.54
McraOR28	167.09	10.85	2.23	1.07
McraOR31	23.81	0.00	1.39	0.00
McraOR37	22.39	0.00	1.37	0.00
McraOR39	22.32	0.00	1.37	0.00
McraOR43	53.95	0.43	1.74	0.16
McraOR51	82.28	3.66	1.92	0.67
McraOR54	94.52	26.00	1.98	1.43
McraOR82	53.88	0.00	1.74	0.00
McraOR83	56.22	0.64	1.76	0.21

Table S13. Adaptive branch-site random effects likelihood (aBSREL) test on eight conserved clades and the Orco clade

Clade Name	**LRT**	**Tested P-values**	**Uncorrected P values**	**ω distribution over sites**
ApisOR20	23.95	0.000	0.000	ω1 = 0.274 (90%)
				ω2 = 19.0 (9.8%)
ApisOR3	8.784	0.035	0.004	ω1 = 0.0618 (95%)
				ω2 = 10.0 (5.1%)
ApisOR40	7.078	0.072	0.010	ω1 = 0.00 (63%)
				ω2 = 5.22 (37%)
ApisOR39	4.432	0.199	0.040	ω1 = 0.120 (98%)
				ω2 = 1560 (2.0%)
ApisOR5	4.468	0.234	0.0390	ω1 = 0.00 (66%)
				ω2 = 4.95 (34%)
ApisOR2	3.464	0.262	0.066	ω1 = 0.00 (92%)

				ω2 = 3.73 (8.2%)
ApisOrco	0.000	1.000	1.000	ω1 = 0.0588 (100%)
ApisOR4	0.000	1.000	1.000	ω1 = 0.305 (100%)
ApisOR43	0.000	1.000	1.000	ω1 = 0.407 (100%)

Table S14. Tested odorants in this study

No.	**Classification**	**Name**	**CAS**	**Purity (%)**	**Company**
1	Terpenoids	(*E*)-4,8-Dimethyl-1,3,7-nonatriene	19945-61-0	95	J&K Scientific Ltd.
2		Ocimene	13877-91-3	90	Sigma
3		(*E*)-β-ocimene	3779-61-1	96	J&K Scientific Ltd.
4		6-Methyl-5-hepten-2-one	110-93-0	97.5	J&K Scientific Ltd.
5		(*S*)-(-)-Limonene	5989-54-8	96	Sigma
6		(*R*)-(+)-Limonene	5989-27-5	97	Sigma
7		(-)-β-Pinene	18172-67-3	99	sigma
8		α-Pinene	80-56-8	98	Sigma
9		Geranyl acetate	105-87-3	97	Sigma
10		*trans*-β-Farnesene	18794-84-8	90	Sigma
11		(*E,E*)-4,8,12-Trimethyltrideca-1,3,7,11-tetraene	62235-06-7	98	Sigma
12		Linalool	78-70-6	97	Sigma
13		(-)-*trans*-Caryophyllene	87-44-5	98.5	Sigma
14		a-Copaene	3856-25-5	98	J&K Scientific Ltd.
15		Nerolidol	7212-44-4	98	Sigma
16	Aromatics	Phenol	108-95-2	99	Sigma
17		Methyl salicylate	119-36-8	99	Sigma
18		Methyl eugenol	93-15-2	98	Sigma
19		2-Methyphenol	95-48-7	99	Sigma

20		Eugenol	97-53-0	99	Sigma
21		*p*-Anisaldehyde	123-11-5	98	Sigma
22		4-Methoxybenzyl alcohol	105-13-5	98	Sigma
23		Benzaldehyde	100-52-7	98	Sigma
24		*p*-Cresol	106-44-5	99	Sigma
25		Benzyl alcohol	100-51-6	98	Sigma
26		Phenylacetaldehyde	122-78-1	97.5	J&K Scientific Ltd.
27		Phenethyl alcohol	1960/12/8	99	Sigma
28		1,4-Dimethoxybenzene	150-78-7	98	Sigma
29		Methyl-2-methoxybenzoate	606-45-1	97	Sigma
30	Aliphatics	Hexanal	66-25-1	98	Sigma
31		*cis*-2-Hexen-1-ol	928-94-9	95	Sigma
32		*trans*-2-Hexen-1-ol	928-95-0	96	Sigma
33		2,3-Butanediol	513-85-9	98	J&K Scientific Ltd.
34		*cis*-3-Hexenyl acetate	3681-71-8	98	Sigma
35		*cis*-3-Hexen-1-ol	928-96-1	98	Sigma
36		*cis*-2-Nonen-1-ol	22104-79-6	97	Sigma
37		*trans*-2-Nonen-1-ol	31502-14-4	96	Sigma
38		*trans*-2-Nonenal	18829-56-6	97	Sigma
39		3-Methyl-2-butenal	107-86-8	97	Sigma
40		*trans*-2-Hexenal	6728-26-3	98	Sigma
41		*trans*-3-Hexen-1-ol	928-97-2	97	Sigma
42		3-Hydroxy-2-butanone	513-86-0	96	Sigma
43		1-Octen-3-ol	3391-86-4	98	Sigma
44		Nepetalactol	109215-55-6	97	J&K Scientific Ltd.
45		Nepetalactone	21651-62-7	97	J&K Scientific Ltd.
46	Heterocyclic derivatives	Indole	120-72-9	99	Sigma
47		Indole-3-aldehyde	487-89-8	98	J&K Scientific Ltd.
48		Methyl jasmonate	39924-52-	95	Sigma

			2		
49		*cis*-Jasmone	488-10-8	94	Sigma

Reference

Abuin, L., Bargeton, B., Ulbrich, M. H., Isacoff, E. Y., Kellenberger, S., & Benton, R. (2011). Functional architecture of olfactory ionotropic glutamate receptors. *Neuron, 69*(1), 44-60. doi:10.1016/j.neuron.2010.11.042

Adams, M. D., Celniker, S. E., Holt, R. A., Evans, C. A., Gocayne, J. D., Amanatides, P. G., Galle, R. F. (2000). The genome sequence of Drosophila melanogaster. *Science, 287*(5461), 2185-2195.

Allmann, S., & Baldwin, I. T. (2010). Insects betray themselves in nature to predators by rapid isomerization of green leaf volatiles. *Science, 329*(5995), 1075-1078. doi:10.1126/science.1191634

Almeida, F. C., Sánchez-Gracia, A., Campos, J. L., & Rozas, J. (2014). Family size evolution in Drosophila chemosensory gene families: a comparative analysis with a critical appraisal of methods. *Genome Biol Evol, 6*(7), 1669-1682. doi:10.1093/gbe/evu130

Anholt, R. R. H. (2020). Chemosensation and Evolution of Drosophila Host Plant Selection. *iScience, 23*(1), 100799. doi:10.1016/j.isci.2019.100799

Auer, T. O., Álvarez-Ocaña, R., Cruchet, S., Benton, R., & Arguello, J. R. (2022). Copy number changes in co-expressed odorant receptor genes enable selection for sensory differences in drosophilid species. *Nat Ecol Evol, 6*(9), 1343-1353. doi:10.1038/s41559-022-01830-y

Bai, Z., Ma, W., Ma, L., Velthof, G. L., Wei, Z., Havlík, P., Zhang, F. (2018). China's livestock transition: Driving forces, impacts, and consequences. *Science Advances, 4*(7), eaar8534. doi:10.1126/sciadv.aar8534

Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., Noble, W. S. (2009). MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res, 37*(Web Server issue), W202-208. doi:10.1093/nar/gkp335

Bansal, R., Mian, M. A., Mittapalli, O., & Michel, A. P. (2014). RNA-Seq reveals a xenobiotic stress response in the soybean aphid, Aphis glycines, when fed aphid-resistant soybean. *BMC Genomics, 15*(1), 972. doi:10.1186/1471-2164-15-972

Bao, W., Kojima, K. K., & Kohany, O. (2015). Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mob DNA, 6*, 11. doi:10.1186/s13100-015-0041-9

Bass, C., Puinean, A. M., Zimmer, C. T., Denholm, I., Field, L. M., Foster, S. P., Williamson, M. S. (2014). The evolution of insecticide resistance in the peach potato aphid, Myzus persicae. *Insect Biochemistry and Molecular Biology, 51*, 41-51. doi:10.1016/j.ibmb.2014.05.003

Bastin-Héline, L., de Fouchier, A., Cao, S., Koutroumpa, F., Caballero-Vidal, G., Robakiewicz, S., Montagné, N. (2019). A novel lineage of candidate pheromone receptors for sex communication in moths. *Elife, 8*. doi:10.7554/eLife.49826

Basu, S., Clark, R. E., Fu, Z., Lee, B. W., & Crowder, D. W. (2021). Insect alarm pheromones in response to predators: Ecological trade-offs and molecular

mechanisms. *Insect Biochem Mol Biol, 128*, 103514. doi:10.1016/j.ibmb.2020.103514

Bates, A. S., Schlegel, P., Roberts, R. J. V., Drummond, N., Tamimi, I. F. M., Turnbull, R., Jefferis, G. (2020). Complete Connectomic Reconstruction of Olfactory Projection Neurons in the Fly Brain. *Curr Biol, 30*(16), 3183-3199.e3186. doi:10.1016/j.cub.2020.06.042

Bayendi Loudit, S. M., Boullis, A., Verheggen, F., & Francis, F. (2018). Identification of the Alarm Pheromone of Cowpea Aphid, and Comparison With Two Other Aphididae Species. *J Insect Sci, 18*(1). doi:10.1093/jisesa/iex097

Beale, M. H., Birkett, M. A., Bruce, T. J. A., Chamberlain, K., Field, L. M., Huttly, A. K., Zhang, Y. (2006). Aphid alarm pheromone produced by transgenic plants affects aphid and parasitoid behavior. *Proc Natl Acad Sci U S A, 103*(27), 10509-10513. doi:10.1073/pnas.0603998103

Ben-Arie, N., Lancet, D., Taylor, C., Khen, M., Walker, N., Ledbetter, D. H., et al. (1994). Olfactory receptor gene cluster on human chromosome 17: possible duplication of an ancestral receptor repertoire. *Hum Mol Genet, 3*(2), 229-235. doi:10.1093/hmg/3.2.229

Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res, 27*(2), 573-580. doi:10.1093/nar/27.2.573

Benton, R. (2006). On the ORigin of smell: odorant receptors in insects. *Cell Mol Life Sci, 63*(14), 1579-1585. doi:10.1007/s00018-006-6130-7

Benton, R., Sachse, S., Michnick, S. W., & Vosshall, L. B. (2006). Atypical membrane topology and heteromeric function of Drosophila odorant receptors in vivo. *PLoS Biol, 4*(2), e20. doi:10.1371/journal.pbio.0040020

Benton, R., Vannice, K. S., Gomez-Diaz, C., & Vosshall, L. B. (2009). Variant ionotropic glutamate receptors as chemosensory receptors in Drosophila. *Cell, 136*(1), 149-162. doi:10.1016/j.cell.2008.12.001

Bernasconi, M. L., Turlings, T. C., Ambrosetti, L., Bassetti, P., & Dorn, S. (1998). Herbivore-induced emissions of maize volatiles repel the corn leaf aphid, Rhopalosiphum maidis. *Entomologia Experimentalis et Applicata, 87*(2), 133-142.

Bian, L., Sun, X. L., Cai, X. M., & Chen, Z. M. (2014). Slow Release of Plant Volatiles Using Sol-Gel Dispensers. *J Econ Entomol, 107*(6), 2023-2029. doi:10.1603/EC14054

Biello, R., Singh, A., Godfrey, C. J., Fernández, F. F., Mugford, S. T., Powell, G., Mathers, T. C. (2021). A chromosome-level genome assembly of the woolly apple aphid, Eriosoma lanigerum Hausmann (Hemiptera: Aphididae). *Mol Ecol Resour, 21*(1), 316-326. doi:10.1111/1755-0998.13258

Birkett, M. A., Campbell, C. A., Chamberlain, K., Guerrieri, E., Hick, A. J., Martin, J. L., Woodcock, C. M. (2000). New roles for cis-jasmone as an insect semiochemical and in plant defense. *Proc Natl Acad Sci U S A, 97*(16), 9329-9334. doi:10.1073/pnas.160241697

Blackman, R. L., & Eastop, V. F. (2000). *Aphids on the world's crops.*: John Wiley

and Sons, Chichester.

Blackman, R. L., & Eastop, V. F. (2020). Aphids on the world's plants: an online identification and information guide. Retrieved 20 Dec 2021 http://www.aphidonworldsplants.info/

Boudarene, L., Baaliouamer, A., Meklati, B. Y., & Scharff, C. (2004). Composition of the Seed Oils from Algerian Cedrus atlantica G. Manetti. *Journal of Essential Oil Research, 16*(1), 61-63. doi:10.1080/10412905.2004.9698651

Boulain, H., Legeai, F., Jaquiéry, J., Guy, E., Morlière, S., Simon, J. C., & Sugio, A. (2019). Differential Expression of Candidate Salivary Effector Genes in Pea Aphid Biotypes With Distinct Host Plant Specificity. *Front Plant Sci, 10*, 1301. doi:10.3389/fpls.2019.01301

Bowers, W. S., Nault, L. R., Webb, R. E., & Dutky, S. R. (1972). Aphid alarm pheromone: isolation, identification, synthesis. *Science, 177*(4054), 1121-1122. doi:10.1126/science.177.4054.1121

Bowers, W. S., Nishino, C., Montgomery, M. E., Nault, L. R., & Nielson, M. W. (1977). Sesquiterpene progenitor, germacrene A: an alarm pheromone in aphids. *Science, 196*(4290), 680-681. doi:10.1126/science.558651

Bowling, R. D., Brewer, M. J., Kerns, D. L., Gordy, J., Seiter, N., Elliott, N. E., Maxson, E. (2016). Sugarcane Aphid (Hemiptera: Aphididae): A New Pest on Sorghum in North America. *J Integr Pest Manag, 7*(1), 12. doi:10.1093/jipm/pmw011

Brand, P., Robertson, H. M., Lin, W., Pothula, R., Klingeman, W. E., Jurat-Fuentes, J. L., & Johnson, B. R. (2018). The origin of the odorant receptor gene family in insects. *Elife, 7*, e38340. doi:10.7554/eLife.38340

Bruce, T. J. (2015). Interplay between insects and plants: dynamic and complex interactions that have coevolved over millions of years but act in milliseconds. *J Exp Bot, 66*(2), 455-465. doi:10.1093/jxb/eru391

Bruce, T. J., Martin, J. L., Pickett, J. A., Pye, B. J., Smart, L. E., & Wadhams, L. J. (2003). cis-Jasmone treatment induces resistance in wheat plants against the grain aphid, Sitobion avenae (Fabricius) (Homoptera: Aphididae). *Pest Manag Sci, 59*(9), 1031-1036. doi:10.1002/ps.730

Bruce, T. J., Wadhams, L. J., & Woodcock, C. M. (2005). Insect host location: a volatile situation. *Trends Plant Sci, 10*(6), 269-274. doi:10.1016/j.tplants.2005.04.003

Bruno, D., Grossi, G., Salvia, R., Scala, A., Farina, D., Grimaldi, A., Falabella, P. (2018). Sensilla Morphology and Complex Expression Pattern of Odorant Binding Proteins in the Vetch Aphid Megoura viciae (Hemiptera: Aphididae). *Front Physiol, 9*, 777. doi:10.3389/fphys.2018.00777

Budelli, G., Ni, L., Berciu, C., van Giesen, L., Knecht, Z. A., Chang, E. C., Garrity, P. A. (2019). Ionotropic Receptors Specify the Morphogenesis of Phasic Sensors Controlling Rapid Thermal Preference in Drosophila. *Neuron, 101*(4), 738-747.e733. doi:10.1016/j.neuron.2018.12.022

Burton, J. N., Adey, A., Patwardhan, R. P., Qiu, R., Kitzman, J. O., & Shendure, J. (2013). Chromosome-scale scaffolding of de novo genome assemblies

based on chromatin interactions. *Nat Biotechnol, 31*(12), 1119-1125. doi:10.1038/nbt.2727

Bushra, S., & Tariq, M. (2014). How aphid alarm pheromone can control aphids: a review. *Archives of Phytopathology and Plant Protection, 47*(13), 1563-1573.

Butterwick, J. A., Del Marmol, J., Kim, K. H., Kahlson, M. A., Rogow, J. A., Walz, T., & Ruta, V. (2018). Cryo-EM structure of the insect olfactory receptor Orco. *Nature, 560*(7719), 447-452. doi:10.1038/s41586-018-0420-8

Caillaud, M. C., & Via, S. (2000). Specialized Feeding Behavior Influences Both Ecological Specialization and Assortative Mating in Sympatric Host Races of Pea Aphids. *Am Nat, 156*(6), 606-621. doi:10.1086/316991

Cao, D., Liu, Y., Walker, W. B., Li, J., & Wang, G. (2014). Molecular characterization of the Aphis gossypii olfactory receptor gene families. *PLoS One, 9*(6), e101187. doi:10.1371/journal.pone.0101187

Cao, S., Huang, T., Shen, J., Liu, Y., & Wang, G. (2020). An Orphan Pheromone Receptor Affects the Mating Behavior of Helicoverpa armigera. *Frontiers in Physiology, 11*. doi:10.3389/fphys.2020.00413

Cao, S., Liu, Y., Guo, M., & Wang, G. (2016). A Conserved Odorant Receptor Tuned to Floral Volatiles in Three Heliothinae Species. *PLoS One, 11*(5), e0155029. doi:10.1371/journal.pone.0155029

Capella-Gutierrez, S., Silla-Martinez, J. M., & Gabaldon, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics, 25*(15), 1972-1973. doi:10.1093/bioinformatics/btp348

Carey, A. F., Wang, G., Su, C. Y., Zwiebel, L. J., & Carlson, J. R. (2010). Odorant reception in the malaria mosquito Anopheles gambiae. *Nature, 464*(7285), 66-71. doi:10.1038/nature08834

Cassau, S., & Krieger, J. (2021). The role of SNMPs in insect olfaction. *Cell Tissue Res, 383*(1), 21-33. doi:10.1007/s00441-020-03336-0

Catangui, M. A., Beckendorf, E. A., & Riedell, W. E. (2009). Soybean Aphid Population Dynamics, Soybean Yield Loss, and Development of Stage-Specific Economic Injury Levels. *Agronomy Journal, 101*(5), 1080-1092. doi:https://doi.org/10.2134/agronj2008.0233x

Cattaneo, A. M., Gonzalez, F., Bengtsson, J. M., Corey, E. A., Jacquin-Joly, E., Montagné, N., . . . Bobkov, Y. V. (2017). Candidate pheromone receptors of codling moth Cydia pomonella respond to pheromones and kairomones. *Sci Rep, 7*, 41105. doi:10.1038/srep41105

Chahda, J. S., Soni, N., Sun, J. S., Ebrahim, S. A. M., Weiss, B. L., & Carlson, J. R. (2019). The molecular and cellular basis of olfactory response to tsetse fly attractants. *PLoS Genet, 15*(3), e1008005. doi:10.1371/journal.pgen.1008005

Chaisson, M. J., & Tesler, G. (2012). Mapping single molecule sequencing reads using basic local alignment with successive refinement (BLASR): application and theory. *BMC Bioinformatics, 13*, 238. doi:10.1186/1471-2105-13-238

Chang, H., Liu, Y., Ai, D., Jiang, X., Dong, S., & Wang, G. (2017). A Pheromone

Antagonist Regulates Optimal Mating Time in the Moth Helicoverpa armigera. *Curr Biol, 27*(11), 1610-1615 e1613. doi:10.1016/j.cub.2017.04.035

Chang, H., Liu, Y., Yang, T., Pelosi, P., Dong, S., & Wang, G. (2015). Pheromone binding proteins enhance the sensitivity of olfactory receptors to sex pheromones in Chilo suppressalis. *Sci Rep, 5*, 13093. doi:10.1038/srep13093

Chen, A., Zhang, H., Shan, T., Shi, X., & Gao, X. (2020). The overexpression of three cytochrome P450 genes CYP6CY14, CYP6CY22 and CYP6UN1 contributed to metabolic resistance to dinotefuran in melon/cotton aphid, Aphis gossypii Glover. *Pestic Biochem Physiol, 167*, 104601. doi:10.1016/j.pestbp.2020.104601

Chen, C., Chen, H., Zhang, Y., Thomas, H. R., Frank, M. H., He, Y., & Xia, R. (2020). TBtools: An Integrative Toolkit Developed for Interactive Analyses of Big Biological Data. *Mol Plant, 13*(8), 1194-1202. doi:10.1016/j.molp.2020.06.009

Chen, K., Durand, D., & Farach-Colton, M. (2000). NOTUNG: a program for dating gene duplications and optimizing gene family trees. *J Comput Biol, 7*(3-4), 429-447. doi:10.1089/106652700750050871

Chen, L., Tian, K., Xu, X., Fang, A., Cheng, W., Wang, G., Wu, J. (2020). Detecting Host-Plant Volatiles with Odorant Receptors from Grapholita molesta (Busck) (Lepidoptera: Tortricidae). *J Agric Food Chem, 68*(9), 2711-2717. doi:10.1021/acs.jafc.9b07305

Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics, 34*(17), i884-i890. doi:10.1093/bioinformatics/bty560

Chen, W., Shakir, S., Bigham, M., Richter, A., Fei, Z., & Jander, G. (2019). Genome sequence of the corn leaf aphid (Rhopalosiphum maidis Fitch). *GigaScience, 8*(4). doi:10.1093/gigascience/giz033

Chen, Y., & Amrein, H. (2017). Ionotropic Receptors Mediate Drosophila Oviposition Preference through Sour Gustatory Receptor Neurons. *Current Biology, 27*(18), 2741-2750.e2744. doi:10.1016/j.cub.2017.08.003

Chen, Y., He, M., Li, Z.-Q., Zhang, Y.-N., & He, P. (2016). Identification and tissue expression profile of genes from three chemoreceptor families in an urban pest, Periplaneta americana. *Scientific Reports, 6*, 27495. doi:10.1038/srep27495

Chin, C. S., Peluso, P., Sedlazeck, F. J., Nattestad, M., Concepcion, G. T., Clum, A., . . . Schatz, M. C. (2016). Phased diploid genome assembly with single-molecule real-time sequencing. *Nat Methods, 13*(12), 1050-1054. doi:10.1038/nmeth.4035

Chyb, S. (2004). Drosophila gustatory receptors: from gene identification to functional expression. *J Insect Physiol, 50*(6), 469-477. doi:10.1016/j.jinsphys.2004.03.012

Clyne, P. J., Warr, C. G., & Carlson, J. R. (2000). Candidate taste receptors in Drosophila. *Science, 287*(5459), 1830-1834.

doi:10.1126/science.287.5459.1830

Clyne, P. J., Warr, C. G., Freeman, M. R., Lessing, D., Kim, J., & Carlson, J. R. (1999). A novel family of divergent seven-transmembrane proteins: candidate odorant receptors in Drosophila. *Neuron, 22*(2), 327-338. doi:10.1016/s0896-6273(00)81093-4

Croset, V., Rytz, R., Cummins, S. F., Budd, A., Brawand, D., Kaessmann, H., . . . Benton, R. (2010). Ancient protostome origin of chemosensory ionotropic glutamate receptors and the evolution of insect taste and olfaction. *PLoS Genet, 6*(8), e1001064. doi:10.1371/journal.pgen.1001064

Crossley, M. S., & Hogg, D. B. (2015). Rag Virulence Among Soybean Aphids (Hemiptera: Aphididae) in Wisconsin. *J Econ Entomol, 108*(1), 326-338. doi:10.1093/jee/tou022

Cui, W., Wang, B., Guo, M., Liu, Y., Jacquin-Joly, E., Yan, S., & Wang, G. (2018). A receptor-neuron correlate for the detection of attractive plant volatiles in Helicoverpa assulta (Lepidoptera: Noctuidae). *Insect Biochem Mol Biol, 97*, 31-39. doi:10.1016/j.ibmb.2018.04.006

Dahanukar, A., Lei, Y.-T., Kwon, J. Y., & Carlson, J. R. (2007). Two Gr Genes Underlie Sugar Reception in Drosophila. *Neuron, 56*(3), 503-516. doi:10.1016/j.neuron.2007.10.024

Dardouri, T., Gautier, H., Ben Issa, R., Costagliola, G., & Gomez, L. (2019). Repellence of Myzus persicae (Sulzer): evidence of two modes of action of volatiles from selected living aromatic plants. *Pest Manag Sci, 75*(6), 1571-1584. doi:10.1002/ps.5271

De Biasio, F., Riviello, L., Bruno, D., Grimaldi, A., Congiu, T., Sun, Y. F., & Falabella, P. (2015). Expression pattern analysis of odorant-binding proteins in the pea aphid Acyrthosiphon pisum. *Insect science, 22*(2), 220-234. doi:10.1111/1744-7917.12118

De Bie, T., Cristianini, N., Demuth, J. P., & Hahn, M. W. (2006). CAFE: a computational tool for the study of gene family evolution. *Bioinformatics, 22*(10), 1269-1271. doi:10.1093/bioinformatics/btl097

de Fouchier, A., Walker, W. B., 3rd, Montagné, N., Steiner, C., Binyameen, M., Schlyter, F., Jacquin-Joly, E. (2017). Functional evolution of Lepidoptera olfactory receptors revealed by deorphanization of a moth repertoire. *Nat Commun, 8*, 15709. doi:10.1038/ncomms15709

de Oliveira, R. S., Penaflor, M., Goncalves, F. G., Sampaio, M. V., Korndorfer, A. P., Silva, W. D., & Bento, J. M. S. (2020). Silicon-induced changes in plant volatiles reduce attractiveness of wheat to the bird cherry-oat aphid Rhopalosiphum padi and attract the parasitoid Lysiphlebus testaceipes. *PLoS One, 15*(4), e0231005. doi:10.1371/journal.pone.0231005

de Vos, M., Cheng, W. Y., Summers, H. E., Raguso, R. A., & Jander, G. (2010). Alarm pheromone habituation in Myzus persicae has fitness consequences and causes extensive gene expression changes. *Proc Natl Acad Sci U S A, 107*(33), 14673-14678. doi:10.1073/pnas.1001539107

Dedryver, C. A., Le Ralec, A., & Fabre, F. (2010). The conflicting relationships between aphids and men: a review of aphid damage and control strategies.

C R Biol, 333(6-7), 539-553. doi:10.1016/j.crvi.2010.03.009

Dewhirst, S. Y., Birkett, M. A., Fitzgerald, J. D., Stewart-Jones, A., Wadhams, L. J., Woodcock, C. M., Pickett, J. A. (2008). Dolichodial: a new aphid sex pheromone component? *J Chem Ecol, 34*(12), 1575-1583. doi:10.1007/s10886-008-9561-9

Dewhirst, S. Y., Birkett, M. A., Loza-Reyes, E., Martin, J. L., Pye, B. J., Smart, L. E., Pickett, J. A. (2012). Activation of defence in sweet pepper, Capsicum annum, by cis-jasmone, and its impact on aphid and aphid parasitoid behaviour. *Pest Manag Sci, 68*(10), 1419-1429. doi:10.1002/ps.3326

Dickson, R., Laird, E., & Pesho, G. (1955). The spotted alfalfa aphid (yellow clover aphid on alfalfa). *Hilgardia, 24*(5), 93-118. doi:10.3733/hilg.v24n05p093

Dixon, A. (1998). *Aphid Ecology: An Optimization Approach*: London: Springer.

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics, 29*(1), 15-21. doi:10.1093/bioinformatics/bts635

Dobritsa, A. A., van der Goes van Naters, W., Warr, C. G., Steinbrecht, R. A., & Carlson, J. R. (2003). Integrating the molecular and cellular basis of odor coding in the Drosophila antenna. *Neuron, 37*(5), 827-841. doi:10.1016/s0896-6273(03)00094-1

Dus, M., Min, S., Keene, A. C., Lee, G. Y., & Suh, G. S. (2011). Taste-independent detection of the caloric content of sugar in Drosophila. *Proc Natl Acad Sci U S A, 108*(28), 11644-11649. doi:10.1073/pnas.1017096108

Duvaux, L., Geissmann, Q., Gharbi, K., Zhou, J. J., Ferrari, J., Smadja, C. M., & Butlin, R. K. (2015). Dynamics of copy number variation in host races of the pea aphid. *Mol Biol Evol, 32*(1), 63-80. doi:10.1093/molbev/msu266

Dweck, H. K., Ebrahim, S. A., Farhan, A., Hansson, B. S., & Stensmyr, M. C. (2015). Olfactory proxy detection of dietary antioxidants in Drosophila. *Curr Biol, 25*(4), 455-466. doi:10.1016/j.cub.2014.11.062

Edwards, L. J., Siddall, J. B., Dunham, L. L., Uden, P., & Kislow, C. J. (1973). Trans-β-farnesene, Alarm Pheromone of the Green Peach Aphid, Myzus persicae (Sulzer). *Nature, 241*(5385), 126-127. doi:10.1038/241126b0

El-Gebali, S., Mistry, J., Bateman, A., Eddy, S. R., Luciani, A., Potter, S. C., Finn, R. D. (2019). The Pfam protein families database in 2019. *Nucleic Acids Res, 47*(D1), D427-D432. doi:10.1093/nar/gky995

Elgar, M. A., Zhang, D., Wang, Q., Wittwer, B., Thi Pham, H., Johnson, T. L., Coquilleau, M. (2018). Insect Antennal Morphology: The Evolution of Diverse Solutions to Odorant Perception. *Yale J Biol Med, 91*(4), 457-469.

Emden, H. F. v., & Harrington, R. (2017). *Aphids as crop pests*: CAB International.

Emms, D. M., & Kelly, S. (2017). STRIDE: Species Tree Root Inference from Gene Duplication Events. *Mol Biol Evol, 34*(12), 3267-3278. doi:10.1093/molbev/msx259

Emms, D. M., & Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol, 20*(1), 238. doi:10.1186/s13059-019-1832-y

Eyres, I., Jaquiery, J., Sugio, A., Duvaux, L., Gharbi, K., Zhou, J. J., Ferrari, J.

(2016). Differential gene expression according to race and host plant in the pea aphid. *Molecular ecology, 25*(17), 4197-4215. doi:10.1111/mec.13771

Eyun, S. I., Soh, H. Y., Posavi, M., Munro, J. B., Hughes, D. S. T., Murali, S. C., Lee, C. E. (2017). Evolutionary History of Chemosensory-Related Gene Families across the Arthropoda. *Mol Biol Evol, 34*(8), 1838-1862. doi:10.1093/molbev/msx147

Fan, J., Xue, W., Duan, H., Jiang, X., Zhang, Y., Yu, W., Chen, J. (2017). Identification of an intraspecific alarm pheromone and two conserved odorant-binding proteins associated with (E)-beta-farnesene perception in aphid Rhopalosiphum padi. *J Insect Physiol, 101*, 151-160. doi:10.1016/j.jinsphys.2017.07.014

Fan, J., Zhang, Y., Francis, F., Cheng, D., Sun, J., & Chen, J. (2015). Orco mediates olfactory behaviors and winged morph differentiation induced by alarm pheromone in the grain aphid, Sitobion avenae. *Insect Biochemistry and Molecular Biology, 64*, 16-24. doi:10.1016/j.ibmb.2015.07.006

Favret, C. (2013). *Aphid species file* Retrieved from: http://Aphid.SpeciesFile.org

Ferrari, J., Via, S., & Godfray, H. C. (2008). Population differentiation and genetic variation in performance on eight hosts in the pea aphid complex. *Evolution, 62*(10), 2508-2524. doi:10.1111/j.1558-5646.2008.00468.x

Fleischer, J., & Krieger, J. (2018). Insect Pheromone Receptors - Key Elements in Sensing Intraspecific Chemical Signals. *Front Cell Neurosci, 12*, 425. doi:10.3389/fncel.2018.00425

Fleischer, J., Pregitzer, P., Breer, H., & Krieger, J. (2018). Access to the odor world: olfactory receptors and their role for signal transduction in insects. *Cell Mol Life Sci, 75*(3), 485-508. doi:10.1007/s00018-017-2627-5

Forstner, M., Breer, H., & Krieger, J. (2009). A receptor and binding protein interplay in the detection of a distinct pheromone component in the silkmoth Antheraea polyphemus. *Int J Biol Sci, 5*(7), 745-757. doi:10.7150/ijbs.5.745

Fox, A. N., Pitts, R. J., Robertson, H. M., Carlson, J. R., & Zwiebel, L. J. (2001). Candidate odorant receptors from the malaria vector mosquito Anopheles gambiae and evidence of down-regulation in response to blood feeding. *Proc Natl Acad Sci U S A, 98*(25), 14693-14697. doi:10.1073/pnas.261432998

Fox, E. G. P., & Adams, R. M. M. (2022). On the Biological Diversity of Ant Alkaloids. *Annu Rev Entomol, 67*, 367-385. doi:10.1146/annurev-ento-072821-063525

Francis, F., Vandermoten, S., Verheggen, F., Lognay, G., & Haubruge, E. (2005). Is the (E)-β-farnesene only volatile terpenoid in aphids? *Journal of Applied Entomology, 129*(1), 6-11. doi:https://doi.org/10.1111/j.1439-0418.2005.00925.x

Frank, D., Watkins, P., Ball, A., Krishnamurthy, R., Piyasiri, U., Sewell, J., Warner, R. (2016). Impact of Brassica and Lucerne Finishing Feeds and Intramuscular Fat on Lamb Eating Quality and Flavor. A Cross-Cultural Study Using Chinese and Non-Chinese Australian Consumers. *J Agric*

Food Chem, 64(36), 6856-6868. doi:10.1021/acs.jafc.6b02018

Gadenne, C., Barrozo, R. B., & Anton, S. (2016). Plasticity in Insect Olfaction: To Smell or Not to Smell? *Annu Rev Entomol, 61*, 317-333. doi:10.1146/annurev-ento-010715-023523

Galperin, M. Y., Makarova, K. S., Wolf, Y. I., & Koonin, E. V. (2015). Expanded microbial genome coverage and improved protein family annotation in the COG database. *Nucleic Acids Res, 43*(Database issue), D261-269. doi:10.1093/nar/gku1223

Gao, H., Lai, S., Zhai, Y., Lv, Z., Zheng, L., Yu, Y., & Ren, F.-S. (2020). Comparison of the antennal sensilla and compound eye sensilla in four Drosophila (Diptera: Drosophilidae) species. *Florida Entomologist, 102*(4), 747-754.

German, P. F., van der Poel, S., Carraher, C., Kralicek, A. V., & Newcomb, R. D. (2013). Insights into subunit interactions within the insect olfactory receptor complex using FRET. *Insect Biochem Mol Biol, 43*(2), 138-145. doi:10.1016/j.ibmb.2012.11.002

Goldman-Huertas, B., Mitchell, R. F., Lapoint, R. T., Faucher, C. P., Hildebrand, J. G., & Whiteman, N. K. (2015). Evolution of herbivory in Drosophilidae linked to loss of behaviors, antennal responses, odorant receptors, and ancestral diet. *Proc Natl Acad Sci U S A, 112*(10), 3026-3031. doi:10.1073/pnas.1424656112

Goldman, A. L., Van der Goes van Naters, W., Lessing, D., Warr, C. G., & Carlson, J. R. (2005). Coexpression of two functional odor receptors in one neuron. *Neuron, 45*(5), 661-666. doi:10.1016/j.neuron.2005.01.025

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol, 29*(7), 644-652. doi:10.1038/nbt.1883

Gray, S. M. (2008). Aphid transmission of plant viruses. *Curr Protoc Microbiol, Chapter 16*, Unit 16B.11.11-16B.11.10. doi:10.1002/9780471729259.mc16b01s10

Gu, S., Wu, K., Guo, Y., Field, L., Pickett, J., Zhang, Y., & Zhou, J. (2013). Identification and expression profiling of odorant binding proteins and chemosensory proteins between two wingless morphs and a winged morph of the cotton aphid Aphis gossypii glover. *PLoS One, 8*(9), e73524. doi:10.1371/journal.pone.0073524

Guo, H., & Wang, C. (2019). The ethological significance and olfactory detection of herbivore-induced plant volatiles in interactions of plants, herbivorous insects, and parasitoids.

Guo, M., Du, L., Chen, Q., Feng, Y., Zhang, J., Zhang, X., Liu, Y. (2020). Odorant Receptors for Detecting Flowering Plant Cues Are Functionally Conserved across Moths and Butterflies. *Molecular biology and evolution*. doi:10.1093/molbev/msaa300

Guo, X., Yu, Q., Chen, D., Wei, J., Yang, P., Yu, J., Kang, L. (2020). 4-Vinylanisole is an aggregation pheromone in locusts. *Nature, 584*(7822), 584-588. doi:10.1038/s41586-020-2610-4

Haas, B. J., Salzberg, S. L., Zhu, W., Pertea, M., Allen, J. E., Orvis, J., Wortman, J. R. (2008). Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biol, 9*(1), R7. doi:10.1186/gb-2008-9-1-r7

Hales, D. F., Gillespie, P. S., Wade, S., & Dominiak, B. C. (2017). First detection of'Megoura crassicauda'Maudvilko (Hemiptera: Aphididae) in Australia and a review of its biology. *General and Applied Entomology: The Journal of the Entomological Society of New South Wales, 45*, 77-81.

Hallem, E. A., & Carlson, J. R. (2006). Coding of odors by a receptor repertoire. *Cell, 125*(1), 143-160. doi:10.1016/j.cell.2006.01.050

Hallem, E. A., Ho, M. G., & Carlson, J. R. (2004). The molecular basis of odor coding in the Drosophila antenna. *Cell, 117*(7), 965-979. doi:10.1016/j.cell.2004.05.012

Hallgren, J., Tsirigos, K. D., Pedersen, M. D., Almagro Armenteros, J. J., Marcatili, P., Nielsen, H., . . . Winther, O. (2022). DeepTMHMM predicts alpha and beta transmembrane proteins using deep neural networks. *bioRxiv*, 2022.2004.2008.487609. doi:10.1101/2022.04.08.487609

Han, B., Zhang, Q.-H., & Byers, J. A. (2012). Attraction of the tea aphid, Toxoptera aurantii, to combinations of volatiles and colors related to tea plants. *Entomologia Experimentalis et Applicata, 144*(3), 258-269. doi:10.1111/j.1570-7458.2012.01303.x

Han, Y., & Wessler, S. (2010). MITE-Hunter: a program for discovering miniature inverted-repeat transposable elements from genomic sequences. *Nucleic Acids Res, 38*(22), e199. doi:10.1093/nar/gkq862

Hansen, A. K., & Moran, N. A. (2011). Aphid genome expression reveals host-symbiont cooperation in the production of amino acids. *Proc Natl Acad Sci U S A, 108*(7), 2849-2854. doi:10.1073/pnas.1013465108

Hansson, B. S., & Stensmyr, M. C. (2011). Evolution of insect olfaction. *Neuron, 72*(5), 698-711. doi:10.1016/j.neuron.2011.11.003

Harraca, V., Ryne, C., & Ignell, R. (2010). Nymphs of the common bed bug (Cimex lectularius) produce anti-aphrodisiac defence against conspecific males. *BMC Biol, 8*, 121. doi:10.1186/1741-7007-8-121

Hassan, B. A., Hussain, A., Zhang, M., Üçpunar, H. K., Svensson, T., Quillery, E., Grunwald Kadow, I. C. (2016). Ionotropic Chemosensory Receptors Mediate the Taste and Smell of Polyamines. *PLOS Biology, 14*(5), e1002454. doi:10.1371/journal.pbio.1002454

Haverkamp, A., Hansson, B. S., & Knaden, M. (2018). Combinatorial Codes and Labeled Lines: How Insects Use Olfactory Cues to Find and Judge Food, Mates, and Oviposition Sites in Complex Environments. *Front Physiol, 9*, 49. doi:10.3389/fphys.2018.00049

He, C., & Zhang, X. (2006). Field evaluation of lucerne (Medicago sativa L.) for resistance to aphids in northern China. *Australian Journal of Agricultural Research, 57*(4), 471-475. doi:https://doi.org/10.1071/AR05255

He, Z., Zhang, H., Gao, S., Lercher, M. J., Chen, W. H., & Hu, S. (2016). Evolview v2: an online visualization and management tool for customized and

annotated phylogenetic trees. *Nucleic Acids Res, 44*(W1), W236-241. doi:10.1093/nar/gkw370

Hegde, M., Oliveira, J. N., da Costa, J. G., Loza-Reyes, E., Bleicher, E., Santana, A. E., Bruce, T. J. (2012). Aphid antixenosis in cotton is activated by the natural plant defence elicitor cis-jasmone. *Phytochemistry, 78*, 81-88.

Hendry, A. P. (2001). A large family of Drosophila odorant-binding proteins expressed in gustatory and olfactory sensilla. *Genetica, 112/113*, 515-534. doi:10.1023/a:1013367100865

Herbert, J. J., Mizell, R. F., 3rd, & McAuslane, H. J. (2009). Host preference of the crapemyrtle aphid (Hemiptera: Aphididae) and host suitability of crapemyrtle cultivars. *Environ Entomol, 38*(4), 1155-1160. doi:10.1603/022.038.0423

Hill, C. A., Fox, A. N., Pitts, R. J., Kent, L. B., Tan, P. L., Chrystal, M. A., Zwiebel, L. J. (2002). G protein-coupled receptors in Anopheles gambiae. *Science, 298*(5591), 176-178. doi:10.1126/science.1076196

Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q., & Vinh, L. S. (2018). UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Mol Biol Evol, 35*(2), 518-522. doi:10.1093/molbev/msx281

Hodge, S., & Powell, G. (2010). Conditional facilitation of an aphid vector, Acyrthosiphon pisum, by the plant pathogen, pea enation mosaic virus. *J Insect Sci, 10*, 155. doi:10.1673/031.010.14115

Hopkins, D. P., Cameron, D. D., & Butlin, R. K. (2017). The chemical signatures underlying host plant discrimination by aphids. *Sci Rep, 7*(1), 8498. doi:10.1038/s41598-017-07729-0

Hore, G., Maity, A., Naskar, A., Ansar, W., Ghosh, S., Saha, G. K., & Banerjee, D. (2017). Scanning electron microscopic studies on antenna of Hemipyrellia ligurriens (Wiedemann, 1830) (Diptera: Calliphoridae)-A blow fly species of forensic importance. *Acta Trop, 172*, 20-28. doi:10.1016/j.actatropica.2017.04.005

Hore, G., Saha, G. K., & Banerjee, D. (2018). Sensory organs of forensically important fly Ophyra capensis (Wiedemann, 1818) (Diptera: Muscidae): A scanning electron microscopic study. *Acta Trop, 185*, 400-411. doi:10.1016/j.actatropica.2018.06.023

Hou, X., Zhang, D., Yuvaraj, J., Corcoran, J., Andersson, M., & Löfstedt, C. (2020). Functional characterization of odorant receptors from the moth Eriocrania semipurpurella: A comparison of results in the Xenopus oocyte and HEK cell systems. *Insect Biochem Mol Biol, 117*, 103289. doi:10.1016/j.ibmb.2019.103289

Hu, J., Fan, J., Sun, Z., & Liu, S. (2020). NextPolish: a fast and efficient genome polishing tool for long-read assembly. *Bioinformatics, 36*(7), 2253-2255. doi:10.1093/bioinformatics/btz891

Huang, T., Zhang, R., Yang, L., Cao, S., Francis, F., Wang, B., & Wang, G. (2022). Identification and functional characterization of ApisOr23 in pea aphid Acyrthosiphon pisum. *Journal of Integrative Agriculture, 21*(5), 1414-1423. doi:https://doi.org/10.1016/S2095-3119(20)63577-8

Hulle, M., Chaubet, B., Turpeau, E., & Simon, J.-C. (2020). Encyclop'Aphid: a website on aphids and their natural enemies.

International Aphid Genomics, C. (2010). Genome sequence of the pea aphid Acyrthosiphon pisum. *PLoS Biol, 8*(2), e1000313. doi:10.1371/journal.pbio.1000313

Irwin, J. A. G., Lloyd, D. L., & Lowe, K. F. (2001). Lucerne biology and genetic improvement - an analysis of past activities and future goals in Australia. *Australian Journal of Agricultural Research, 52*(7), 699-712. doi:https://doi.org/10.1071/AR00181

Jacques, S., Sperschneider, J., Garg, G., Thatcher, L. F., Gao, L.-L., Kamphuis, L. G., & Singh, K. B. (2020). A functional genomics approach to dissect spotted alfalfa aphid resistance in Medicago truncatula. *Scientific Reports, 10*(1), 22159. doi:10.1038/s41598-020-78904-z

Jacquin-Joly, E., & Merlin, C. (2004). Insect olfactory receptors: contributions of molecular biology to chemical ecology. *J Chem Ecol, 30*(12), 2359-2397. doi:10.1007/s10886-004-7941-3

Jaquiery, J., Peccoud, J., Ouisse, T., Legeai, F., Prunier-Leterme, N., Gouin, A., Rispe, C. (2018). Disentangling the Causes for Faster-X Evolution in Aphids. *Genome Biol Evol, 10*(2), 507-520. doi:10.1093/gbe/evy015

Jaquiery, J., Stoeckel, S., Nouhaud, P., Mieuzet, L., Maheo, F., Legeai, F., Simon, J. C. (2012). Genome scans reveal candidate regions involved in the adaptation to host plant in the pea aphid complex. *Molecular ecology, 21*(21), 5251-5264. doi:10.1111/mec.12048

Jia, H. R., Sun, Y. F., Luo, S. P., & Wu, K. M. (2019). Characterization of antennal chemosensilla and associated odorant binding as well as chemosensory proteins in the Eupeodes corollae (Diptera: Syrphidae). *J Insect Physiol, 113*, 49-58. doi:10.1016/j.jinsphys.2018.08.002

Jiang, X., Zhang, Q., Qin, Y., Yin, H., Zhang, S., Li, Q., Chen, J. (2019). A chromosome-level draft genome of the grain aphid Sitobion miscanthi. *GigaScience, 8*(8). doi:10.1093/gigascience/giz101

Jiao, Y., Moon, S. J., & Montell, C. (2007). A Drosophila gustatory receptor required for the responses to sucrose, glucose, and maltose identified by mRNA tagging. *Proceedings of the National Academy of Sciences, 104*(35), 14110-14115. doi:10.1073/pnas.0702421104

Jones, R. A. C. (2004). Occurrence of virus infection in seed stocks and 3-year-old pastures of lucerne (Medicago sativa). *Australian Journal of Agricultural Research, 55*(7), 757-764. doi:https://doi.org/10.1071/AR04011

Jones, W. D., Cayirlioglu, P., Grunwald Kadow, I., & Vosshall, L. B. (2006). Two chemosensory receptors together mediate carbon dioxide detection in Drosophila. *Nature, 445*(7123), 86-90. doi:10.1038/nature05466

Jones, W. D., Nguyen, T. A., Kloss, B., Lee, K. J., & Vosshall, L. B. (2005). Functional conservation of an insect odorant receptor gene across 250 million years of evolution. *Curr Biol, 15*(4), R119-121. doi:10.1016/j.cub.2005.02.007

Joseph, R. M., & Carlson, J. R. (2015). Drosophila Chemoreceptors: A Molecular

Interface Between the Chemical World and the Brain. *Trends in genetics : TIG, 31*(12), 683-695. doi:10.1016/j.tig.2015.09.005

Julca, I., Marcet-Houben, M., Cruz, F., Vargas-Chavez, C., Johnston, J. S., Gómez-Garrido, J., Gabaldón, T. (2020). Phylogenomics Identifies an Ancestral Burst of Gene Duplications Predating the Diversification of Aphidomorpha. *Mol Biol Evol, 37*(3), 730-756. doi:10.1093/molbev/msz261

Kamphuis, L. G., Lichtenzveig, J., Peng, K., Guo, S.-M., Klingler, J. P., Siddique, K. H. M., Singh, K. B. (2013). Characterization and genetic dissection of resistance to spotted alfalfa aphid (Therioaphis trifolii) in Medicago truncatula. *Journal of Experimental Botany, 64*(16), 5157-5172. doi:10.1093/jxb/ert305

Kanehisa, M., & Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res, 28*(1), 27-30. doi:10.1093/nar/28.1.27

Kappers, I. F., Aharoni, A., van Herpen, T. W., Luckerhoff, L. L., Dicke, M., & Bouwmeester, H. J. (2005). Genetic engineering of terpenoid metabolism attracts bodyguards to Arabidopsis. *Science, 309*(5743), 2070-2072. doi:10.1126/science.1116232

Karasev, A. V., & Gray, S. M. (2013). Continuous and emerging challenges of Potato virus Y in potato. *Annu Rev Phytopathol, 51*, 571-586. doi:10.1146/annurev-phyto-082712-102332

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol, 30*(4), 772-780. doi:10.1093/molbev/mst010

Keesey, I. W., Doll, G., Chakraborty, S. D., Baschwitz, A., Lemoine, M., Kaltenpoth, M., Hansson, B. S. (2020). Alcohol boosts pheromone production in male flies and makes them sexier. *bioRxiv*, 2020.2008.2009.242784. doi:10.1101/2020.08.09.242784

Keesey, I. W., Zhang, J., Depetris-Chauvin, A., Obiero, G. F., Gupta, A., Gupta, N., . . . Hansson, B. S. (2022). Functional olfactory evolution in Drosophila suzukii and the subgenus Sophophora. *iScience, 25*(5), 104212. doi:https://doi.org/10.1016/j.isci.2022.104212

Keilwagen, J., Wenk, M., Erickson, J. L., Schattat, M. H., Grau, J., & Hartung, F. (2016). Using intron position conservation for homology-based gene prediction. *Nucleic Acids Res, 44*(9), e89. doi:10.1093/nar/gkw092

Khashaveh, A., An, X., Shan, S., Xiao, Y., Wang, Q., Wang, S., Zhang, Y. (2020). Deorphanization of an odorant receptor revealed new bioactive components for green mirid bug Apolygus lucorum (Hemiptera: Miridae). *Pest Manag Sci, 76*(5), 1626-1638. doi:10.1002/ps.5682

Kim, D., Paggi, J. M., Park, C., Bennett, C., & Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol, 37*(8), 907-915. doi:10.1038/s41587-019-0201-4

Kim, H., & Lee, S. (2008). Molecular systematics of the genus Megoura (Hemiptera: Aphididae) using mitochondrial and nuclear DNA sequences. *Mol Cells, 25*(4), 510-522.

Kivimäenpää, M., Babalola, A. B., Joutsensaari, J., & Holopainen, J. K. (2020).

Methyl Salicylate and Sesquiterpene Emissions Are Indicative for Aphid Infestation on Scots Pine. *Forests, 11*(5), 573.

Knecht, Z. A., Silbering, A. F., Cruz, J., Yang, L., Croset, V., Benton, R., & Garrity, P. A. (2017). Ionotropic Receptor-dependent moist and dry cells control hygrosensation in. *Elife, 6*. doi:10.7554/eLife.26654

Knecht, Z. A., Silbering, A. F., Ni, L., Klein, M., Budelli, G., Bell, R., Garrity, P. A. (2016). Distinct combinations of variant ionotropic glutamate receptors mediate thermosensation and hygrosensation in. *Elife, 5*. doi:10.7554/eLife.17879

Koh, T. W., He, Z., Gorur-Shandilya, S., Menuz, K., Larter, N. K., Stewart, S., & Carlson, J. R. (2014). The Drosophila IR20a clade of ionotropic receptors are candidate taste and pheromone receptors. *Neuron, 83*(4), 850-865. doi:10.1016/j.neuron.2014.07.012

Konopka, J. K., Task, D., Afify, A., Raji, J., Deibel, K., Maguire, S., Potter, C. J. (2021). Olfaction in Anopheles mosquitoes. *Chem Senses, 46*. doi:10.1093/chemse/bjab021

Korgaonkar, A., Han, C., Lemire, A. L., Siwanowicz, I., Bennouna, D., Kopec, R. E., Stern, D. L. (2021). A novel family of secreted insect proteins linked to plant gall development. *Curr Biol, 31*(9), 1836-1849.e1812. doi:10.1016/j.cub.2021.01.104

Kosakovsky Pond, S. L., Poon, A. F. Y., Velazquez, R., Weaver, S., Hepler, N. L., Murrell, B., Muse, S. V. (2020). HyPhy 2.5-A Customizable Platform for Evolutionary Hypothesis Testing Using Phylogenies. *Mol Biol Evol, 37*(1), 295-299. doi:10.1093/molbev/msz197

Kovaka, S., Zimin, A. V., Pertea, G. M., Razaghi, R., Salzberg, S. L., & Pertea, M. (2019). Transcriptome assembly from long-read RNA-seq alignments with StringTie2. *Genome Biol, 20*(1), 278. doi:10.1186/s13059-019-1910-1

Kreher, S. A., Kwon, J. Y., & Carlson, J. R. (2005). The molecular basis of odor coding in the Drosophila larva. *Neuron, 46*(3), 445-456. doi:10.1016/j.neuron.2005.04.007

Kreher, S. A., Mathew, D., Kim, J., & Carlson, J. R. (2008). Translation of sensory input into behavioral output via an olfactory system. *Neuron, 59*(1), 110-124. doi:10.1016/j.neuron.2008.06.010

Krieger, J., Klink, O., Mohl, C., Raming, K., & Breer, H. (2003). A candidate olfactory receptor subtype highly conserved across different insect orders. *J Comp Physiol A Neuroethol Sens Neural Behav Physiol, 189*(7), 519-526. doi:10.1007/s00359-003-0427-x

Kroes, A., Weldegergis, B. T., Cappai, F., Dicke, M., & van Loon, J. J. A. (2017). Terpenoid biosynthesis in Arabidopsis attacked by caterpillars and aphids: effects of aphid density on the attraction of a caterpillar parasitoid. *Oecologia, 185*(4), 699-712. doi:10.1007/s00442-017-3985-2

Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol., 33*(7), 1870-1874. doi:10.1093/molbev/msw054

Kurtovic, A., Widmer, A., & Dickson, B. J. (2007). A single class of olfactory

neurons mediates behavioural responses to a Drosophila sex pheromone. *Nature, 446*(7135), 542-546. doi:10.1038/nature05672

Kymre, J. H., Liu, X., Ian, E., Berge, C. N., Wang, G., Berg, B. G., Chu, X. (2021). Distinct protocerebral neuropils associated with attractive and aversive female-produced odorants in the male moth brain. *Elife, 10*. doi:10.7554/eLife.65683

Lake, A. (1989). Spotted alfalfa aphid survival and reproduction on annual medics with various levels of aphid resistance. *Australian Journal of Agricultural Research, 40*(1), 117-123. doi:https://doi.org/10.1071/AR9890117

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat Methods, 9*(4), 357-359. doi:10.1038/nmeth.1923

Larsson, M. C., Domingos, A. I., Jones, W. D., Chiappe, M. E., Amrein, H., & Vosshall, L. B. (2004). Or83b encodes a broadly expressed odorant receptor essential for Drosophila olfaction. *Neuron, 43*(5), 703-714. doi:10.1016/j.neuron.2004.08.019

Leal, W. S. (2013). Odorant reception in insects: roles of receptors, binding proteins, and degrading enzymes. *Annu Rev Entomol, 58*, 373-391. doi:10.1146/annurev-ento-120811-153635

Legan, A. W., Jernigan, C. M., Miller, S. E., Fuchs, M. F., & Sheehan, M. J. (2021). Expansion and Accelerated Evolution of 9-Exon Odorant Receptors in Polistes Paper Wasps. *Mol Biol Evol, 38*(9), 3832-3846. doi:10.1093/molbev/msab023

Legeai, F., Shigenobu, S., Gauthier, J. P., Colbourne, J., Rispe, C., Collin, O., Tagu, D. (2010). AphidBase: a centralized bioinformatic resource for annotation of the pea aphid genome. *Insect Mol Biol, 19 Suppl 2*, 5-12. doi:10.1111/j.1365-2583.2009.00930.x

Li, B., Bickel, R. D., Parker, B. J., Saleh Ziabari, O., Liu, F., Vellichirammal, N. N., Brisson, J. A. (2020). A large genomic insertion containing a duplicated follistatin gene is linked to the pea aphid male wing dimorphism. *Elife, 9*. doi:10.7554/eLife.50608

Li, B., & Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics, 12*, 323. doi:10.1186/1471-2105-12-323

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics, 25*(14), 1754-1760. doi:10.1093/bioinformatics/btp324

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics, 25*(16), 2078-2079. doi:10.1093/bioinformatics/btp352

Li, Y., Park, H., Smith, T. E., & Moran, N. A. (2019). Gene Family Evolution in the Pea Aphid Based on Chromosome-Level Genome Assembly. *Mol Biol Evol, 36*(10), 2143-2156. doi:10.1093/molbev/msz138

Li, Z., Zhang, S., Cai, X., Luo, J., Dong, S., Cui, J., & Chen, Z. (2018). Distinct binding affinities of odorant-binding proteins from the natural predator Chrysoperla sinica suggest different strategies to hunt prey. *J Insect Physiol,*

111, 25-31. doi:10.1016/j.jinsphys.2018.10.004

Lin, C., Prokop-Prigge, K. A., Preti, G., & Potter, C. J. (2015). Food odors trigger Drosophila males to deposit a pheromone that guides aggregation and female oviposition decisions. *Elife, 4*, e08688. doi:10.7554/eLife.08688

Liu, H., Wu, S., Li, A., & Ruan, J. (2021). SMARTdenovo: a de novo assembler using long noisy reads. *Gigabyte, 1*. doi:10.46471/gigabyte.15

Liu, J., Zhao, X., Zhan, Y., Wang, K., Francis, F., & Liu, Y. (2021). New slow release mixture of (E)-β-farnesene with methyl salicylate to enhance aphid biocontrol efficacy in wheat ecosystem. *Pest Manag Sci, 77*(7), 3341-3348. doi:10.1002/ps.6378

Liu, S., Chang, H., Liu, W., Cui, W., Liu, Y., Wang, Y., Wang, G. (2020). Essential role for SNMP1 in detection of sex pheromones in Helicoverpa armigera. *Insect Biochem Mol Biol, 127*, 103485. doi:10.1016/j.ibmb.2020.103485

Liu, Y., Cui, Z., Si, P., Liu, Y., Zhou, Q., & Wang, G. (2020). Characterization of a specific odorant receptor for linalool in the Chinese citrus fly Bactrocera minax (Diptera: Tephritidae). *Insect Biochem Mol Biol, 122*, 103389. doi:10.1016/j.ibmb.2020.103389

Liu, Y., Cui, Z., Wang, G., Zhou, Q., & Liu, Y. (2020). Cloning and Functional Characterization of Three Odorant Receptors From the Chinese Citrus fly Bactrocera minax (Diptera: Tephritidae). *Front Physiol, 11*, 246. doi:10.3389/fphys.2020.00246

Liu, Y., Liu, C., Lin, K., & Wang, G. (2013). Functional specificity of sex pheromone receptors in the cotton bollworm Helicoverpa armigera. *PLoS One, 8*(4), e62094. doi:10.1371/journal.pone.0062094

Liu, Y., Liu, H., Wang, H., Huang, T., Liu, B., Yang, B., Wang, G. (2020). Apolygus lucorum genome provides insights into omnivorousness and mesophyll feeding. *Molecular ecology resources*. doi:10.1111/1755-0998.13253

Liu, Y., Liu, Y., Jiang, X., & Wang, G. (2018). Cloning and functional characterization of three new pheromone receptors from the diamondback moth, Plutella xylostella. *J Insect Physiol, 107*, 14-22. doi:10.1016/j.jinsphys.2018.02.005

Liu, Z., Hu, T., Guo, H. W., Liang, X. F., & Cheng, Y. Q. (2021). Ultrastructure of the Olfactory Sensilla across the Antennae and Maxillary Palps of Bactrocera dorsalis (Diptera: Tephritidae). *Insects, 12*(4). doi:10.3390/insects12040289

Livak, K. J., & Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods, 25*(4), 402-408. doi:10.1006/meth.2001.1262

Lokeshwari, D., Krishna Kumar, N. K., & Manjunatha, H. (2016). Multiple Mutations on the Second Acetylcholinesterase Gene Associated With Dimethoate Resistance in the Melon Aphid, Aphis gossypii (Hemiptera: Aphididae). *J Econ Entomol, 109*(2), 887-897. doi:10.1093/jee/tov403

Loughrin, J. H., Manukian, A., Heath, R. R., & Tumlinson, J. H. (1995). Volatiles emitted by different cotton varieties damaged by feeding beet armyworm larvae. *J Chem Ecol, 21*(8), 1217-1227. doi:10.1007/BF02228321

Loughrin, J. H., Potter, D. A., & Hamilton-Kemp, T. R. (1995). Volatile compounds induced by herbivory act as aggregation kairomones for the Japanese beetle (Popillia japonica Newman). *J Chem Ecol, 21*(10), 1457-1467. doi:10.1007/BF02035145

Lundin, C., Käll, L., Kreher, S. A., Kapp, K., Sonnhammer, E. L., Carlson, J. R., Nilsson, I. (2007). Membrane topology of the Drosophila OR83b odorant receptor. *FEBS Lett, 581*(29), 5601-5604. doi:10.1016/j.febslet.2007.11.007

Mallinger, R. E., Hogg, D. B., & Gratton, C. (2011). Methyl salicylate attracts natural enemies and reduces populations of soybean aphids (Hemiptera: Aphididae) in soybean agroecosystems. *J Econ Entomol, 104*(1), 115-124. doi:10.1603/ec10253

Manzano-Mari, N. A., Coeur d'acier, A., Clamens, A. L., Orvain, C., Cruaud, C., Barbe, V., & Jousselin, E. (2020). Serial horizontal transfer of vitamin-biosynthetic genes enables the establishment of new nutritional symbionts in aphids' di-symbiotic systems. *ISME J, 14*(1), 259-273. doi:10.1038/s41396-019-0533-6

Marçais, G., & Kingsford, C. (2011). A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics, 27*(6), 764-770. doi:10.1093/bioinformatics/btr011

Martin, F., Boto, T., Gomez-Diaz, C., & Alcorta, E. (2013). Elements of olfactory reception in adult Drosophila melanogaster. *Anat Rec (Hoboken), 296*(9), 1477-1488. doi:10.1002/ar.22747

Mathers, T. C. (2020). Improved Genome Assembly and Annotation of the Soybean Aphid (Aphis glycines Matsumura). *G3 (Bethesda, Md.), 10*(3), 899-906. doi:10.1534/g3.119.400954

Mathers, T. C., Chen, Y., Kaithakottil, G., Legeai, F., Mugford, S. T., Baa-Puyoulet, P., Hogenhout, S. A. (2017). Rapid transcriptional plasticity of duplicated gene clusters enables a clonally reproducing aphid to colonise diverse plant species. *Genome Biol, 18*(1), 27. doi:10.1186/s13059-016-1145-3

Mathers, T. C., Mugford, S. T., Hogenhout, S. A., & Tripathi, L. (2020). Genome Sequence of the Banana Aphid, Pentalonia nigronervosa Coquerel (Hemiptera: Aphididae) and Its Symbionts. *G3 (Bethesda), 10*(12), 4315-4321. doi:10.1534/g3.120.401358

Mathers, T. C., Wouters, R. H. M., Mugford, S. T., Swarbreck, D., van Oosterhout, C., & Hogenhout, S. A. (2021). Chromosome-Scale Genome Assemblies of Aphids Reveal Extensively Rearranged Autosomes and Long-Term Conservation of the X Chromosome. *Mol Biol Evol, 38*(3), 856-875. doi:10.1093/molbev/msaa246

Matthes, M. C., Bruce, T. J., Ton, J., Verrier, P. J., Pickett, J. A., & Napier, J. A. (2010). The transcriptome of cis-jasmone-induced resistance in Arabidopsis thaliana and its role in indirect defence. *Planta, 232*(5), 1163-1180. doi:10.1007/s00425-010-1244-4

Mauck, K. E., De Moraes, C. M., & Mescher, M. C. (2010). Deceptive chemical signals induced by a plant virus attract insect vectors to inferior hosts.

Proceedings of the National Academy of Sciences, 107(8), 3600-3605.

McBride, C. S. (2007). Rapid evolution of smell and taste receptor genes during host specialization in Drosophila sechellia. *Proc Natl Acad Sci U S A, 104*(12), 4996-5001. doi:10.1073/pnas.0608424104

McBride, C. S., Arguello, J. R., & O'Meara, B. C. (2007). Five Drosophila genomes reveal nonneutral evolution and the signature of host specialization in the chemoreceptor superfamily. *Genetics, 177*(3), 1395-1416. doi:10.1534/genetics.107.078683

Miazzi, F., Hoyer, C., Sachse, S., Knaden, M., Wicher, D., Hansson, B. S., & Lavista-Llanos, S. (2019). Optimization of Insect Odorant Receptor Trafficking and Functional Expression Via Transient Transfection in HEK293 Cells. *Chem Senses, 44*(9), 673-682. doi:10.1093/chemse/bjz062

Michaud, J. P. (2022). The Ecological Significance of Aphid Cornicles and Their Secretions. *Annu Rev Entomol, 67*, 65-81. doi:10.1146/annurev-ento-033021-094437

Missbach, C., Dweck, H. K., Vogel, H., Vilcinskas, A., Stensmyr, M. C., Hansson, B. S., & Grosse-Wilde, E. (2014). Evolution of insect olfactory receptors. *Elife, 3*, e02115. doi:10.7554/eLife.02115

Mistry, J., Finn, R. D., Eddy, S. R., Bateman, A., & Punta, M. (2013). Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. *Nucleic Acids Res, 41*(12), e121. doi:10.1093/nar/gkt263

Mitchell, R. F., Schneider, T. M., Schwartz, A. M., Andersson, M. N., & McKenna, D. D. (2020). The diversity and evolution of odorant receptors in beetles (Coleoptera). *Insect Mol Biol, 29*(1), 77-91. doi:10.1111/imb.12611

Montagné, N., Chertemps, T., Brigaud, I., François, A., François, M. C., de Fouchier, A., Jacquin-Joly, E. (2012). Functional characterization of a sex pheromone receptor in the pest moth Spodoptera littoralis by heterologous expression in Drosophila. *Eur J Neurosci, 36*(5), 2588-2596. doi:10.1111/j.1460-9568.2012.08183.x

Montagné, N., de Fouchier, A., Newcomb, R. D., & Jacquin-Joly, E. (2015). Advances in the identification and characterization of olfactory receptors in insects. *Prog Mol Biol Transl Sci, 130*, 55-80. doi:10.1016/bs.pmbts.2014.11.003

Moon, S. J., Kottgen, M., Jiao, Y., Xu, H., & Montell, C. (2006). A taste receptor required for the caffeine response in vivo. *Curr Biol, 16*(18), 1812-1817. doi:10.1016/j.cub.2006.07.024

Moran, N. A., & Jarvik, T. (2010). Lateral transfer of genes from fungi underlies carotenoid production in aphids. *Science, 328*(5978), 624-627. doi:10.1126/science.1187113

Moreno, A., Garzo, E., Fernandez-Mata, G., Kassem, M., Aranda, M. A., & Fereres, A. (2011). Aphids secrete watery saliva into plant tissues from the onset of stylet penetration. *Entomologia Experimentalis et Applicata, 139*(2), 145-153. doi:10.1111/j.1570-7458.2011.01117.x

Murugathas, T., Zheng, H. Y., Colbert, D., Kralicek, A. V., Carraher, C., & Plank, N. O. V. (2019). Biosensing with Insect Odorant Receptor Nanodiscs and

Carbon Nanotube Field-Effect Transistors. *ACS Appl Mater Interfaces, 11*(9), 9530-9538. doi:10.1021/acsami.8b19433

Nault, L. R., Montgomery, M. E., & Bowers, W. S. (1976). Ant-Aphid Association: Role of Aphid Alarm Pheromone. *Science, 192*(4246), 1349-1351. doi:doi:10.1126/science.1273595

Nebreda, M., Moreno, A., Pérez, N., Palacios, I., Seco-Fernández, V., & Fereres, A. (2004). Activity of aphids associated with lettuce and broccoli in Spain and their efficiency as vectors of Lettuce mosaic virus. *Virus Res, 100*(1), 83-88. doi:10.1016/j.virusres.2003.12.016

Nef, S., Allaman, I., Fiumelli, H., De Castro, E., & Nef, P. (1996). Olfaction in birds: differential embryonic expression of nine putative odorant receptor genes in the avian olfactory system. *Mech Dev, 55*(1), 65-77. doi:10.1016/0925-4773(95)00491-2

Ngai, J., Chess, A., Dowling, M. M., Necles, N., Macagno, E. R., & Axel, R. (1993). Coding of olfactory information: topography of odorant receptor expression in the catfish olfactory epithelium. *Cell, 72*(5), 667-680. doi:10.1016/0092-8674(93)90396-8

Nguyen, L. T., Schmidt, H. A., von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol, 32*(1), 268-274. doi:10.1093/molbev/msu300

Nicholson, S. J., Nickerson, M. L., Dean, M., Song, Y., Hoyt, P. R., Rhee, H., . . . Puterka, G. J. (2015). The genome of Diuraphis noxia, a global aphid pest of small grains. *BMC Genomics, 16*, 429. doi:10.1186/s12864-015-1525-1

Nishino, C., Bowers, W. S., Montgomery, M. E., Nault, L. R., & Nielson, M. W. (1977). Alarm pheromone of the spotted alfalfa aphid,Therioaphis maculata buckton (Homoptera: Aphididae). *Journal of Chemical Ecology, 3*(3), 349-357. doi:10.1007/BF00988450

Norman, V. C., Butterfield, T., Drijfhout, F., Tasman, K., & Hughes, W. O. (2017). Alarm Pheromone Composition and Behavioral Activity in Fungus-Growing Ants. *J Chem Ecol, 43*(3), 225-235. doi:10.1007/s10886-017-0821-4

Northey, T., Venthur, H., De Biasio, F., Chauviac, F.-X., Cole, A., Ribeiro, K. A. L., Zhou, J.-J. (2016). Crystal Structures and Binding Dynamics of Odorant-Binding Protein 3 from two aphid species Megoura viciae and Nasonovia ribisnigri. *Scientific Reports, 6*, 24739. doi:10.1038/srep24739

Nouvian, M., Reinhard, J., & Giurfa, M. (2016). The defensive response of the honeybee Apis mellifera. *J Exp Biol, 219*(Pt 22), 3505-3517. doi:10.1242/jeb.143016

Nozawa, M., & Nei, M. (2007). Evolutionary dynamics of olfactory receptor genes in Drosophila species. *Proc Natl Acad Sci U S A, 104*(17), 7122-7127. doi:10.1073/pnas.0702133104

O'Donnell, D. J. (1987). Larval development and the determination of the number of instars in aphid parasitoids (Hymenoptera: Aphidiidae). *International Journal of Insect Morphology and Embryology, 16*(1), 3-15.

doi:https://doi.org/10.1016/0020-7322(87)90052-3

Oh, H. W., Jeong, S. A., Kim, J., & Park, K. C. (2019). Morphological and functional heterogeneity in olfactory perception between antennae and maxillary palps in the pumpkin fruit fly, Bactrocera depressa. *Arch Insect Biochem Physiol, 101*(4), e21560. doi:10.1002/arch.21560

Olafson, P. U. (2013). Molecular characterization and immunolocalization of the olfactory co-receptor Orco from two blood-feeding muscid flies, the stable fly (Stomoxys calcitrans, L.) and the horn fly (Haematobia irritans irritans, L.). *Insect Mol Biol, 22*(2), 131-142. doi:10.1111/imb.12009

Pan, H., Yang, X., Romeis, J., Siegfried, B. D., & Zhou, X. (2020). Dietary RNAi toxicity assay exhibits differential responses to ingested dsRNAs among lady beetles. *Pest Manag Sci, 76*(11), 3606-3614. doi:10.1002/ps.5894

Papke, R. L., & Stokes, C. (2010). Working with OpusXpress: methods for high volume oocyte experiments. *Methods, 51*(1), 121-133. doi:10.1016/j.ymeth.2010.01.012

Pareja, M., Mohib, A., Birkett, M. A., Dufour, S., & Glinwood, R. T. (2009). Multivariate statistics coupled to generalized linear models reveal complex use of chemical cues by a parasitoid. *Animal Behaviour, 77*(4), 901-909. doi:10.1016/j.anbehav.2008.12.016

Parra, G., Bradnam, K., & Korf, I. (2007). CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics, 23*(9), 1061-1067. doi:10.1093/bioinformatics/btm071

Pask, G. M., Slone, J. D., Millar, J. G., Das, P., Moreira, J. A., Zhou, X., . . . Ray, A. (2017). Specialized odorant receptors in social insects that detect cuticular hydrocarbon cues and candidate pheromones. *Nat Commun, 8*(1), 297. doi:10.1038/s41467-017-00099-1

Paulo, D. F., Junqueira, A. C. M., Arp, A. P., Vieira, A. S., Ceballos, J., Skoda, S. R., Azeredo-Espin, A. M. L. (2021). Disruption of the odorant coreceptor Orco impairs foraging and host finding behaviors in the New World screwworm fly. *Sci Rep, 11*(1), 11379. doi:10.1038/s41598-021-90649-x

Peccoud, J., Ollivier, A., Plantegenest, M., & Simon, J. C. (2009). A continuum of genetic divergence from sympatric host races to species in the pea aphid complex. *Proc Natl Acad Sci U S A, 106*(18), 7495-7500. doi:10.1073/pnas.0811117106

Pelosi, P., Iovinella, I., Zhu, J., Wang, G., & Dani, F. R. (2018). Beyond chemoreception: diverse tasks of soluble olfactory proteins in insects. *Biol Rev Camb Philos Soc, 93*(1), 184-200. doi:10.1111/brv.12339

Pelosi, P., Iovinella, I., Zhu, J., Wang, G., & Dani, F. R. (2018). Beyond chemoreception: diverse tasks of soluble olfactory proteins in insects. *Biological reviews of the Cambridge Philosophical Society, 93*(1), 184-200. doi:10.1111/brv.12339

Pelosi, P., Zhou, J. J., Ban, L. P., & Calvello, M. (2006). Soluble proteins in insect chemical communication. *Cell Mol Life Sci, 63*(14), 1658-1676. doi:10.1007/s00018-005-5607-0

Pertea, G., Huang, X., Liang, F., Antonescu, V., Sultana, R., Karamycheva, S.,

Quackenbush, J. (2003). TIGR Gene Indices clustering tools (TGICL): a software system for fast clustering of large EST datasets. *Bioinformatics, 19*(5), 651-652. doi:10.1093/bioinformatics/btg034

Petersen, T. N., Brunak, S., von Heijne, G., & Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods, 8*(10), 785-786. doi:10.1038/nmeth.1701

Pickett, J. A., & Griffiths, D. C. (1980). Composition of aphid alarm pheromones. *Journal of Chemical Ecology, 6*(2), 349-360. doi:10.1007/bf01402913

Pickett, J. A., Rasmussen, H. B., Woodcock, C. M., Matthes, M., & Napier, J. A. (2003). Plant stress signalling: understanding and exploiting plant-plant interactions. *Biochem Soc Trans, 31*(Pt 1), 123-127. doi:10.1042/bst0310123

Pitts, R. J., Fox, A. N., & Zwiebel, L. J. (2004). A highly conserved candidate chemoreceptor expressed in both olfactory and gustatory tissues in the malaria vector Anopheles gambiae. *Proc Natl Acad Sci U S A, 101*(14), 5058-5063. doi:10.1073/pnas.0308146101

Powell, W., & Pickett, J. A. (2003). Manipulation of parasitoids for aphid pest management: progress and prospects. *Pest Manag Sci, 59*(2), 149-155. doi:10.1002/ps.550

Prieto-Godino, L. L., Rytz, R., Cruchet, S., Bargeton, B., Abuin, L., Silbering, A. F., Benton, R. (2017). Evolution of Acid-Sensing Olfactory Circuits in Drosophilids. *Neuron, 93*(3). doi:10.1016/j.neuron.2016.12.024

Pym, A., Umina, P. A., Reidy-Crofts, J., Troczka, B. J., Matthews, A., Gardner, J., Bass, C. (2022). Overexpression of UDP-glucuronosyltransferase and cytochrome P450 enzymes confers resistance to sulfoxaflor in field populations of the aphid, Myzus persicae. *Insect Biochem Mol Biol, 143*, 103743. doi:10.1016/j.ibmb.2022.103743

Qiao, H., Tuccori, E., He, X., Gazzano, A., Field, L., Zhou, J., & Pelosi, P. (2009). Discrimination of alarm pheromone (E)-beta-farnesene by aphid odorant-binding proteins. *Insect Biochemistry and Molecular Biology, 39*(5-6), 414-419. doi:10.1016/j.ibmb.2009.03.004

Qin, Y., Yang, Z., Song, D., Wang, Q., Gu, S., Li, W., Yang, X. (2020). Bioactivities of synthetic salicylate-substituted carboxyl (E)-β-Farnesene derivatives as ecofriendly agrochemicals and their binding mechanism with potential targets in aphid olfactory system. *Pest Management Science, 76*(7), 2465-2472. doi:10.1002/ps.5787

Quiroz, A., & Niemeyer, H. (1998). Olfactometer-assessed responses of aphid Rhopalosiphum padi to wheat and oat volatiles. *Journal of Chemical Ecology, 24*, 113-124.

Radović, J., Sokolović, D., & Marković, J. (2009). Alfalfa-most important perennial forage legume in animal husbandry. *Biotehnologija u stocarstvu, 25*(5-6-1), 465-475. doi:10.2298/bah0906465r

Ragsdale, D. W., Voegtlin, D. J., & O'Neil, R. J. (2004). Soybean Aphid Biology in North America. *Annals of the Entomological Society of America, 97*(2), 204-208. doi:10.1093/aesa/97.2.204

Ramsden, M., Menendez, R., Leather, S., & Wäckers, F. (2017). Do natural enemies really make a difference? Field scale impacts of parasitoid wasps and hoverfly larvae on cereal aphid populations. *Agricultural and Forest Entomology, 19*(2), 139-145.

Ray, S., Helms, A. M., Matulis, N. L., Davidson-Lowe, E., Grisales, W., & Ali, J. G. (2020). Asymmetry in Herbivore Effector Responses: Caterpillar Frass Effectors Reduce Performance of a Subsequent Herbivore. *J Chem Ecol, 46*(1), 76-83. doi:10.1007/s10886-019-01131-y

Renou, M., & Anton, S. (2020). Insect olfactory communication in a complex and changing world. *Curr Opin Insect Sci, 42*, 1-7. doi:10.1016/j.cois.2020.04.004

Roberts, R. E., Biswas, T., Yuvaraj, J. K., Grosse-Wilde, E., Powell, D., Hansson, B. S., Andersson, M. N. (2022). Odorant receptor orthologues in conifer-feeding beetles display conserved responses to ecologically relevant odours. *Molecular ecology, 31*(13), 3693-3707. doi:https://doi.org/10.1111/mec.16494

Robertson, H. M. (2015). The Insect Chemoreceptor Superfamily Is Ancient in Animals. *Chem Senses, 40*(9), 609-614. doi:10.1093/chemse/bjv046

Robertson, H. M. (2019). Molecular Evolution of the Major Arthropod Chemoreceptor Gene Families. *Annu Rev Entomol, 64*, 227-242. doi:10.1146/annurev-ento-020117-043322

Robertson, H. M., Robertson, E. C. N., Walden, K. K. O., Enders, L. S., & Miller, N. J. (2019). The chemoreceptors and odorant binding proteins of the soybean and pea aphids. *Insect Biochem. Mol. Biol., 105*, 69-78. doi:10.1016/j.ibmb.2019.01.005

Robertson, H. M., Warr, C. G., & Carlson, J. R. (2003). Molecular evolution of the insect chemoreceptor gene superfamily in Drosophila melanogaster. *Proc Natl Acad Sci U S A, 100 Suppl 2*(Suppl 2), 14537-14542. doi:10.1073/pnas.2335847100

Rose, U. S., & Tumlinson, J. H. (2004). Volatiles released from cotton plants in response to Helicoverpa zea feeding damage on cotton flower buds. *Planta, 218*(5), 824-832. doi:10.1007/s00425-003-1162-9

Rytz, R., Croset, V., & Benton, R. (2013). Ionotropic receptors (IRs): chemosensory ionotropic glutamate receptors in Drosophila and beyond. *Insect Biochem Mol Biol, 43*(9), 888-897. doi:10.1016/j.ibmb.2013.02.007

Sanderson, M. J. (2003). r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics, 19*(2), 301-302. doi:10.1093/bioinformatics/19.2.301

Sandler, B. H., Nikonova, L., Leal, W. S., & Clardy, J. (2000). Sexual attraction in the silkworm moth: structure of the pheromone-binding-protein–bombykol complex. *Chemistry & Biology, 7*(2), 143-151. doi:10.1016/s1074-5521(00)00078-8

Sarang, K., Rudziński, K. J., & Szmigielski, R. (2021). Green Leaf Volatiles in the Atmosphere—Properties, Transformation, and Significance. *Atmosphere, 12*(12), 1655.

Sato, K., Pellegrino, M., Nakagawa, T., Nakagawa, T., Vosshall, L. B., & Touhara, K. (2008). Insect olfactory receptors are heteromeric ligand-gated ion channels. *Nature, 452*(7190), 1002-1006. doi:10.1038/nature06850

Schmidt, H. R., & Benton, R. (2020). Molecular mechanisms of olfactory detection in insects: beyond receptors. *Open Biol, 10*(10), 200252. doi:10.1098/rsob.200252

Schröder, M. L., Glinwood, R., Webster, B., Ignell, R., & Krüger, K. (2015). Olfactory responses of R hopalosiphum padi to three maize, potato, and wheat cultivars and the selection of prospective crop border plants. *Entomologia Experimentalis et Applicata, 157*(2), 241-253.

Schwartzberg, E. G., Boroczky, K., & Tumlinson, J. H. (2011). Pea aphids, Acyrthosiphon pisum, suppress induced plant volatiles in broad bean, Vicia faba. *J Chem Ecol, 37*(10), 1055-1062. doi:10.1007/s10886-011-0006-5

Seo, M. J., & Youn, Y. N. (2002). Effective preservation methods of the Asian ladybird, Harmonia axyridis (Coleoptera: Coccinellidae), as an application strategy for the biological control of aphids. *Journal of Asia-Pacific Entomology, 5*(2), 209-214.

Servant, N., Varoquaux, N., Lajoie, B. R., Viara, E., Chen, C. J., Vert, J. P., Barillot, E. (2015). HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol, 16*, 259. doi:10.1186/s13059-015-0831-x

Shambaugh, G. F., Frazier, J. L., Castell, A. E., & Coons, L. B. (1978). Antennal sensilla of seventeen aphid species (Homoptera: Aphidinae). *International Journal of Insect Morphology and Embryology, 7*(5-6), 389-404.

Shang, F., Niu, J., Ding, B. Y., Zhang, W., Wei, D. D., Wei, D., Wang, J. J. (2020). The miR-9b microRNA mediates dimorphism and development of wing in aphids. *Proc Natl Acad Sci U S A*. doi:10.1073/pnas.1919204117

Shih, P. Y., Sugio, A., & Simon, J. C. (2023). Molecular Mechanisms Underlying Host Plant Specificity in Aphids. *Annu Rev Entomol, 68*, 431-450. doi:10.1146/annurev-ento-120220-020526

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics, 31*(19), 3210-3212. doi:10.1093/bioinformatics/btv351

Slone, J. D., Pask, G. M., Ferguson, S. T., Millar, J. G., Berger, S. L., Reinberg, D., Zwiebel, L. J. (2017). Functional characterization of odorant receptors in the ponerine ant, Harpegnathos saltator. *Proc Natl Acad Sci U S A, 114*(32), 8586-8591. doi:10.1073/pnas.1704647114

Smadja, C., Shi, P., Butlin, R. K., & Robertson, H. M. (2009). Large gene family expansions and adaptive evolution for odorant and gustatory receptors in the pea aphid, Acyrthosiphon pisum. *Mol Biol Evol, 26*(9), 2073-2086. doi:10.1093/molbev/msp116

Smadja, C. M., Canback, B., Vitalis, R., Gautier, M., Ferrari, J., Zhou, J. J., & Butlin, R. K. (2012). Large-scale candidate gene scan reveals the role of chemoreceptor genes in host plant specialization and speciation in the pea aphid. *Evolution, 66*(9), 2723-2738. doi:10.1111/j.1558-

5646.2012.01612.x

Smallegange, R. C., Kelling, F. J., & Den Otter, C. J. (2008). Types and numbers of sensilla on antennae and maxillary palps of small and large houseflies, Musca domestica (Diptera, Muscidae). *Microsc Res Tech, 71*(12), 880-886. doi:10.1002/jemt.20636

Smart, R., Kiely, A., Beale, M., Vargas, E., Carraher, C., Kralicek, A. V., Warr, C. G. (2008). Drosophila odorant receptors are novel seven transmembrane domain proteins that can signal independently of heterotrimeric G proteins. *Insect Biochem Mol Biol, 38*(8), 770-780. doi:10.1016/j.ibmb.2008.05.002

Smith, C. M., & Chuang, W.-P. (2014). Plant resistance to aphid feeding: behavioral, physiological, genetic and molecular cues regulate aphid host selection and feeding. *Pest Management Science, 70*(4), 528-540. doi:https://doi.org/10.1002/ps.3689

Smith, T. E., Li, Y., Perreau, J., & Moran, N. A. (2022). Elucidation of host and symbiont contributions to peptidoglycan metabolism based on comparative genomics of eight aphid subfamilies and their Buchnera. *PLoS Genet, 18*(5), e1010195. doi:10.1371/journal.pgen.1010195

Sobhy, I. S., Woodcock, C. M., Powers, S. J., Caulfield, J. C., Pickett, J. A., & Birkett, M. A. (2017). cis-Jasmone Elicits Aphid-Induced Stress Signalling in Potatoes. *J Chem Ecol, 43*(1), 39-52. doi:10.1007/s10886-016-0805-9

Soffan, A., Subandiyah, S., Makino, H., Watanabe, T., & Horiike, T. (2018). Evolutionary Analysis of the Highly Conserved Insect Odorant Coreceptor (Orco) Revealed a Positive Selection Mode, Implying Functional Flexibility. *J Insect Sci, 18*(6). doi:10.1093/jisesa/iey120

Song, H. G., Young Kwon, J., Soo Han, H., Bae, Y. C., & Moon, C. (2008). First Contact to Odors: Our Current Knowledge about Odorant Receptor. *Sensors (Basel), 8*(10), 6303-6320. doi:10.3390/s8106303

Song, L., Wang, X., Liu, Y., Sun, Y., & Ban, L. (2020). Characterization of antennal sensilla and immunolocalization of odorant-binding proteins on spotted alfalfa aphid, Therioaphis trifolii (Monell). *Frontiers in Physiology, 11*, 606575.

Song, X., Qin, Y. G., Yin, Y., & Li, Z. X. (2021). Identification and Behavioral Assays of Alarm Pheromone in the Vetch Aphid Megoura viciae. *J Chem Ecol, 47*(8-9), 740-746. doi:10.1007/s10886-021-01297-4

Sparks, J. T., Bohbot, J. D., & Dickens, J. C. (2015). Olfactory disruption: toward controlling important insect vectors of disease. *Prog Mol Biol Transl Sci, 130*, 81-108. doi:10.1016/bs.pmbts.2014.11.004

Speth, Z., Kaur, G., Mazolewski, D., Sisomphou, R., Siao, D. D. C., Pooraiiouby, R., Nuss, A. B. (2021). Characterization of Anopheles stephensi Odorant Receptor 8, an Abundant Component of the Mouthpart Chemosensory Transcriptome. *Insects, 12*(7). doi:10.3390/insects12070593

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics, 30*(9), 1312-1313. doi:10.1093/bioinformatics/btu033

Stanke, M., Diekhans, M., Baertsch, R., & Haussler, D. (2008). Using native and

syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics, 24*(5), 637-644. doi:10.1093/bioinformatics/btn013

Steinbrecht, R. A. (1997). Pore structures in insect olfactory sensilla: a review of data and concepts. *International Journal of Insect Morphology and Embryology, 26*(3-4), 229-245.

Suh, E., Bohbot, J., & Zwiebel, L. J. (2014). Peripheral olfactory signaling in insects. *Curr Opin Insect Sci, 6*, 86-92. doi:10.1016/j.cois.2014.10.006

Sun, H., Liu, F., Ye, Z., Baker, A., & Zwiebel, L. J. (2020). Mutagenesis of the orco odorant receptor co-receptor impairs olfactory function in the malaria vector Anopheles coluzzii. *Insect Biochem Mol Biol, 127*, 103497. doi:10.1016/j.ibmb.2020.103497

Sun, J. S., Xiao, S., & Carlson, J. R. (2018). The diverse small proteins called odorant-binding proteins. *Open biology, 8*(12), 180208. doi:10.1098/rsob.180208

Sun, M., Liu, Y., Walker, W. B., Liu, C., Lin, K., Gu, S., Wang, G. (2013). Identification and characterization of pheromone receptors and interplay between receptors and pheromone binding proteins in the diamondback moth, Plutella xyllostella. *PLoS One, 8*(4), e62098. doi:10.1371/journal.pone.0062098

Sung, H. Y., Jeong, Y. T., Lim, J. Y., Kim, H., Oh, S. M., Hwang, S. W., Moon, S. J. (2017). Heterogeneity in the Drosophila gustatory receptor complexes that detect aversive compounds. *Nat Commun, 8*(1), 1484. doi:10.1038/s41467-017-01639-5

Sunnucks, P., Driver, F., Brown, W. V., Carver, M., Hales, D. F., & Milne, W. M. (1997). Biological and genetic characterization of morphologically similar Therioaphis trifolii (Hemiptera: Aphididae) with different host utilization. *Bulletin of Entomological Research, 87*(4), 425-436. doi:10.1017/S0007485300037433

Syed, Z., Kopp, A., Kimbrell, D. A., & Leal, W. S. (2010). Bombykol receptors in the silkworm moth and the fruit fly. *Proc Natl Acad Sci U S A, 107*(20), 9436-9439. doi:10.1073/pnas.1003881107

Takabayashi, J. (2022). Herbivory-Induced Plant Volatiles Mediate Multitrophic Relationships in Ecosystems. *Plant Cell Physiol, 63*(10), 1344-1355. doi:10.1093/pcp/pcac107

Takabayashi, J., & Shiojiri, K. (2019). Multifunctionality of herbivory-induced plant volatiles in chemical communication in tritrophic interactions. *Curr Opin Insect Sci, 32*, 110-117. doi:10.1016/j.cois.2019.01.003

Takemoto, H., & Takabayashi, J. (2015). Parasitic Wasps Aphidius ervi are More Attracted to a Blend of Host-Induced Plant Volatiles than to the Independent Compounds. *J Chem Ecol, 41*(9), 801-807. doi:10.1007/s10886-015-0615-5

Tang, S., Lomsadze, A., & Borodovsky, M. (2015). Identification of protein coding regions in RNA transcripts. *Nucleic Acids Research, 43*(12), e78-e78.

Task, D., Lin, C., Vulpe, A., Afify, A., Ballou, S., Brbic, M., Potter, C. J. (2022). Chemoreceptor co-expression in Drosophila melanogaster olfactory

neurons. *Elife, 11*. doi:10.7554/eLife.72599

Thorpe, P., Escudero-Martinez, C. M., Cock, P. J. A., Eves-van den Akker, S., & Bos, J. I. B. (2018). Shared Transcriptional Control and Disparate Gain and Loss of Aphid Parasitism Genes. *Genome Biol Evol, 10*(10), 2716-2733. doi:10.1093/gbe/evy183

Thorvaldsdóttir, H., Robinson, J. T., & Mesirov, J. P. (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform, 14*(2), 178-192. doi:10.1093/bib/bbs017

Troemel, E. R., Chou, J., Dwyer, N. D., Colbert, H. A., & Bargmann, C. I. (1995). Divergent seven transmembrane receptors are candidate chemosensory receptors in C. elegans. *Cell, 83*(2), 207-218. doi:10.1016/0092-8674(95)90162-0

Tsirigos, K. D., Peters, C., Shu, N., Käll, L., & Elofsson, A. (2015). The TOPCONS web server for consensus prediction of membrane protein topology and signal peptides. *Nucleic Acids Res, 43*(W1), W401-407. doi:10.1093/nar/gkv485

Tsitoura, P., Andronopoulou, E., Tsikou, D., Agalou, A., Papakonstantinou, M. P., Kotzia, G. A., Iatrou, K. (2010). Expression and membrane topology of Anopheles gambiae odorant receptors in lepidopteran insect cells. *PLoS One, 5*(11), e15428. doi:10.1371/journal.pone.0015428

Turlings, T. C., & Tumlinson, J. H. (1992). Systemic release of chemical signals by herbivore-injured corn. *Proc Natl Acad Sci U S A, 89*(17), 8399-8402. doi:10.1073/pnas.89.17.8399

Turlings, T. C. J., & Erb, M. (2018). Tritrophic Interactions Mediated by Herbivore-Induced Plant Volatiles: Mechanisms, Ecological Relevance, and Application Potential. *Annu Rev Entomol, 63*, 433-452. doi:10.1146/annurev-ento-020117-043507

van Poecke, R. M., & Dicke, M. (2004). Indirect defence of plants against herbivores: using Arabidopsis thaliana as a model plant. *Plant Biol (Stuttg), 6*(4), 387-401. doi:10.1055/s-2004-820887

Vandermoten, S., Mescher, M. C., Francis, F., Haubruge, E., & Verheggen, F. J. (2012). Aphid alarm pheromone: an overview of current knowledge on biosynthesis and functions. *Insect Biochem Mol Biol, 42*(3), 155-163. doi:10.1016/j.ibmb.2011.11.008

Vet, L. E., & Dicke, M. (1992). Ecology of infochemical use by natural enemies in a tritrophic context. *Annual Review of Entomology, 37*(1), 141-172.

Vosshall, L. B., & Hansson, B. S. (2011). A unified nomenclature system for the insect olfactory coreceptor. *Chem Senses, 36*(6), 497-498. doi:10.1093/chemse/bjr022

Wada-Katsumata, A., Robertson, H. M., Silverman, J., & Schal, C. (2018). Changes in the Peripheral Chemosensory System Drive Adaptive Shifts in Food Preferences in Insects. *Front Cell Neurosci, 12*, 281. doi:10.3389/fncel.2018.00281

Wang, B., Cao, S., Liu, W., & Wang, G. (2020). Effect of OBPs on the response of olfactory receptors. *Methods Enzymol, 642*, 279-300.

doi:10.1016/bs.mie.2020.04.065

Wang, B., Huang, T., Yao, Y., Francis, F., Yan, C., Wang, G., & Wang, B. (2022). A conserved odorant receptor identified from antennal transcriptome of Megoura crassicauda that specifically responds to cis-jasmone. *Journal of Integrative Agriculture, 21*(7), 2042-2054. doi:https://doi.org/10.1016/S2095-3119(21)63712-7

Wang, B., Liu, Y., He, K., & Wang, G. (2016). Comparison of research methods for functional characterization of insect olfactory receptors. *Scientific Reports, 6*(1). doi:10.1038/srep32806

Wang, B., Liu, Y., & Wang, G. (2017). Chemosensory genes in the antennal transcriptome of two syrphid species, Episyrphus balteatus and Eupeodes corollae (Diptera: Syrphidae). *BMC Genomics, 18*(1), 586. doi:10.1186/s12864-017-3939-4

Wang, C., Li, G., Miao, C., Zhao, M., Wang, B., & Guo, X. (2020). Nonanal modulates oviposition preference in female Helicoverpa assulta (Lepidoptera: Noctuidae) via the activation of peripheral neurons. *Pest Manag Sci*. doi:10.1002/ps.5870

Wang, G., Vasquez, G. M., Schal, C., Zwiebel, L. J., & Gould, F. (2011). Functional characterization of pheromone receptors in the tobacco budworm Heliothis virescens. *Insect Mol Biol, 20*(1), 125-133. doi:10.1111/j.1365-2583.2010.01045.x

Wang, L., Xie, J., Luo, Z., Niu, Y., Coulter, J., Zhang, R., & Lingling, L. (2020). Forage yield, water use efficiency, and soil fertility response to alfalfa growing age in the semiarid Loess Plateau of China. *Agricultural Water Management, 243*. doi:10.1016/j.agwat.2020.106415

Wang, Q., Zhou, J. J., Liu, J. T., Huang, G. Z., Xu, W. Y., Zhang, Q., Gu, S. H. (2019). Integrative transcriptomic and genomic analysis of odorant binding proteins and chemosensory proteins in aphids. *Insect Mol Biol, 28*(1), 1-22. doi:10.1111/imb.12513

Wang, X., Gao, Y., Chen, Z., Li, J., Huang, J., Cao, J., Ban, L. (2019). (E)-β-farnesene synthase gene affects aphid behavior in transgenic Medicago sativa. *Pest Manag Sci, 75*(3), 622-631. doi:10.1002/ps.5153

Wang, Y., Tang, H., Debarry, J. D., Tan, X., Li, J., Wang, X., Paterson, A. H. (2012). MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res, 40*(7), e49. doi:10.1093/nar/gkr1293

Webster, B., Bruce, T., Dufour, S., Birkemeyer, C., Birkett, M., Hardie, J., & Pickett, J. (2008). Identification of volatile compounds used in host location by the black bean aphid, Aphis fabae. *J Chem Ecol, 34*(9), 1153-1161. doi:10.1007/s10886-008-9510-7

Wen, M., Li, E., Chen, Q., Kang, H., Zhang, S., Li, K., Ren, B. (2019). A herbivore-induced plant volatile of the host plant acts as a collective foraging signal to the larvae of the meadow moth, Loxostege sticticalis (Lepidoptera: Pyralidae). *J Insect Physiol, 118*, 103941. doi:10.1016/j.jinsphys.2019.103941

Wenger, J. A., Cassone, B. J., Legeai, F., Johnston, J. S., Bansal, R., Yates, A. D., Michel, A. (2020). Whole genome sequence of the soybean aphid, Aphis glycines. *Insect Biochem Mol Biol, 123*, 102917. doi:10.1016/j.ibmb.2017.01.005

Wicher, D., Schäfer, R., Bauernfeind, R., Stensmyr, M. C., Heller, R., Heinemann, S. H., & Hansson, B. S. (2008). Drosophila odorant receptors are both ligand-gated and cyclic-nucleotide-activated cation channels. *Nature, 452*(7190), 1007-1011. doi:10.1038/nature06861

Wistrand, M., Käll, L., & Sonnhammer, E. L. (2006). A general model of G protein-coupled receptor sequences and its application to detect remote homologs. *Protein Sci, 15*(3), 509-521. doi:10.1110/ps.051745906

Wyatt, T. (2003). Pheromones and animal behaviour: communication by smell and taste (book). *41*. doi:10.1017/CBO9780511615061.002

Xiao, Y., An, X., Khashaveh, A., Shan, S., Wang, Q., Wang, S., Zhang, Y. (2020). Broadly Tuned Odorant Receptor AlinOR59 Involved in Chemoreception of Floral Scent in Adelphocoris lineolatus. *J Agric Food Chem, 68*(47), 13815-13823. doi:10.1021/acs.jafc.0c04434

Xie, W., He, C., Fei, Z., & Zhang, Y. (2020). Chromosome-level genome assembly of the greenhouse whitefly (Trialeurodes vaporariorum Westwood). *Mol Ecol Resour, 20*(4), 995-1006. doi:10.1111/1755-0998.13159

Xu, L., Dong, Z., Fang, L., Luo, Y., Wei, Z., Guo, H., Wang, Y. (2019). OrthoVenn2: a web server for whole-genome comparison and annotation of orthologous clusters across multiple species. *Nucleic Acids Res, 47*(W1), W52-w58. doi:10.1093/nar/gkz333

Xue, W., Fan, J., Zhang, Y., Xu, Q., Han, Z., Sun, J., & Chen, J. (2016). Identification and Expression Analysis of Candidate Odorant-Binding Protein and Chemosensory Protein Genes by Antennal Transcriptome of Sitobion avenae. *PLoS One, 11*(8), e0161839. doi:10.1371/journal.pone.0161839

Yang, C., Pan, H., Liu, Y., & Zhou, X. (2014). Selection of reference genes for expression analysis using quantitative real-time PCR in the pea aphid, Acyrthosiphon pisum (Harris) (Hemiptera, Aphidiae). *PLoS One, 9*(11), e110454. doi:10.1371/journal.pone.0110454

You, Y., Smith, D. P., Lv, M., & Zhang, L. (2016). A broadly tuned odorant receptor in neurons of trichoid sensilla in locust, Locusta migratoria. *Insect Biochem Mol Biol, 79*, 66-72. doi:10.1016/j.ibmb.2016.10.008

Zdobnov, E. M., & Apweiler, R. (2001). InterProScan--an integration platform for the signature-recognition methods in InterPro. *Bioinformatics, 17*(9), 847-848. doi:10.1093/bioinformatics/17.9.847

Zhang, D., Li, X., Liu, X., Wang, Q., & Pape, T. (2016). The antenna of horse stomach bot flies: morphology and phylogenetic implications (Oestridae, Gasterophilinae: Gasterophilus Leach). *Sci Rep, 6*, 34409. doi:10.1038/srep34409

Zhang, R., Wang, B., Grossi, G., Falabella, P., Liu, Y., Yan, S., Wang, G. (2017). Molecular Basis of Alarm Pheromone Detection in Aphids. *Curr Biol,*

27(1), 55-61. doi:10.1016/j.cub.2016.10.013

Zhang, R., Liu, Y., Yan, S., & Wang, G. (2019). Identification and functional characterization of an odorant receptor in pea aphid, Acyrthosiphon pisum. *Insect Sci, 26*(1), 58-67. doi:10.1111/1744-7917.12510

Zhang, S., Gao, X., Wang, L., Jiang, W., Su, H., Jing, T., Yang, Y. (2022). Chromosome-level genome assemblies of two cotton-melon aphid Aphis gossypii biotypes unveil mechanisms of host adaption. *Mol Ecol Resour, 22*(3), 1120-1134. doi:10.1111/1755-0998.13521

Zhao, H., Ullah, H., McNeill, M. R., Du, G., Hao, K., Tu, X., & Zhang, Z. (2019). Inhibitory Effects of Plant Trypsin Inhibitors Msti-94 and Msti-16 on Therioaphis trifolii (Monell) (Homoptera: Aphididae) in Alfalfa. *Insects, 10*(6), 154.

Zhao, J., Zhang, Y., Fan, D., & Feng, J. (2017). Identification and Expression Profiling of Odorant-Binding Proteins and Chemosensory Proteins of Daktulosphaira vitifoliae (Hemiptera: Phylloxeridae). *J Econ Entomol, 110*(4), 1813-1820. doi:10.1093/jee/tox121

Zheng, Y., Hu, Y., Yan, S., Zhou, H., Song, D., Yin, M., & Shen, J. (2019). A polymer/detergent formulation improves dsRNA penetration through the body wall and RNAi-induced mortality in the soybean aphid Aphis glycines. *Pest Manag Sci, 75*(7), 1993-1999. doi:10.1002/ps.5313

Zhong, T., Yin, J., Deng, S., Li, K., & Cao, Y. (2012). Fluorescence competition assay for the assessment of green leaf volatiles and trans-β-farnesene bound to three odorant-binding proteins in the wheat aphid Sitobion avenae (Fabricius). *Journal of insect physiology, 58*(6), 771-781. doi:10.1016/j.jinsphys.2012.01.011

Zhou, J., Vieira, F., He, X., Smadja, C., Liu, R., Rozas, J., & Field, L. (2010). Genome annotation and comparative analyses of the odorant-binding proteins and chemosensory proteins in the pea aphid Acyrthosiphon pisum. *Insect Mol Biol, 19 Suppl 2*, 113-122. doi:10.1111/j.1365-2583.2009.00919.x

Zhu, G., Zheng, M., Sun, J., Khuhro, S., Yan, Q., Huang, Y., Dong, S. (2019). CRISPR/Cas9 mediated gene knockout reveals a more important role of PBP1 than PBP2 in the perception of female sex pheromone components in Spodoptera litura. *Insect Biochem Mol Biol, 115*, 103244. doi:10.1016/j.ibmb.2019.103244