Identification and preliminary characterization of chemosensory-related proteins in the gall fly, *Procecidochares utilis* by transcriptomic analysis

**CURRENT STATUS:** POSTED

Lifang Li  
Yunnan Agricultural University

Xi Gao  
Yunnan Agricultural University

Huamin Gui  
Yunnan Agricultural University

Mingxian Lan  
Yunnan Agricultural University

Jiaying Zhu  
Southwest Forestry University

Yonghui Xie  
Kunming Branch of Yunnan Provincial Tobacco Company

Youguo Zhan  
Kunming Branch of Yunnan Provincial Tobacco Company

Zhijiang Wang  
Kunming Branch of Yunnan Provincial Tobacco Company

Li Zhengyue  
Yunnan Agricultural University

Min Ye  
Yunnan Agricultural University

Guoxing Wu  

wugx1@163.com

**Corresponding Author**

**ORCiD:** https://orcid.org/0000-0002-7308-2161
SUBJECT AREAS
Entomology

KEYWORDS
Procecidochares utilis, Transcriptome, Chemosensory genes, Expression profile
Abstract

**Background:** Chemoreception is critical for insect behaviors such as foraging, host searching and oviposition, and finding mating partners. The process of chemoreception is mediated by a series of proteins, including odorant-binding proteins (OBPs), gustatory receptors (GRs), odorant receptors (ORs), ionotrophic receptors (IRs), chemosensory proteins (CSPs) and sensory neuron membrane proteins (SNMPs). The tephritid stem gall fly, *Procecidochares utilis* Stone, is a type of egg parasitic insect, which is an effective biological control reagent for the invasive weed *Ageratina adenophora* in many countries. However, the study of molecular components related to the olfactory system of *P. utilis* has not been investigated. Here, we report the developmental transcriptome (egg, first-third instar larvae, pupae, fame and male adult) of *P. utilis* using next-generation sequencing technology and identified the major chemosensory related genes.

**Results:** In this study, a total of 133 chemosensory genes in *P. utilis* transcriptomes were identified by bioinformatics analysis, including 40 OBPs, 29 GRs, 24 ORs, 28 IRs, 6 CSPs, and 6 SNMPs in *P. utilis*. The sequences of these candidate chemosensory genes were confirmed by BLAST, and phylogenetic analysis was performed. The expression profiles of all candidate genes at different developmental stages were analyzed by differentially expressed genes (DEGs) analysis and then the expression profiles of the OBPs in the seven developmental stages were confirmed by real-time quantitative RT-PCR (qPCR). The results showed that the expression of candidate OBP genes across different developmental stages was consistent with the differentially expressed genes (DEGs) analysis using the fragments per kilobase per million fragments (FPKM) value.

**Conclusions:** A large number of chemosensory genes were identified. This study will provide a significant contribution to the molecular mechanism of chemoreception and help advance the use of *P. utilis* as biological control agents.

**Introduction:**
The chemosensory system is essential for the survival and reproduction of insects. Insects rely on the chemosensory system to identify and analyze volatiles, as well as to distinguish unique chemical signals in their living environment, which play vital roles in their ecological adaptations, host plant
location, mate selection, reproductive behavior and other important physiological processes [1, 2]. The system is mediated by several important chemosensory proteins, including odorant-binding proteins (OBPs), chemosensory proteins (CSPs), olfactory receptors (ORs), ionotropic receptors (IRs), and sensory neuron membrane proteins (SNMPs) [2].

OBPs are small, globular, and water-soluble proteins that are recognized as playing an important role in the first step of olfactory recognition [3–5]. Insect OBPs bind and transport hydrophobic odorant molecules through the aqueous lymph within its sensilla to the olfactory receptors (ORs) and activate ORs in the dendritic membranes of the olfactory neurons [6, 7]. Generally, insect OBPs contain three paired disulfide bridges formed by six highly conserved cysteine residues at the conserved position [8]. Since OBP was first identified in Antheraea polyphemus [9], a large number of OBP genes have been identified from different insect species [5, 10]. Studies have shown that OBPs are not only distributed in the lymph of insect chemosensory organs, such as the olfactory sensilla, but also in non-chemosensory organs, such as reproductive organs, eggs, abdomen, thorax and fat body [4, 6, 8, 10]. This suggests that OBPs may have different physiological functions including development, insecticide resistance and pheromone delivery in addition to chemoreception [4, 6, 11].

Insect CSPs are small, soluble, acidic proteins composed of five α-helices, four conserved cysteines, with two disulfide bridges [12]. Like OBPs, the CSPs are also regarded as the first step for the transportation of odorants in chemosensory recognition [13]. CSPs are broadly expressed in chemosensory organs, including antennae, probosces, legs, maxillary palps and wings, as well as non-chemosensory organs, such as pheromone glands and ejaculatory bulb, which indicate that CSPs may perform multiple physiological and behavioral functions [3].

Insect ORs were first discovered in the Drosophila melanogaster genome [14, 15]. Since then, a large number of receptor genes were identified from other Dipteran species such as Bactrocera minax [7], Bactrocera dorsalis [16], Chlorops oryzae [17], Ceratitis capitata [18], Anopheles gambiae [19], Musca domestica [20], Calliphora stygia [21], Glossina morsitans morsitans [22], Episyrphus balteatus and Eupeodes corolla [23]. Insect ORs contain seven conservative transmembrane domains, which are located in the dendrite membrane of olfactory sensory neurons (OSNs) [24, 25]. Structurally, ORs
likely form heteromers of ligand gated ion channels composed of a highly conserved co-receptor (Orco; formerly called OR83b) and an odorant-specific OR, which provides the complex with its ligand specificity [25, 26]. In contrast to general ORs, Orcos are more highly conserved across insect species [25, 27]. ORs are specifically important in the process of insect and host odor recognition, which are broadly tuned to several compounds, including alcohols, ketones and esters [25, 27–30].

Insect GRs, also first identified in the D. melanogaster genome, belong to an ancient superfamily of chemoreceptors which are broadly expressed in gustatory receptor neurons in taste organs, including antennae, legs, mouthparts, wings and ovipositor [31, 32]. GRs and ORs are evolutionarily related within the same insect chemoreceptor superfamily, with ORs specifically being a highly expanded lineage of genes within the larger GR family [32, 33]. GRs, similar to ORs, may be ligand gated ion channels, most of which are divergent and have low sequence identity between insect species [32, 33]. Functional and behavioral studies revealed that GRs respond to many sugars and bitter compounds [34, 35]. Nevertheless, GR21a and GR63a are highly conserved genes associated with carbon dioxide detection and involved in the feeding activities of insects [26, 36, 37]. In addition, some GRs (Gr39a, Gr32a, Gr33a and Gr68a) are also involved in pheromone detection and are necessary for sexual behaviors in Drosophila [26, 38–42].

IRs, another kind of chemosensory genes that are related to ionotropic glutamate receptors (iGluRs), function as ion channels [26]. Insects IRs can be divided into two subfamilies: the conserved “antennal IRs” and the species-specific “divergent IRs” [27, 43]. The “antennal IRs” show orthologs in many different insect species and most of them are primarily or specifically expressed in antennae [7, 25]. Antennal IRs play important roles in thermosensation and odour perception, such as ammonia, amines, polyamines, acids and volatile chemicals impacting courtship [43]. Genomic analyses have indicated that “divergent IRs” have undergone significant expansion in Diptera, and showed unclear orthologous relationships in other insects, forming a number of species-specific clades among insects in phylogenetic trees [25, 26, 43, 44]. “Divergent IRs” are associated with other important biological processes such as taste sensation [43]. Compared to ORs, IRs are more narrowly tuned to a few compounds [25, 45].
SNMPs are the transmembrane domain-containing proteins and belong to a large gene family of CD36 receptors [7, 17, 46]. There are two types of SNMPs: SNMP1 and SNMP2. SNMP1 function as a coreceptor, which is expressed in the dendritic membranes of pheromone-sensitive neurons of insect antennal olfactory sensory neurons, play an important role in pheromone detection [26, 47, 48]. SNMP2 is expressed in the sensilla support cells or other chemosensory organs, which has been proposed to be associated with the clearance of the sensillum lymph [26, 49].

The tephritid stem gall fly, P. utilis Stone (Diptera) originates from Mexico and was introduced into Hawaii, Australia, New Zealand, South Africa and China for the purpose of controlling its host plant, Crofton weed (Ageratina adenophora) [50-52]. P. utilis is a holometabolous insect with four developmental stages: egg (three-five days), larva (25–30 days), pupa (20–25 days) and adult (7–11) [53]. Female adults lay eggs on or near the apical bud of crofton weed. After hatching, larvae tunnel into the stem and feed on stem tissues, ultimately resulting in the formation of galls [52, 54]. Subsequently, larvae and pupae are protected within the gall structures of the host plant until they complete the entire larval and pupal stage [52, 54]. The P. utilis gall is a nutrient sink inhibiting vertical stem growth of the plant, and reduces seed germinability and even causes the death of plants [52]. In Yunnan and many other regions of China, P. utilis has already been successfully used in biological control [50]. Despite its importance in biocontrol, research on P. utilis has mainly focused on the following parameters: biological and ecological characteristics, reproductive behavior and the control effect on the host A. adenophora. Unlike other tephritid insects, the female adults of P. utilis oviposit only on or near the apical bud of A. adenophora and their larvae only fed on this troublesome weed [54, 55], suggesting that some chemosensory genes may play critical role in its searching behavior for specific host plants. Nevertheless, few chemosensory-related genes have been studied in P. utilis.

In this study, we sequenced and analyzed the P. utilis egg, first- to third -instar larvae, pupae, male and female adult transcriptomes using Illumina Hiseq™ 4000 sequencing and identified candidate chemosensory genes (40 OBPs, 29 GRs, 24 ORs, 28 IRs, 6 CSPs, and 6 SNMPs). We then constructed comprehensive and comparative phylogenetic trees to infer the putative functions of each gene, and
identified the expression patterns of the chemosensory via fragments per kilobase per million reads (FPKM) and used quantitative real-time RT-PCR (qPCR) to confirm the expression patterns of OBP genes in each developmental stage. The identification of putative chemosensory-related genes will provide a basis for further study the molecular mechanism of olfactory recognition and also helps advance the use of P. utilis as biological control agents.

Results

Transcriptome assembly

The transcriptomes of eggs (E), first instar larvae (1stL), second instar larvae (2ndL), third instar larvae (3rdL), pupae (P), female (F) and male (M) adults of P. utilis were separately sequenced using Illumina HiSeq 4000 platform. A total of 45.58 million (eggs-1), 45.06 million (eggs-2), 43.36 million (eggs-3), 46.92 million (1stL-1), 46.00 million (1stL-2), 43.61 million (1stL-3), 40.57 million (2ndL-1), 39.90 million (2ndL-2), 49.31 million (2ndL-3), 38.17 million (3rdL-1), 43.07 million (3rdL-2), 39.68 million (3rdL-3), 38.85 million (P-1), 37.63 million (P-2), 45.78 million (P-3), 45.22 million (F-1), 43.24 million (F-2), 33.11 million (F-3), 39.84 million (M-1), 34.35 million (M-2), and 42.90 million (M-3) raw reads were obtained, respectively (Additional file 1: Table S1). After filtering, total of 45.51 million (eggs-1), 45.99 million (eggs-2), 43.30 million (eggs-3), 46.85 million (1stL-1), 45.93 million (1stL-2), 43.55 million (1stL-3), 40.52 million (2ndL-1), 39.85 million (2ndL-2), 49.24 million (2ndL-3), 38.10 million (3rdL-1), 43.00 million (3rdL-2), 39.63 million (3rdL-3), 38.85 million (P-1), 37.56 million (P-2), 45.75 million (P-3), 45.19 million (F-1), 43.23 million (F-2), 33.08 million (F-3), 39.79 million (M-1), 34.33 million (M-2), and 42.84 million (M-3) clean reads were generated, respectively. These transcripts were clustered and assembled into 54,878 unigenes, with a mean length of 952 bp and an N50 of 1690 bp (Additional file 1: Table S1).

Functional annotation of the unigenes in P. utilis

A total of 24,414 (44.49%) unigenes were annotated by searching against four main databases using BLAST with an e-value of 1e-5. Specifically, 23,962 (43.66%) unigenes were matched in the non-redundant (Nr) database, which accounted for the largest match. 18759 (34.18%) were annotated in the Encyclopedia of Genes and Genomes (KEGG) database, 12,459 (22.70%) were annotated by
blasting against the SwissPort database. Additionally, the Clusters of Orthologous Groups (COG) database had the lowest number of annotated unigenes with 12,187 (22.21%) unigenes (Additional file 1: Table S1). Species with the highest proportion of similar genes were Rhagoletis zephyria (22.34%), followed by C. capitata (12.99%), B. dorsalis (10.12%) and Bactrocera cucurbitae (7.59%) (Additional file 2: Figure S1A).

Gene ontology (GO) analysis was used to categorize annotated genes into 57 functional groups in three main categories: ‘biological process’ (19,488), ‘cellular component’ (10,936), and ‘molecular function’ (7,259). Within the biological process class, the subcategories ‘metabolic process’, ‘cellular process’ and ‘single-organism process’ contained the majority of the unigenes. In ‘cellular component’ class, the subcategories ‘cell’ and ‘cell part’ contained the largest numbers of unigenes. ‘Catalytic activity’ and ‘binding’ were the most numerous subcategories in the “molecular function” category (Additional file 2: Figure S1B). For KOG functional classification, approximately 12,187 unigenes were annotated and divided into 25 molecular families (Additional file 2: Figure S1C). Among them, the largest category was the ‘general function prediction only’, followed by ‘signal transduction mechanisms’ and ‘posttranslational modification, protein turnover, chaperones’. ‘Nuclear structure’ and ‘cell motility’ were the smallest groups with only 45 and 30 unigenes, respectively (Additional file 2: Figure S1C). KEGG analysis was used to classified annotated genes into different KEGG Pathway functional categories (Additional file 2: Figure S1D). The most representative pathways included ‘global and overview maps’, ‘translation’ and ‘transport and catabolism’ (Additional file 2: Figure S1D).

**Identification of candidate odorant binding proteins**

In this study, a total of 40 predicted OBP transcripts were identified from *P. utilis* larvae, pupae and adults transcriptomes using the BLASTx program. The sequence identities of these candidate OBPs with other Dipteran insects ranged from 40.81% to 95.59% in the NCBI database (Additional file 3: Table S2). Among them, 32 unigenes have a signal peptide sequence, and the remaining 8 sequences have no signal peptides (Additional file 3: Table S2). Sequence analysis showed that 27 OBPs belonged to the Classical OBP subclass with the typical six conserved cysteine residues and fit the
motif “C1-X14-56- C2-X3-C3- X32-45- C4- X7-42- C5-X8-C6” [56, 57] (Additional file 4: Figure S2). Four (Unigene0017986, Unigene0004435, Unigene0052365 and Unigene0025930) OBPs belonged to the Plus-C subclass, with three conserved cysteines in addition to the six conserved cysteines and a conserved proline after the sixth cysteine [58]. Three (Unigene0006628, Unigene0045087, and Unigene0007971) Minus-C OBPs have four conserved cysteines with C2 and C5 missing [59], whereas two (Unigene0007733 and Unigene0046437) Dimer OBPs have two six conserved cysteine [60, 61] (Additional file 4: Figure S2). The remaining four OBPs have two or three conserved cysteine residues, and thus represent unclassified OBPs.

A Maximum Likelihood phylogenetic tree was constructed to infer the relationships between the 40 OBPs of *P. utilis* and 181 OBPs of seven Dipteran species, including *D. melanogaster, B. minax, B. dorsalis* and *C. capitate* (Fig. 1). In the phylogenetic tree, the *P. utilis* OBP genes clustered in four clades: Plus-C, Minus-C, Dimer and Classical groups. The Plus-C OBP subclass described previously consisting of four of the *P. utilis* OBPs (Unigene0017986, Unigene0004435, Unigene0052365 and Unigene0025930) clusters with the *Drosophila* sequence DmelOBP47b, DmelOBP85a, DmelOBP49a and DmelOBP50e. Two unigenes (Unigene0007733 and Unigene0046437) were clustered into the Dimer class with DmelOBP83ef. Interestingly, Unigene0001375, Unigene0013291, Unigene0003151 and Unigene005748 are unclassified OBPs, but one (Unigene0001375) clustered together with Minus-C OBPs, the remaining three clustered together with Classical OBPs (Fig. 1). Most of the *P. utilis* OBP genes are present in the clade of Classical, and the genes of this family show expansion in *P. utilis* including Unigene0013291, Unigene0003151 and Unigene005748 clusters with CcapOBP4, BdorOBP 19a-2, BminOBP 19a.1 and Dmel_OBP19a, of which Dmel_OBP19a is known to be an antennal binding protein [62]. In addition, another of the classical OBP genes Unigene0044361 was clustered with DmelOBP-lush, which is a protein involved in pheromone binding activities [62, 63].

**Identification of candidate gustatory receptors**

A total of 29 candidate GR transcripts were identified from *P. utilis* transcriptome. Ten GR sequences contained a full-length ORF with 4-8 transmembrane domains (TMDs) (Additional file 3: Table S2). The remaining sequences were incomplete, and seven of them possessed a deduced protein longer than
100 amino acids with 1-2 transmembrane domains (Additional file 3: Table S2). A total of 29 candidate GR transcripts were identified from the *P. utilis* transcriptome. Ten GR sequences contained a full-length ORF with 4-8 transmembrane domains (TMDs) (Additional file 5: Table S6). The remaining sequences were incomplete, and seven of them possessed a deduced protein longer than 100 amino acids with 1-2 transmembrane domains (Additional file 3: Table S2). The Maximum Likelihood phylogenetic tree was constructed to classify the functions of GRs in *P. utilis* transcriptome using the GR genes identified in other dipteran species (Fig. 2). Nine (Unigene0037548, Unigene0037352, Unigene0001281, Unigene0037351, Unigene0037353, Unigene0037354, Unigene0026343, Unigene0040419 and Unigene0033944) GRs of *P. utilis* were clustered with the members of the candidate sugar detection GR subfamily including DmelGR64a-f, DmelGR5a and DmelGR61a [32, 64]. Unigene0046450 and Unigene0052196 were grouped with carbon dioxide GRs (DmelGR21a and DmelGR63a) [7, 37]. Unigene0024893 and Unigene0038509 were classified with a *D. melanogaster* GR (GR43a). *D. melanogaster* GR43a is a fructose receptor in *Drosophila* [65]. Unigene0022838, Unigene0050093 and Unigene0038872 were clustered with DmelGR66a, which is gustatory receptor specific to bitter tasting substrates in *D. melanogaster* [32, 66, 67]. Unigene0017169 and Unigene0026845 were classified with the DmelGR39b, which is a receptor involved in mediating the acceptance or avoidance of special substances [62] (Fig. 2).

**Identification of Candidate odorant receptors**

A total of 24 candidate GRs were identified from *P. utilis* transcriptomes. The sequence identities of these candidate GRs with other Dipteran insects ranged from 27.19% to 96.61% in the NCBI database (Additional file 3: Table S2). Notably, one of the ORs identified in the *P. utilis* transcriptomes shared the highest identity (96.61%), similar to a co-receptor in *R. zephyria* (XP_017470946.1). Among these OR unigenes, six encode full-length proteins with 374 to 472 amino acid residues with 5–7 TMDs. All of the remaining unigenes contained 1–4 TMDs, except that two unigenes had no domains (Additional file 3: Table S2).

A maximum likelihood tree was subsequently constructed using our identified putative OR proteins and the sequences from four other Dipteran species; *D. melanogaster, B. minax, C. stygia* and *B.
dorsalis (Fig. 3). The phylogenetic tree showed that the highly conserved *P. utilis* co-receptor (Orco), Orco (Unigene0052656) was clustered with Orco in *D. melanogaster, B. minax, C. stygia* and *B. dorsalis*. Three *P. utilis* ORs (Unigene0002977, Unigene0053719 and Unigene0027066) were clustered with DmelOR7a from Drosophila, whereas three *P. utilis* ORs (Unigene0034139, Unigene0049177 and Unigene0049173) were clustered with BdorOR7a from *B. dorsalis*. Unigene0049174 and Unigene0003489 were placed on a single branch. Similarly, Unigene0005203 and Unigene0001830 were also placed on a single branch. The remaining unigenes were spread across various branches, where they generally formed small subgroups together with at least one Dipteran orthologous OR in the phylogenetic tree (Fig. 3).

**Identification of candidate ionotropic receptors**

A total of 28 candidate IRs were identified from *P. utilis* transcriptomes by bioinformatics analysis according to the comparison with known insect IRs. The sequences identities of these candidate IRs with other Diptera insects ranged from 53.41 to 96.6% in the NCBI database (Additional file 3: Table S2). Among these IRs, eight encode full-length proteins with 406 to 778 amino acid residues, the remaining unigenes were incomplete, because of the lack of complete 5’ or 3’ terminus. Of these IRs, the transmembrane domains ranged from 0 to 5 (Additional file 3: Table S2). The maximum likelihood tree was subsequently conducted by IQTREE (version 1.6.10) with best-fitting substitution model. To guarantee the reliability of the phylogenetic tree, all of the IRs in our transcriptomes were aligned with IRs from other Dipterans; *D. melanogaster, B. minax, C. stygia* and *C. capitata* (Fig. 4). In the phylogenetic analyses, all *P. utilis* candidate IRs were clustered with other known Dipterans IRs into separate sub-clades and most of them were clustered with presumed “antennal” orthologues (Fig. 4). For example, four (nigene0031176, Unigene0044165, Unigene0031177, Unigene0031178) candidate IRs were located in the clade of the DmelIR76b group, three (Unigene0033408, Unigene0033409, Unigene0033410) candidate IRs were located in the clade of the DmelIR64a group. Unigene0022450 and Unigene0022451 were placed on DmelIR8a group; Unigene0049020 and Unigene0027207 were placed on DmelIR25a group. Unigene0045883 and Unigene0003817 were located in the clade of the DmelIR21a and DmelIR75a group, respectively. In addition, three unigenes were located in the clade
of the non-NMDA iGluRs. The remaining unigenes were spread across various branches with Dipteran orthologous IRs.

Identification of candidate Chemosensory proteins

In all, six putative unigenes encoding CSPs were identified in *P. utilis* (Additional file 3: Table S2). All of them contain four highly conserved cysteine residues and fit the motif “C1-X6-8-C2-X16-21-C3-X2-C4”, which are characteristic of typical insect CSP (Additional file 5: Figure S3). Of these CSPs, four unigenes encode full-length proteins with predicted signal peptide sequences (Additional file 3: Table S2). Phylogenetic analysis was constructed with CSPs from *D. melanogaster, B. minax, M. domestica* and *G. morsitans morsitans*. The resulting phylogenetic tree suggests that Unigene0017876 were clustered with BminCSP3, BdorCSP2 and BdorCSP3, and Unigene0010654 is clustered with DmelCSP1, BminCSP1 and BdorCSP1. Unigene0033940 is classified with CSP4 from *D. melanogaster, B. minax, B. dorsalis and M. domestica*. In addition, Unigene0012976 and Unigene0044055 were placed on a single branch, and Unigene0027955 is also classified in a single branch (Fig. 5).

Identification of candidate sensory neuron membrane proteins

Six unigenes encoding SNMPs were identified in *P. utilis* transcriptome by bioinformatics analysis and all of them were more conserved across other SNMPs variants, with 80.64 to 91.52% amino acid identity. Among them, two unigenes encode full-length proteins with 2 transmembrane domains (TMDs) (Additional file 3: Table S2). A phylogenetic tree was constructed using the sequences from five Dipteran species (Fig. 6). The result demonstrated that Unigene0035241, Unigene0035242 and Unigene0007016 were similar to BminSNMP1b, BminSNMP1a, and BdorSNMP1-1, the remaining 3 unigenes (Unigene0035275, Unigene0012260 and Unigene0035274) were clustered with other insect SNMP2 orthologous (Fig. 6).

Differentially expressed gene (DEG) analysis

Gene expression levels of all candidate chemosensory genes in eggs, 1st-3rd instar larvae, pupae, male and female adults were estimated based on their FRKM values. The overall expression levels of putative GRs, ORs, IRs, SNMPs and CSPs were relatively low compared with that of OBPs (Additional file 6: Table S3). Of the 40 OBPs identified, 17 were mainly expressed in male adults; five were more
highly expressed in female adults; 12 were more highly expressed in pupae and three were highly expressed in 3rd instar larvae. One OBP (Unigene0051738) was mainly expressed in 2nd instar larvae compared to eggs, 1st and 3rd instar larvae, pupae, male and female adults. The remaining two OBPs, the former one was more highly expressed in 1st instar larvae, the latter one was more highly expressed in egg (Fig. 7a). The RPKM values for putative P. utilis GRs were relatively low and ranged from 0 to 4.49. Five, five, four, four, three and two GRs were relatively high expressed in female adults, 3rd instar larvae, eggs, 1st -2nd instar larvae, male adults and pupae, respectively (Fig. 7b). Most candidate ORs were highly expressed in male adults. Five candidate ORs were relatively high expressed in pupae, three ORs were relatively high expressed in 2nd instar larvae, two ORs were relatively high expressed in eggs and one IRs was relatively high expressed in female adults. The remaining candidate ORs exhibited diverse expression patterns (Fig. 7c). For IRs, most candidate IRs were mainly expressed in male adult and three were highly expressed in pupae, whereas three candidate IRs had the highest expression level in eggs. The remaining candidate IRs exhibited diverse expression profiles (Fig. 7d). P. utilis CSPs were mainly expressed in pupae, only one (Unigene0012976) candidate CSP was highly expressed in 2nd instar larvae and Unigene0017876 was highly expressed in female and male adults (Fig. 7e). For SNMPs, three were highly expressed in pupae and the remaining candidate IRs showed diverse expression patterns (Fig. 7f)

**Specific expression patterns of candidate OBP genes in different developmental stages**

To confirm the expression levels of candidate OBP genes, fluorescence quantitative real-time PCR was performed using the seven different developmental stages including eggs, 1st-3rd instar larvae, pupae, male and female adults. Transcript levels of 35 were successfully detected in the seven developmental stages. Ten OBPs were expressed at significantly higher levels in both male and female adults, and the differences between males and females were not significant (Fig. 8). In addition, six OBPs also were more highly expressed in both male and female adults compared to other stages, but were expressed at extremely high significant levels in male adults, as were Unigene0000012, Unigene0004435, Unigene0004872, Unigene0012354, Unigene0020678 and Unigene0053021. The expression levels of five OBPs (Unigene0001375, Unigene0008856,
Unigene0017986 and Unigene0034863) were significantly higher in pupae than those in other six stages. Furthermore, four OBPs (Unigene0022728, Unigene0031193, Unigene0037789 and Unigene0052365) were significantly expressed in female adults, one OBP (Unigene0007733) was significantly expressed in eggs. Unigene0007971 was equally expressed in all three eggs, male and female adults, and Unigene0046437 was equally expressed in both pupae and male adults. Similarity, Unigene0051738 was equally expressed in both 1st and 2nd instar larvae. In general, the relative expression level of most candidate OBPs at different developmental stages sharing the similar trend of the FPKM values (Additional file 6: Table S3).

Discussion

The stem gall fly, P. utilis is an oligophagous insect, has obvious selectivity for host plant and strong adaptability to the toxic weed A. adenophora. Unlike other insects, the female adult of P. utilis oviposits only on or near the apical bud of A. adenophora and its larvae only fed on A. adenophora [52, 54]. In the majority of insect species, chemical cues drive several aspects of their behavior, such as host plant locations, mate finding, and oviposition site selection. Chemosensory proteins were involved in this process [23]. The differences in oviposition and feeding behaviors of the gall fly may be due to the differences in olfactory sensation specificity. Prior to this study, most research on P. utilis was focused on biological and ecological parameters. To better understand P. utilis chemosensory proteins, we first identified candidate chemosensory genes in transcriptomes of P. utilis eggs, 1st–3rd instar larvae, pupae, male and female adults. Then we further analyzed the expression profiles of OBPs in these seven different developmental stages. This study provides valuable information for future studies of chemical communication in the interactions between P. utilis and host plants.

A total of 40 candidate OBPs were identified from P. utilis developmental transcriptomes. This result is similar to the results found in other Dipteran insects, such as C. capitata (45), B. minax (33), E. balteatus (49), E. corollae (44) [18, 23]. Furthermore, the number of OBPs we found in P. utilis is more than those identified in C. oryzae (26), B. dorsalis (31) [68], B. cucurbitae (13) [62] but fewer than those reported in D. melanogaster (52), A. gambiae (83) and M. domestica (87) [18]. The differences
in the number of OBPs may be due to the evolution of different insects and adaptation to a variety of environments [23]. The phylogenetic analysis showed that Unigene0044361 clustered together with DmelOBPs-lush, which was first found to bind with D. melanogaster and other insect pheromones, such as cis-vaccenyl acetate (cVA), short-chain alcohols and phthalates [8, 69]. Unigene0012354 and Unigene0049983 were clustered with the OBP83a/b orthologs form D. melanogaster, B. minax, C. capitate, and B. dorsalis clusters, which were reported to play crucial roles in volatile pheromones detection [4, 7, 70]. The QPCR was used to detect 32 of 40 OBPs expression profiles at different developmental stages based on the FPKM. The results indicated that ten P. utilis OBPs were significantly expressed in male and female adults compared to other developmental stages. Similar results have also been observed in other insects. For example, in D. melanogaster, Obp19a-d was expressed only in adults [71]. In Chilo suppressalis, six OBPs (CsupPBP4 and CsupOBP1, CsupOBP3, CsupOBP8, CsupOBP11, and CsupOBP24) were specifically expressed in the adult stage [72]. In Chrysomya megacephala, two OBP genes were more highly expressed in adults than in larvae [73]. These studies suggested that these OBP genes may be involved in the search for mating partners and oviposition sites [72, 74]. In addition, in our study, there were significant differences in expression levels between male and female adults; six OBPs (Unigene0000012, Unigene0004435, Unigene0004872, Unigene0012354, Unigene0020678, Unigene0053021) were significantly expressed in male adults, suggesting that these OBPs may be related to the perception of sex pheromones, like in moths, Galeruca daurica and Bradysia odoriphaga [59, 4]. In contrast, four OBPs (Unigene0022728, Unigene0031193, Unigene0037789 and Unigene0052365) were significantly expressed in female adults, and similar results were found in G. daurica and B. dorsalis, indicating that these OBPs play important roles in oviposition site selection and mating behavior [4, 59, 74]. There were five (Unigene0001375, Unigene0008856, Unigene0017986 and Unigene0034863) special OBPs that were highly expressed in pupae, which were also consistent with the results obtained in other species. For example, OBP44a, AsteObp1, GdauOBP15, 17, and 25 were found to be abundantly expressed in the pupal stage of B. dorsalis, Anopheles stephensi and G. daurica, respectively [4, 59, 74]. These results suggest that these OBPs may be related to the beginning of the development of chemosensory tissue
during pupation [74, 75]. Furthermore, one OBP (Unigene0051738) in both 1st and 2nd instar larvae was significantly expressed compared to other developmental stages, and similar results were found in D. melanogaster (Obp56a) and Sclerodermus sp. (OBP3), which suggests that these OBPs may have played a basic and conserved role in feeding or identifying general volatiles [71, 76]. Interestingly, one OBP (Unigene0007733) was abundantly expressed exclusively in eggs, which agrees with another study of G. daurica that reported that GdauOBP28 levels peaked in eggs, which suggests its putative roles in egg development [59]. One (Unigene0046437) in both pupae and male adults, and one (Unigene0007971) in eggs, male and female adults were significantly expressed compared to other developmental stages. Research has shown that four OBPs (OBP19c, OBP44a, OBP99a, and OBP99d) from B. dorsalis were expressed from the prepupa to the adult stage, suggesting that they may be related to the distinction of mating partners and egg-laying substrates [74, 76]. In general, our results show that the expression profile of OBPs are extremely complex at different developmental stages, which may be due to their different roles in P. utilis behaviors.

We identified 29 P. utilis GRs, more than those found in the antennae of B. minax, C. oryzae, B. dorsalis, E. balteatus and E. corolla [7, 16, 17, 23], whereas less than those found in C. capitata, D. melanogaster, and M. domestica [18]. In the phylogenetic analysis, eight P. utilis GRs (Unigene0037548, Unigene0037352, Unigene0001281, Unigene0037351, Unigene0037353, Unigene0037354, Unigene0026343 and Unigene0033944) were clustered with the D. melanogaster sugar receptors DmelGR64a-f, DmelGR5a and DmelGR61a, suggesting that they likely play roles in tasting sugar [32, 64]. Two P. utilis GRs formed a clade with CO₂ receptors, suggesting that they may be involved in CO₂ perception [37]. In addition, some GRs from P. utilis were clustered with fructose-detecting Grs and bitters-detecting GRs from Drosophila, indicating that they may perform similar functions [34]. Some studies have showed that the GRs from Helicoverpa armigera may be related to the host plant defense substances and the environments [13, 77]. Similar functional analysis of the candidate P. utilis GRs will be required to confirm their physiological function.

We identified 24 P. utilis ORs, less than those found in the antennae of B. minax, B. dorsalis, E.
balteatus, E. corolla and D. melanogaster [7, 16, 17, 23]. Compared to the antennae, fewer ORs were obtained in P. utilis, and the FPKM value of these genes in eggs, 1st-3rd instar larvae, pupae, male and female adults were relatively low. This may indicate that different samples including antennae will be required to obtain more functional ORs from P. utilis [17]. Our phylogenetic analysis showed that more members of the P. utilis ORs were part of the OR7a superfamily. Previous studies have showed that OR7a is involved in species aggregation and oviposition site-selection in Drosophila [78]. In addition, Unigene0051341 and Unigene0021745 were clustered with BdorOR13a and BdorOR82a, respectively. Furthermore, one OR (Unigene0052656) was clustered with the Orco, which is necessary for membrane targeting of canonical ORs [26]. Recent studies have found that BdorOR13a co-expressed with BdorOrco responded strongly to 1-octen-3-ol, while BdorOR82 co-expressed with BdorOrco responded markedly to geranyl acetate, suggesting that these ORs were related to the perception of plant volatiles that impact on the host-finding behavior of B. dorsalis [16].

Besides GRs and ORs, a third class of chemosensory receptors in insects is the ionotropic receptor (IRs), which play a vital role in the synaptic ligand gated ion channels participating in chemosensation and have multiple functions, such as sensation of tastes, odours and temperature [17, 26]. The numbers of IRs identified in P. utilis were similar to that of B. minax, E. balteatus and B. dorsalis [7, 23, 16] but fewer than those reported in D. melanogaster, C. capitata and M. domestica [18]. It may indicate that the number of IRs varies from species to species and is related to diet or natural habitats [23]. Our phylogenetic analysis showed that most of P. utilis IRs were clustered with “antennal” orthologues from other Dipterans. According to the function of IRs gene in Drosophila, antenna IRs play key roles in odour and thermosensation [43]. For example, IR76b has been implicated in the sensation of polyamine, low-salt and amino acids [32, 43]. IR8a acts as a coreceptor, along with IR64a that responds to acidic molecules [45]. IR8a and IR75a are also tuned to acidic molecules [79]. IR25a and IR21a are necessary to induce responses to cool conditions [80]. IR76b, together with IR25a and IR8a, is considered to have co-receptor function [43]. The IRs orthologs in P. utilis might perform similar functions.

In present study, we identified six SNMPs which are similari to homologues from B. minax, D.
melanogaster and B. dorsalis [68]. Our phylogenetic analysis showed that Unigene0007016, Unigene0035241 and Unigene0035242 were clustered with DmelSNMP1, BminSNMP1a and 1b, suggesting that these proteins may be associated with pheromone reception [7]. We also identified six P. utilis CSP Unigenes. In the phylogenetic tree analysis, three P. utilis CSPs along with CSPs from B. dorsalis and B. minax were grouped in the same clade, suggesting that some CSPs may have similar functions between three species. Further investigations are needed to reveal the functions of P. utilis CSPs.

Conclusions
A total of 133 chemosensory genes were identified from the P. utilis transcriptomes of different developmental stages. As the first step towards understanding gene functions, chemosensory genes were classified according to their conservation, transmembrane domain prediction and phylogenetic analysis. Then, the expression levels of these genes at different development stages were observed in the transcriptomic data and verified OBP gene transcription patterns. This is the first study to obtain and identify candidate genes associated with chemoreception in P. utilis. Our results not only enrich the gene inventory of P. utilis and provide a foundation for further functional studies of the chemosensory system of P. utilis at the molecular level, but also offer creative approaches for the better application of P. utilis as a biological control agent.

Materials And Methods

Host plants and insects.
The host plants, A. adenophora were cultivated in the greenhouse of Yunnan Agricultural University, which were used as the host to rear the laboratory colony of P. utilis as described by Gao et al. (2014) [81]. Samples of eggs, first- to third-instar larvae, pupae, and newly emerged adults (female and male) were collected and rinsed three times in precooled PBS, respectively. Then all samples were immediately frozen in liquid nitrogen and stored at – 80 °C for RNA extraction.

RNA extraction.
Total RNA of 100 eggs, 30 first - instar larvae, 25 second instar larvae, 20 third instar larvae, 20 pupae, 20 female adults and 20 male adults were individually extracted using TRizol reagent
(Invitrogen, Carlsbad, CA, USA) according to the manufacturer’s instructions and genomic DNA was removed using gDNA eraser (TaKaRa, Dalian, China). A NanoDrop® Series and Agilent Bioanalyzer 2100 (Agilent Technologies, CA, USA) were used for detect the quality and concentration of each RNA sample.

**cDNA library construction and Illumina sequencing**

Three micrograms of total RNA per sample was used to construct the cDNA library. cDNA library was generated using Illumina’s sample preparation instructions (Illumina, San Diego, CA). The libraries were then sequenced on the Illumina HiSeq™ 4000 platform (Illumina, San Diego, CA, United States) and performed at the Gene Denovo Biotechnology Co (Guangzhou, China) to obtain paired-end reads (150 bp).

**De novo assembly and functional annotation**

Datasets of clean reads were obtained by removing adaptor reads, ploy-N and low-quality reads, from the raw data using fastp [82]. Clean reads were then de novo assembled using the Trinity program v2.4.0 with default parameters [83]. The raw sequence data has been uploaded to the National Center for Biotechnology Information (NCBI), under the accession number of SRR11300821-SRR11300841.

The annotation of the assembled sequences was conducted by BLASTn and BLASTx searches (E-value <1e-5) against databases. Databases used for annotation include the Nt (non-redundant nucleotide), Nr (non-redundant database), Swiss-Prot, Pfam, KEGG (Kyoto Encyclopedia of Genes and Genomes) and COG/KOG (Clusters of Orthologous Groups).

**Identification of chemosensory genes**

To identify candidate chemosensory genes (ORs, IRs, GRs, SNMPs, OBPs and CSPs), the known amino acid sequences of ORs, IRs, GRs, SNMPs, OBPs and CSPs from other insect species were retrieved from NCBI with the keywords “odorant-binding protein AND insecta”, “odorant receptor AND insecta”, “ionotropic receptor OR AND insecta”, “gustatory receptor AND insecta”, “sensory neuron membrane protein AND insecta”, and “chemosensory proteins AND insecta”). The retrieved sequences were used to blast against assembled *P. utilis* unigenes using tBLASTn with an E-value cut-off of 1.0e−5. All
identified candidate unigenes were manually checked using the BLASTx program against the nr
database. The Open reading frames (ORFs) of all putative chemosensory genes were predicted by
ORF finder tool (https://www.ncbi.nlm.nih.gov/orffinder/). The transmembrane domains of ORs, IRs,
and GRs were predicted by TMHMM2.0 (http://www.cbs.dtu.dk/services/TMHMM/) [84], and the
putative N-terminal signal peptide of OBPs and CSPs were predicted by SignalP 4.0
(http://www.cbs.dtu.dk/services/SignalP/) with default parameters [85].

**Phylogenetic analysis**

Alignments of amino acid sequences were performed by MAFFT (version 7.149) [86]. Phylogenetic
trees of chemosensory genes from *P. utilis* and other Dipteran insects were constructed in IQ-TREE
(version 1.6.2) using the best-fitting substitution-model with Maximum-likelihood [87]. Branch support
was assessed with 1000 bootstrap replicates. Phylogenetic trees were visualized and edited by
FigTree (version 1.4.3). *P. utilis* OBP were compared to sequences from *D. melanogaster, B. dorsalis,
C. capitate* and *B. minax*. *P. utilis* GR were compared to sequences from *B. minax, D. melanogaster, B.
dorsalis,* and *C. stygia*. *P. utilis* IR were compared to sequences from *B. minax, D. melanogaster, C.
capitate* and *C. stygia*. *P. utilis* OR were compared to sequences from *B. minax, D. melanogaster, B.
dorsalis,* and *C. stygia*. CSP sequences were obtained from *B. minax, D. melanogaster, B. dorsalis, M.
domestica* and *G. morsitans morsitans*. SNMP sequences were obtained from *B. minax, B. dorsalis, M.
domestica* and *A. gambiae*. These protein sequences of chemosensory proteins used for constructing
phylogenetic trees are listed in Additional file 7: Table S4.

**Differential gene expression**

The expression levels of these unigenes in egg, first instar larvae, second instar larvae, third instar
larvae, pupae, female and male adults were calculated using the Fragments Reads Per kb per Million
reads (FPKM) method [88]. Differential expression analysis in two different conditions/groups (genes
and samples) was performed by DESeq2 software [89]. Genes with the parameter of false discovery
rate (FDR) below 0.05 and absolute fold change ≥ 2 were considered to be differentially expressed.
Heatmaps of gene expression for different chemosensory genes among egg, first instar larvae,
second instar larvae, third instar larvae, pupae, female and male adults were generated by R
Quantitative Real-Time PCR

We used qRT-PCR to verify the expression of candidate *P. utilis* chemosensory genes at different developmental stages. Total RNA was isolated from 100 eggs, 25 first instar larvae, 20 second instar larvae, 15 third instar larvae, 23 female adults and 23 male adults respectively using Trizol (Invitrogen) according to the manufacturer's instructions. The cDNA was synthesized with input of 1 μg of total RNA using the Prime ScriptRT Reagent Kit with gDNA Eraser to remove gDNA (TaKaRa, China). Gene-specific primers were designed using the NCBI’s profile server (https://www.ncbi.nlm.nih.gov/tools/primer-blast/index.cgi?LINK_LOC=BlastHome ) (Additional file 8: Table S5). The 18S rRNA gene was used as a reference. QPCR was carried out using the SYBR® Premix EX Taq™ II (Tli RNaseH Plus) (No. RR820A, Takara, Dalian, China) on an Applied Biosystems 7500/7500 Fast Real-time PCR System (ABI, Foster City, CA, USA). The reaction conditions were as follows: 95 °C for 30 s, followed by 40 cycles of 95 °C for 5 s, 60 °C for 30 s.

Melting curve analysis was performed from 65 °C to 95 °C to determine the specificity of qPCR primers. Each sample had three biological replicates. The $2^{-\Delta\Delta Ct}$ method was used to analyze gene expression profiles [90]. The comparative analyses for each target gene among various samples were analyzed with a one-way nested analysis of variance (ANOVA), followed by Tukey's honestly significant difference (HSD) test using the software Prism 6.0 (GraphPad Software, CA). All values are presented as the mean ± SE.

Abbreviations

OBP: Odorant-binding protein; GR: Gustatory receptor; OR: Odorant receptor; IR: Ionotropic receptor; ORF: Open reading frame; SNMP: Sensory neuron membrane protein; CSPs: Chemosensory proteins; DEGs: Differentially expressed genes; FPKM: Fragments per kilobase per million fragments; qRT-PCR: quantitative realtime polymerase chain reaction; Orco: Odorant receptor co-receptor; iGluRs: ionotropic glutamate receptors; E: eggs; 1stL: first- instar larvae; 2ndL: second- instar larvae; 3rdL: third- instar larvae; P: pupae; F: female adults; M: male adults; cDNA: Complementary DNA; NCBI: National Center for Biotechnology Information; COG: Clusters of Orthologous Groups; GO: Gene
Ontology; Nt: non-redundant nucleotide; Nr: non-redundant database; KEGG: Kyoto Encyclopedia of Genes and Genomes; TMD: Transmembrane domain

Declarations

Acknowledgements

We thank M.S. students Ping Gao, Ming-Chun He, Xian-Bin Liao and Shi-Yuan Ao (Yunnan Agricultural University) for help collecting insects. We thank Guangzhou Genedenovo Biotechnology Co., Ltd for assisting in sequencing and bioinformatics analysis.

Authors’ contributions

LL carried out the molecular genetic studies, performed the sequence alignment and drafted the manuscript. XG, JZ and GW conceived and designed the experiments. ML, YX, YZ, HG and ZW collected almost all of the samples and participated in some experiments. ZL and MY analyzed the data and designed figures and tables. All authors read and approved the final manuscript.

Funding

This work was funded by the Natural Science Foundation of China (Grant No. 31860521 and 31501706). The funders had no role in study design, data collection and analysis.

Availability of data and materials

The raw reads of the 21 transcriptomes in this study have been stored in the NCBI SRA database, under the accession number of SRR11300841 (E-1: egg-1), SRR11300840 (E-2: egg-2), SRR11300829 (E-2: egg-3), SRR11300827 (1stl-1: first-instar larvae-1), SRR11300826 (1stl-2: first-instar larvae-2), SRR11300825 (1stl-3), SRR11300824 (2ndl-1), SRR11300823 (2ndl-2), SRR11300822 (2ndl-3), SRR11300821 (3rdl-1), SRR11300839 (3rdl-2), SRR11300838 (3rdl-3), SRR11300837 (P-1), SRR11300836 (P-2), SRR11300835 (P-3), SRR11300834 (F-1), SRR11300833 (F-2), SRR11300832 (F-3), SRR11300831 (M-1), SRR11300830 (M-2), SRR11300828 (M-3).

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.
Author details

1 State Key Laboratory for Conservation and Utilization of Bio-Resources in Yunnan, Yunnan Agricultural University, Kunming, China. 2 Key Laboratory of Forest Disaster Warning and Control of Yunnan Province, Southwest Forestry University, Kunming, China. 3 Kunming Branch of Yunnan Provincial Tobacco Company, Kunming 650021, China.

References

[1] Walker WB, Roy A, Anderson P, Schlyter F, Hansson BS, Larsson MC. Transcriptome analysis of gene families involved in chemosensory function in *Spodoptera littoralis* (Lepidoptera: Noctuidae). BMC Genomics. 2019; 20(1): 428. doi: 10.1186/s12864-019-5815-x.

[2] Younas A, Waris MI, Chang XQ, Shaaban M, Abdelnabby H, Ul Qamar MT, Wang MQ. A chemosensory protein MsepCSP5 involved in chemoreception of oriental armyworm *Mythimna separate*. Int J Biol Sci. 2018; 14(14):1935-1949. doi: 10.7150/ijbs.27315.

[3] Song YQ, Sun HZ, Du J. Identification and tissue distribution of chemosensory protein and odorant binding protein genes in *Tropidothorax elegans* Distant (Hemiptera: Lygaeidae). Sci Rep. 2018; 8(1):7803. doi: 10.1038/s41598-018-26137-6.

[4] Tang B, Tai S, Dai W, Zhang C. Expression and functional analysis of two odorant-binding proteins from *Bradysia odoriphaga* (Diptera: Sciaridae). J Agric Food Chem. 2019; 67(13):3565-3574. doi: 10.1021/acs.jafc.9b00568.

[5] Sun L, Wang Q, Wang Q, Dong K, Xiao Y, Zhang YJ. Identification and characterization of odorant binding proteins in the forelegs of *Adelphocoris lineolatus* (Goeze). Front Physiol. 2017; 8:735. doi: 10.3389/fphys.2017.00735.

[6] Pelosi P, lovinella I, Zhu J, Wang G, Dani FR. Beyond chemoreception: diverse tasks of soluble olfactory proteins in insects. Biol. Rev Camb Philos Soc. 2018; 93(1):184-200. doi: 10.1111/brv.12339.

[7] Wu Z, Kang C, Qu M, Chen J, Chen M, Bin S, Lin J. Candidates for chemosensory genes identified in the Chinese citrus fly, *Bactrocera minax*, through a transcriptomic analysis. BMC Genomics. 2019; 20(1):646. doi: 10.1186/s12864-019-6022-5.

[8] Chen XF, Xu L, Zhang YX, Wei D, Wang JJ, Jiang HB. Genome-wide identification and expression
profiling of odorant-binding proteins in the oriental fruit fly, *Bactrocera dorsalis*. Comp Biochem Physiol Part D Genomics Proteomics. 2019; 31:100605. doi: 10.1016/j.cbd.2019.100605.

[9] Vogt RG, Riddiford LM. Pheromone binding and inactivation by moth antennae. Nature. 1981;293(5828):161-3. doi: 10.1038/293161a0.

[10] Galindo K, Smith DP. A large family of divergent *Drosophila* odorant-binding proteins expressed in gustatory and olfactory sensilla. Genetics. 2001; 159(3):1059-72.

[11] Jeong YT, Shim J, Oh SR, Yoon HI, Kim CH, Moon SJ, Montell C. An odorant-binding protein required for suppression of sweet taste by bitter chemicals. Neuron. 2013; 79(4):725-37. doi: 10.1016/j.neuron.2013.06.025.

[12] Waris MI, Younas A, Adeel MM, Duan SG, Quershi SR, Kaleem Ullah RM, Wang MQ. The role of chemosensory protein 10 in the detection of behaviorally active compounds in brown planthopper, *Nilaparvata lugens*. Insect Sci. 2018 Dec 28. doi: 10.1111/1744-7917.12659.

[13] Tian Z, Sun L, Li Y, Quan L, Zhang H, Yan W, Yue Q, Qiu G. Antennal transcriptome analysis of the chemosensory gene families in *Carposina sasakii* (Lepidoptera: Carposinidae). BMC Genomics. 2018;19(1):544. doi: 10.1186/s12864-018-4900-x.

[14] Clyne PJ, Warr CG, Freeman MR, Lessing D, Kim J, Carlson JR. A novel family of divergent seven-transmembrane proteins: candidate odorant receptors in *Drosophila*. Neuron. 1999;22(2):327-38. doi: 10.1016/s0896-6273(00)81093-4.

[15] Benton R, Sachse S, Michnick SW, Vosshall LB. Atypical membrane topology and heteromeric function of *Drosophila* odorant receptors in vivo. PLoS Biol. 2006;4(2):e20. doi: 10.1371/journal.pbio.0040020.

[16] Miyazaki H, Otake J, Mitsuno H, Ozaki K, Kanzaki R, Chui-Ting Chieng A, et al. Functional characterization of olfactory receptors in the Oriental fruit fly *Bactrocera dorsalis* that respond to plant volatiles. Insect Biochem Mol Biol. 2018;101:32-46. doi: 10.1016/j.ibmb.2018.07.002

[17] Qiu L, Tao S, He H, Ding W, Li Y. Transcriptomics reveal the molecular underpinnings of chemosensory proteins in *Chlorops oryzae*. BMC Genomics. 2018;19(1):890. doi: 10.1186/s12864-018-5315-4.
[18] Papanicolaou A, Schetelig MF, Arensburger P, Atkinson PW, Benoit JB, Bourtzis K, Castañera P, et al. The whole genome sequence of the Mediterranean fruit fly, Ceratitis capitata (Wiedemann), reveals insights into the biology and adaptive evolution of a highly invasive pest species. Genome Biol. 2016;17(1):192. doi: 10.1186/s13059-016-1049-2.

[19] Rinker DC, Zhou X, Pitts RJ, Rokas A, Zwiebel LJ. Antennal transcriptome profiles of anopheline mosquitoes reveal human host olfactory specialization in Anopheles gambiae. BMC Genomics. 2013;14:749. doi: 10.1186/1471-2164-14-749.

[20] Scott JG, Warren WC, Beukeboom LW, Bopp D, Clark AG, Giers SD, et al. Genome of the house fly, Musca domestica L., a global vector of diseases with adaptations to a septic environment. Genome Biol. 2014;15(10):466. doi: 10.1186/s13059-014-0466-3.

[21] Leitch O, Papanicolaou A, Lennard C, Kirkbride KP, Anderson A. Chemosensory genes identified in the antennal transcriptome of the blowfly Calliphora stygia. BMC Genomics. 2015;16(1):255. doi: 10.1186/s12864-015-1466-8.

[22] Obiero GF, Mireji PO, Nyanjom SR, Christoffels A, Robertson HM, Masiga DK. Odorant and gustatory receptors in the tsetse Fly Glossina morsitans morsitans. PLoS Negl Trop Dis. 2014;8(4):e2663. doi: 10.1371/journal.pntd.0002663.

[23] Wang B, Liu Y, Wang GR. Chemosensory genes in the antennal transcriptome of two syrphid species, Episyrphus balteatus and Eupeodes corolla (Diptera: Syrphidae). BMC Genomics. 2017;18(1):586.doi: 10.1186/s12864-017-3939-4.

[24] Du LX, Liu Y, Zhang J, Gao XW, Wang B, Wang GR. Identification and characterization of chemosensory genes in the antennal transcriptome of Spodoptera exigua. Comp Biochem Physiol Part D Genomics Proteomics. 2018;27:54-65. doi: 10.1016/j.cbd.2018.05.001.

[25] Gomez-Diaz C, Martin F, Garcia-Fernandez JM, Alcorta E. The Two main olfactory receptor families in Drosophila, ORs and IRs: A Comparative Approach. Front Cell Neurosci. 2018;12:253. doi: 10.3389/fncel.2018.00253.

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

[27] Liu JB, Wu H, Yi JQ, Song ZW, Li DS, Zhang GR. Transcriptome characterization and gene expression analysis related to chemoreception in *Trichogramma chilonis*, an egg parasitoid. Gene. 2018;678:288-301. doi: 10.1016/j.gene.2018.07.065.

[28] Xu JW, Zhu XY, Chao QJ, Zhang YJ, Yang YX, Wang RR, et al. Chemosensory gene families in the oligophagous pear pest *Cacopsylla chinensis* (Hemiptera: Psyllidae). Insects. 2019;10(6):175. doi: 10.3390/insects10060175.

[29] Hallem EA, Carlson J R. Coding of odors by a receptor repertoire. Cell. 2006;125(1):143-160. doi: 10.1016/j.cell.2006.01.050.

[30] Hallem EA, Ho MG, Carlson JR. The molecular basis of odor coding in the *Drosophila* antenna. Cell. 2004;117(7):965-79. doi: 10.1016/j.cell.2004.05.012.

[31] Clyne PJ, Warr CG, Carlson JR. Candidate taste receptors in *Drosophila*. Science. 2000;287(5459):1830-4. doi: 10.1126/science.287.5459.1830.

[32] Scott K. Gustatory Processing in *Drosophila melanogaster*. Annu Rev Entomol. 2018;63:15-30. doi: 10.1146/annurev-ento-020117-043331.

[33] Robertson HM, Warr CG, Carlson JR. Molecular evolution of the insect chemoreceptor gene superfamily in *Drosophila melanogaster*. Proc Natl Acad Sci U S A. 2003;100 (Suppl 2):14537-42. doi: 10.1073/pnas.2335847100.

[34] Weiss LA, Dahanukar A, Kwon JY, Banerjee D, Carlson JR. The molecular and cellular basis of bitter taste in *Drosophila*. Neuron. 2011;69(2):258-272.doi: 10.1016/j.neuron.2011.01.001.

[35] Hiroi M, Meunier N, Marion-Poll F, Tanimura T. Two antagonistic gustatory receptor neurons responding to sweet-salty and bitter taste in *Drosophila*. Neurobiol. 2004;61(3):333-42. doi: 10.1002/neu.20063.

[36] Kwon JY, Dahanukar A, Weiss LA, Carlson JR. The molecular basis of CO$_2$ reception in *Drosophila*. Proc Natl Acad Sci USA. 2007;104(9):3574-3578. doi:10.1073/pnas.0700079104.

[37] Jones WD, Cayirlioglu P, Kadow IG, Vosshall LB. Two chemosensory receptors together mediate carbon dioxide detection in *Drosophila*. Nature. 2007;445(7123):86-90. doi:10.1038/nature05466.
[38] Kohl J, Huoviala P, Jeferis GS. Pheromone processing in Drosophila. Curr Opin Neurobiol. 2015;34:149-157. doi:10.1016/j.conb.2015.06.009.

[39] Joseph RM, Carlson JR. Drosophila chemoreceptors: a molecular interface between the chemical world and the brain. Trends Genet. 2015;31(12):683-695. doi:10.1016/j.tig.2015.09.005.

[40] Watanabe K, Toba G, Koganezawa M, Yamamoto D. Gr39a, a highly diversified gustatory receptor in Drosophila, has a role in sexual behavior. Behav Genet. 2011;41(5):746-753. doi:10.1007/s10519-011-9461-6.

[41] Hu Y, Han Y, Shao Y, Wang X, Ma Y, Ling E, Xue L. Gr33a modulates Drosophila male courtship preference. Sci Rep. 2015;5:7777. doi: 10.1038/srep07777.

[42] Moon SJ, Lee Y, Jiao Y, Montell C. A Drosophila gustatory receptor essential for aversive taste and inhibiting maleto-male courtship. Curr Biol. 2009;19(19):1623-1627. doi:10.1016/j.cub.2009.07.061.

[43] Rimal S, Lee Y. The multidimensional ionotropic receptors of Drosophila melanogaster. Insect Mol Biol. 2018; 27 (1), 1-7. doi: 10.1111/imb.12347.

[44] Benton R, Vannice KS, Gomez-Diaz C, Vosshall LB. Variant ionotropic glutamate receptors as chemosensory receptors in Drosophila. Cell. 2009;136(1):149-62. doi: 10.1016/j.cell.2008.12.001.

[45] Silbering AF, Rytz R, Grosjean Y, Abuin L, Ramdya P, Jefferis GS, Benton R. Complementary function and integrated wiring of the evolutionarily distinct Drosophila olfactory subsystems. J Neurosci. 2011;31(38):13357-75. doi: 10.1523/JNEUROSCI.2360-11.2011.

[46] Wei HS, Li KB, Zhang S, Cao YZ, Yin J. Identification of candidate chemosensory genes by transcriptome analysis in Loxostege sticticalis Linnaeus. PLoS One. 2017;12(4):e0174036. doi: 10.1371/journal.pone.0174036.

[47] Rogers ME, Sun M, Lerner MR, Vogt RG. Snmp-1, a novel membrane protein of olfactory neurons of the silk moth Antheraea polyphemus with homology to the CD36 family of membrane proteins. J Biol Chem. 1997;272(23):14792-14799. doi: 10.1074/jbc.272.23.14792.

[48] Rogers ME, Steinbrech RA, Vogt RG. Expression of SNMP-1 in olfactory neurons and sensilla of male and female antennae of the silkmoth Antheraea polyphemus. Cell Tissue Res. 2001;303(3):433-446. doi:10.1007/s004410000305.
[49] Forstner M, Gohl T, Gondesen I, Raming K, Breer H, Krieger J. Differential expression of SNMP-1 and SNMP-2 proteins in pheromone-sensitive hairs of moths. Chem Senses. 2005;33(3):291-9. doi:10.1093/chemse/bjm087.

[50] Sun XY, Lu ZH, Sang WG. Review on studies of Eupatorium adenophoruman important invasive species in China. Journal of Forestry Research. 2004; 15(4):319-322.

[51] Bess HA, Haramoto FH. Biological control of pamakani, Eupatorium adenophorum, in Hawaii by a tephritid gall fly, Procecidochares utilis. 3. Status of the weed, fly and parasites of the fly in 1966-71 versus 1950-57. Proceedings of the Hawaiian Entomological Society. 1972;40(2):244-249.

[52] Buccellato L. An insect-plant pathogen interaction between two biocontrol agents, the stem-gall fly and leaf-spot pathogen released against crofton weed, Ageratina adenophora. University of the Witwatersrand. 2012.

[53] Zhang ZY, Wei Yi, He DY. Biology of a gall fly, procecidochares utilis [dip.: tephritidae] and its impact on crofton weed, Eupatorium adenophorum. Chinese journal of biological control.1988; 4(1): 10-13.

[54] Bennett PH, VanStaden J. Gall formation in crofton weed, Eupatorium adenophorum Spreng. (syn. Ageratina adenophora), by the Eupatorium gall fly Procecidochares utilis Stone (Diptera: Trypetidae). Australian Journal of Botany. 1986;34, 473-480.

[55] Bush GL, Taylor SC. The Cytogenetics of Procecidochares. I. The Mitotic and Polytene Chromosomes of the Pamakani Fly, P. utilis Stone (Tephritidae-Diptera). Caryologia –Firenze. 2014; 22(4):311-321.

[56] Xu YL, He P, Zhang L, Fang SQ, Dong SL, Zhang YJ, Li F. Large-scale identification of odorant-binding proteins and chemosensory proteins from expressed sequence tags in insects. BMC Genomics. 2009;10:632. doi: 10.1186/1471-2164-10-632.

[57] Cui HH, Gu SH, Zhu XQ, Wei Y, Liu HW, Khalid HD, Guo YY, Zhang YJ. Odorant-binding and chemosensory proteins identified in the antennal transcriptome of Adelphocoris suturalis Jakovlev. Comp Biochem Physiol Part D Genomics Proteomics. 2017;24:139-145. doi: 10.1016/j.cbd.2016.03.001.
[58] Zhou JJ, Huang W, Zhang GA, Pickett JA, Field LM. “Plus-C” odorant-binding protein genes in two *Drosophila* species and the malaria mosquito *Anopheles gambiae*. Gene. 2004;327(1):117-29. doi: 10.1016/j.gene.2003.11.007.

[59] Li L, Zhou YT, Tan Y, Zhou XR, Pang BP. Identification of odorant-binding protein genes in *Galeruca daurica* (Coleoptera: Chrysomelidae) and analysis of their expression profiles. Bull Entomol Res. 2017;107(4):550-561. doi: 10.1017/S0007485317000402.

[60] Brito NF, Moreira MF, Melo AC. A look inside odorant-binding proteins in insect chemoreception. J Insect Physiol. 2016; 95:51-65. doi: 10.1016/j.jinsphys.2016.09.008.

[61] Vieira FG, Rozas J. Comparative genomics of the odorant-binding and chemosensory protein gene families across the Arthropoda: origin and evolutionary history of the chemosensory system. Genome Biol Evol. 2011;3:476-90. doi: 10.1093/gbe/evr033.

[62] Elfekih S, Chen CY, Hsu JC, Belcaid M, Haymer D. Identification and preliminary characterization of chemosensory perception-associated proteins in the melon fly *Bactrocera cucurbitae* using RNA-seq. Sci Rep. 2016; 6: 19112. doi: 10.1038/srep19112.

[63] Xu P, Atkinson R, Jones DN, Smith DP, *Drosophila* OBP LUSH is required for activity of pheromone-sensitive neurons. Neuron. 2005; 45(2):193-200. doi: 10.1016/j.neuron.2004.12.031.

[64] Fujii S, Yavuz A, Slone J, Jagge C, Song X, Amrein H. Drosophila sugar receptors in sweet taste perception, olfaction, and internal nutrient sensing. Curr Biol. 2015;25(5):621–27. doi: 10.1016/j.cub.2014.12.058.

[65] Miyamoto T, Slone J, Song X, Amrein H. A fructose receptor functions as a nutrient sensor in the *Drosophila* brain. Cell. 2012; 151(5):1113-25. doi: 10.1016/j.cell.2012.10.024.

[66] Lee Y, Moon SJ, Montell C. Multiple gustatory receptors required for the caffeine response in *Drosophila*. Proc Natl Acad Sci USA. 2009;106(11):4495-500. doi: 10.1073/pnas.0811744106.

[67] Moon SJ, Köttgen M, Jiao Y, Xu H, Montell C. A taste receptor required for the caffeine response in vivo. Curr Biol. 2006;16(18):1812-7. doi: 10.1016/j.cub.2006.07.024.

[68] Wu Z, Zhang H, Wang Z, Bin S, He H, Lin J. Discovery of chemosensory genes in the oriental Fruit Fly, *Bactrocera dorsalis*. PLoS One. 2015;10(6):e0129794.. doi:10.1371/journal.pone.0129794
[69] Kim MS, Repp A, Smith DP. LUSH odorant-binding protein mediates chemosensory responses to alcohols in *Drosophila melanogaster*. Genetics. 1998; 150(2):711-721. doi: 10.1039/a908148h.

[70] Siciliano P, He XL, Woodcock C, Pickett JA, Field LM, Birkett MA, et al. Identification of pheromone components and their binding affinity to the odorant binding protein CcapOBP83a-2 of the Mediterranean fruit fly, *Ceratitis capitata*. Insect Biochem Mol Biol. 2014;48(100):51-62. doi: 10.1016/j.ibmb.2014.02.005.

[71] Zhou S, Stone EA, Mackay TF, Anholt RR. Plasticity of the chemoreceptor repertoire in *Drosophila melanogaster*. PLoS Genet. 2009;5(10):e1000681. doi: 10.1371/journal.pgen.1000681.

[72] Yang K, Liu Y, Niu DJ, Wei D, Li F, Wang GR, Dong SL. Identification of novel odorant binding protein genes and functional characterization of OBP8 in *Chilo suppressalis* (Walker). Gene. 2016;591(2), 425-432. doi.org/10.1016/j.gene.2016.06.052.

[73] Wang XY, Xiong M, Lei CL, Zhu F. The developmental transcriptome of the synanthropic fly *Chrysomya megacephala* and insights into olfactory proteins. BMC Genomics. 2015; 16(1), 20. doi: 10.1186/s12864-014-1200-y.

[74] Zhang J, Luo D, Wu P, Li H, Zhang H, Zheng W. Identification and expression profiles of novel odorant binding proteins and functional analysis of OBP99a in *Bactrocera dorsalis*. Arch Insect Biochem Physiol. 2018; 98(1):e21452. doi.org/10.1002/arch.21452.

[75] Sengul MS, Tu Z. Identification and characterization of odorant-binding protein 1 gene from the Asian malaria mosquito, *Anopheles stephensi*. Insect Molecular Biology. 2010;19(1), 49–60. doi: 10.1111/j.1365-2583.2009.00929.x.

[76] Zhou J, Zhang N, Wang P, Zhang S, Li D, Liu K, Ai H. Identification of host-plant volatiles and characterization of two novel general odorant-binding proteins from the legume pod borer, *Maruca vitrata* Fabricius (Lepidoptera: Crambidae). PLoS One. 2015;10(10), e0141208. doi: 10.1371/journal.pone.0141208.

[77] Xu W, Alexie P, Zhang HJ, et al. Expansion of a bitter taste receptor family in a polyphagous insect herbivore. Sci Rep. 2016;6(1):23666. doi: 10.1038/srep23666.

[78] Lin CC, Prokop-Prigge KA, Preti G, Potter CJ. Food odors trigger *Drosophila* males to deposit a
pheromone that guides aggregation and female oviposition decisions. elife. 2015;4:e08688. doi: 10.7554/elife.08688.

[79] Prieto-Godino LL, Rytz R, Bargeton B, Abuin L, Arguello JR, Dal Peraro M, et al. Olfactory receptor pseudopseudogenes. Nature. 2016;539(7627): 93–97. doi: 10.1038/nature19824.

[80] Ni L, Klein M, Svec KV, Budelli G, Chang EC, Ferrer AJ, et al. The Ionotropic Receptors IR21a and IR25a mediate cool sensing in Drosophila. eLife. 2016;5: e13254. doi: 10.7554/eLife.13254.

[81] Gao X, Zhu JY, Ma S, Zhang Z, Xiao C, Li Q, et al. Transcriptome profiling of the crofton weed gall fly Procecidochares utilis. Genet Mol Res.2014; 13(2):2857-2864. doi: 10.4238/2014.March.19.1.

[82] Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. BioRxiv. 2018;274100. doi:org/10.1101/274100.

[83] Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genom. Nat Biotechnol. 2011;29(7): 644. doi:10.1038/nbt.1883.

[84] Krogh A, Larsson B, von Heijne G, Sonnhammer ELL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol. 2001; 305(3):567–580. doi: 10.1006/jmbi.2000.4315.

[85] Petersen TN, Brunak S, von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods. 2011;8(10):785–6. doi: 10.1038/nmeth.1701.

[86] Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 2002;30(14):3059-66. doi: 10.1093/nar/gkf436.

[87] Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 2015;32(1):268-74. doi: 10.1093/molbev/msu300.

[88] Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods. 2008;5(7):621-8.

[89] Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNAseq
data with DESeq2. Genome Biol. 2014;15(12):550. doi: 10.1186/s13059-014-0550-8.

[90] Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-DeltaDeltaC(T)) method. Methods. 2001;25(4):402-8. doi: 10.1006/meth.

Figures

A maximum likelihood phylogenetic tree of candidate P. utilis OBPs (red) and other Dipteran OBPs. D. melanogaster (Dmel, black), B. minax (Bmin, purple), B. dorsalis (Bdor, blue) and C. capitate (Ccap, deep yellow). The Classic OBPs clade, Plus-C OBPs clade, Minus-C OBPs and Dimer OBPs clade are shown. Bootstrap values after 1000 replications.
Figure 2

A maximum likelihood phylogenetic tree of candidate P. utilis GRs (red) and other Dipteran GRs. D. melanogaster (Dmel, black), B. minax (Bmin, purple), B. dorsalis (Bdor, blue) and C. stygia (csty, deep yellow). The carbon dioxide GRs clade, sugar GRs clade and bitter GRs clade are shown. Bootstrap values after 1000 replications.
Figure 3

A maximum likelihood phylogenetic tree of candidate P. utilis ORs (red) and other Dipteran ORs. D. melanogaster (Dmel, black), B. minax (Bmin, blue), B. dorsalis (Bdor, green) and C. stygia (Csty, purple). The ORco clade are shown. Bootstrap values after 1000 replications.
Figure 4

A maximum likelihood phylogenetic tree of candidate P. utilis IRs (red) and other Diptera IRs. D. melanogaster (Dmel, black), B. minax (Bmin, purple), C. capitate (Ccap, blue) and C. stygia (Csty, deep yellow). The IR25a/IR8a clade, IR76b clade, IR64a clade, IR75d clade and some antennal-associated orthologue clade are shown. Bootstrap values after 1000 replications.
A maximum likelihood phylogenetic tree of candidate P. utilis CSPs (red) and other Dipteran CSPs. D. melanogaster (Dmel, black), B. minax (Bmin, purple), B. dorsalis (Bdor, blue), C. stygia (Csty, green), M. domestica (Mdom, orange) and Anopheles funestus (Afun, gray).

Bootstrap values after 1000 replications.
Figure 6

A maximum likelihood phylogenetic tree of candidate P. utilis SNMPs (red) and other Dipteran SNMPs. D. melanogaster (Dmel, black), B. minax (Bmin, purple), B. dorsalis (Bdor, blue), M. domestica (Mdom, deep yellow), A. gambiae (Agam, green) and A. aegypti (Aaeg, deep yellow). Bootstrap values after 1000 replications.
Figure 7

Expression profiles of chemosensory genes in P. utilis. a: OBPs; b: GRs; c: ORs; d: IRs; e: CSPs; f: SNMPs. E: eggs; 1stL: first-instar larvae; 2ndL: second-instar larvae; 3rdL: third-instar larvae; P: pupae; F: female adults; M: male adults.
Figure 8

Expression profiles of OBPs at different developmental stages, using RT-qPCR. E: eggs; 1stL: first-instar larvae; 2ndL: second-instar larvae; 3rdL: third-instar larvae; P: pupae; F: female adults; M: male adults.

Supplementary Files
This is a list of supplementary files associated with this preprint. Click to download.
Additional file 5 Figure S3.pdf
Additional file 2 Figure S1.pdf
Additional file 8 Table S5.xlsx
Additional file 3 Table S2.xlsx
Additional file 7 Table S4.xlsx
Additional file 6 Table S3.xlsx
Additional file 1 Table S1.xlsx