SCIENTIFIC REP♀RTS

OPEN

Antennal transcriptome and expression analyses of olfactory genes in the sweetpotato weevil *Cylas formicarius*

Shu-Ying Bin, Meng-Qiu Qu, Xin-Hua Pu, Zhong-Zhen Wu & Jin-Tian Lin

The sweetpotato weevil, *Cylas formicarius* (Fabricius), is a serious pest of sweetpotato. Olfaction-based approaches, such as use of synthetic sex pheromones to monitor populations and the bait-and-kill method to eliminate males, have been applied successfully for population management of *C. formicarius*. However, the molecular basis of olfaction in *C. formicarius* remains unknown. In this study, we produced antennal transcriptomes from males and females of *C. formicarius* using high-throughput sequencing to identify gene families associated with odorant detection. A total of 54 odorant receptors (ORs), 11 gustatory receptors (GRs), 15 ionotropic receptors (IRs), 3 sensory neuron membrane proteins (SNMPs), 33 odorant binding proteins (OBPs), and 12 chemosensory proteins (CSPs) were identified. Tissue-specific expression patterns revealed that all 54 ORs and 11 antennal IRs, one SNMP, and three OBPs were primarily expressed in antennae, suggesting their putative roles in olfaction. Sex-specific expression patterns of these antenna-predominant genes suggest that they have potential functions in sexual behaviors. This study provides a framework for understanding olfaction in coleopterans as well as future strategies for controlling the sweetpotato weevil pest.

Olfaction plays an essential role in the life cycle of insects that use a wide range of environmental chemical cues to locate and evaluate food, mates, and egg-laying sites as well as to avoid predators and other dangers. Olfaction is therefore an important research field in insect biology[1]. Insect odor reception and signal transduction occurs in the dendritic membrane of olfactory sensory neurons (OSNs) in the antennae[2, 3]. The key molecular components are a diverse array of odorant receptors (ORs), ionotropic receptors (IRs), and odorant-binding proteins (OBPs)[1]. After entering the sensillum lymph through cuticular pores, odorant molecules are recognized, bound, and transported by OBPs or potential chemosensory proteins (CSPs) across the lymph to ORs in the dendritic membrane of OSNs. The OSNs are then activated and generate an electrical signal that is processed and transmitted to higher-order neural centers[1, 4, 5]. Sensory neuron membrane proteins (SNMPs)[6–8] and ionotropic receptors (IRs)[9–14] have also been reported in OSNs with acquired olfactory functions, as well as particular gustatory receptors (GRs) responding to carbon dioxide[15, 16].

Coleoptera (beetles and weevils) is the largest and most diverse order of insects on earth, representing almost 40 percent of all described insect species (https://www.britannica.com/animal/beetle). The order contains a large number of important pests of agriculture, forestry and stored products worldwide. Sex and aggregation pheromones are critical in communication among individuals within a coleopteran species for mating and locating host plants[17, 18].

Olfactory-related genes have been identified from several Coleopteran species, including *Tribolium castaneum*[19, 20], *Megacyllene caryae*[21], *Ips typographus*[22], *Dendroctonus ponderosae*[22], *Agrilus planipennis*[23], *Batocera horsfieldi*[24], *Anomala corpulenta*[25, 26], *Dendroctonus valens*[27], *Tenebrio molitor*[28], *Colaphellus bowringi*[29], *Ambrostoma quadriimpressum*[30], *Rhynchophorus ferrugineus*[31], *Phyllotreta striolata*[32], *Anoplophora glabripennis*[33], *Brontispa longissima*[34]. To date, identifying functional olfactory molecules in Coleoptera is limited in a single species. In *M. caryae*, three ORs (McarOR3, McarOR20, and McarOR5) were functionally characterized that

Institute for Management of Invasive Alien Species, 314 Yingdong teaching building, Zhongkai University of Agriculture and Engineering, Guangzhou, 510225, P.R. China. Correspondence and requests for materials should be addressed to Z.-Z.W. (email: zhongzhen_wu@163.com) or J.-T.L. (email: linjtian@163.com).

they were tuned respectively to three aggregation pheromone components ((S)-2-methyl-1-butanol, (2S, 3R)-2, 3-hexanediol and 2-phenylethanol)[21].

The sweetpotato weevil, *Cylas formicarius* (Fabricius) (Coleoptera: Brentidae), is a serious pest of sweetpotato (*Ipomea batatans*)[35]. *C. formicarius* damages sweet potatoes both in the field and in storage[36] and it is a quarantine pest. Owing to the cryptic feeding habits of the larvae and the nocturnal activities of the adults, *C. formicarius* is difficult to control using conventional chemical insecticides. Due to the life-history characteristics of *C. formicarius*, chemical control has typically been achieved with residual insecticides such as spinosad and azadirachtin. However, this management approach has led to control failures due to development of insecticide resistance[37]. Olfaction-based approaches, using synthetic sex pheromones and host volatiles to interfere with the pests' ability to find suitable mates and hosts, have been used successfully in "push-pull" control strategies[38]. The male *C. formicarius* is very sensitive to the pheromone active component (Z)-3-dodecen-1-ol (E)-2-butenoate, released by virgin females[39, 40]. Pheromone-baited traps used for population monitoring and mass trapping can provide effective control of this pest[41, 42]. However, the molecular mechanisms involved with *C. formicarius* finding mates or host plants are still unknown. Hence, a detailed knowledge of olfaction in insects is imperative.

The goal of the present study was to identify the genes involved in olfaction from the male and female antennae of *C. formicarius* using high-throughput sequencing. The expression profiles in different tissues were studied using semi-quantitative RT-PCR and real-time quantitative-PCR and their putative olfactory functions are proposed. Evolutionary relationships with other Coleoptera olfaction genes are discussed.

Results

Illumina sequencing. A total of 51,183,400 and 48,433,914 raw reads were obtained from male and female antennae cDNA libraries of *C. formicarius*, respectively. Trimming adaptor sequences, eliminating low quality reads and contaminating sequences produced 51,025,928 and 48,238,192 clean reads from male and female antennae respectively. After a combined assembly, the two datasets resulted in 66,531 unigenes with a mean length of 1,384 bp, an N50 of 2,924 bp, and an N90 of 506 bp (see Supplementary Table S1).

Functional annotation of assembled unigenes. Annotation was conducted by BLASTx and BLASTn program with the E-value cut-off of 10^{-5}, 31,532 (47.39%) unigenes were annotated by at least one of the databases: 30,003 (45.10%) unigenes were annotated by the NCBI-Nr database, 16,661 (25.04%) unigenes by the NCBI-Nt database, 22,720 (34.15%) by SwissPort, 9,010 (13.54%) by GO, 11,650 (17.51%) by the COG database, 22,812 (34.29%) by KEGG (Fig. 1A). BLASTx homology searches in the NCBI-Nr database showed that *C. formicarius* antennal transcriptomes had a best blast match to coleopteran sequences, primarily the mountain pine beetle *D. ponderosae*, (38.52%) and the red flour beetle *T. castaneum* (37.45%) (Fig. 1B). With the GO classification, all unigenes in *C. formicarius* antennal transcriptomes were classified into 3 functional categories: molecular function, biological process, and cellular component (Fig. 1C). In molecular function, the most abundant transcripts in the antennae were linked to binding and catalytic activity. In biological process, the most represented biological processes were cellular, metabolic, and single-organism processes. In the cellular component terms, cell, cell part, and membrane constituted the most abundant categories. These GO assignments are in accordance with those reported previously for dipteran[43–45], lepidopteran[46–48], and other coleopteran antennal transcriptomes[22, 27, 33].

Gene families associated with odorant detection. Gene families involved in odorant detection, including ORs (54 transcripts), GRs (11 transcripts), IRs (15 transcripts), SNMPs (3 transcripts), OBPs (26 transcripts) and CSPs (12 transcripts) were identified in male and female *C. formicarius* antennal transcriptomes. Information of candidate ORs, GRs, IRs, SNMPs, OBPs, CSPs and housekeeping genes including the gene name, unigene sequences, lengths, predicted protein sequences and the annotation in NCBI-Nr database, predicted protein domains and expression abundance are listed in the Supplementary Dataset.

Identification of odorant receptors. Bioinformatic analysis of the *C. formicarius* male and female antennal transcriptomes identified 54 candidate OR transcripts, which were classified as belonging to the 7-transmembrane receptors superfamily. Of these, 40 represented full-length open reading frames (ORF), based upon the presence of predicted start and stop codons and 5′ and 3′ untranslated regions (UTR). The highly conserved co-receptor (Orco) was identified in the *C. formicarius* transcriptomes, sharing 84.41% amino acid sequence identity with *C. bowringi* Orco. The specific ORs (53 ORs) in *C. formicarius* antennae shared low identity (19.26% to 59.52% identity) with known coleopteran ORs in the NCBI (see Supplementary Dataset). A phylogenetic analysis was conducted using a data set containing the sequences of all 54 ORs in *C. formicarius* and other coleopteran ORs from antennal transcriptomes or expressed in antenna (Fig. 2). Phylogenetic analysis demonstrated the clustering of CforORs with the previously defined coleopteran OR subgroups 1, 2, 7a and 7b, as well as the Orco subgroup[19], and the vast majority of CforORs were assigned to OR subgroup 7a (65%, 35 ORs).

As expected, the CforOrco gene exhibited the highest abundance both in male and female antennae (male RPKM: 645.76; female RPKM: 703.3). The transcriptional profiles of CforOR genes were characterized using qPCR, and the results revealed that all of the 53 CforORs except CforOR53 displayed predominantly antenna linked or otherwise biased expression levels. Although we did not identify sex-specific genes in these *C. formicarius* olfactory receptors, there were four (CforOR4, 47, 52, and 54) with significantly higher expression in the male antennae and four (CforOR13, 20, 23 and 24) with significantly higher expression in the female antennae, respectively (Fig. 3). Additionally, qPCR data for five CforORs (4, 13, 23, 52 and 54) mirrored the RNA-seq data.

Gustatory receptors. In the antennal transcriptomes, 11 candidate GR transcripts were identified. Most candidate CforGRs were partial fragments (only CforGR63a represent full-length proteins), encoding overlapping but distinct sequences. This establishes the proteins as being fragments of independent genes. A phylogeny
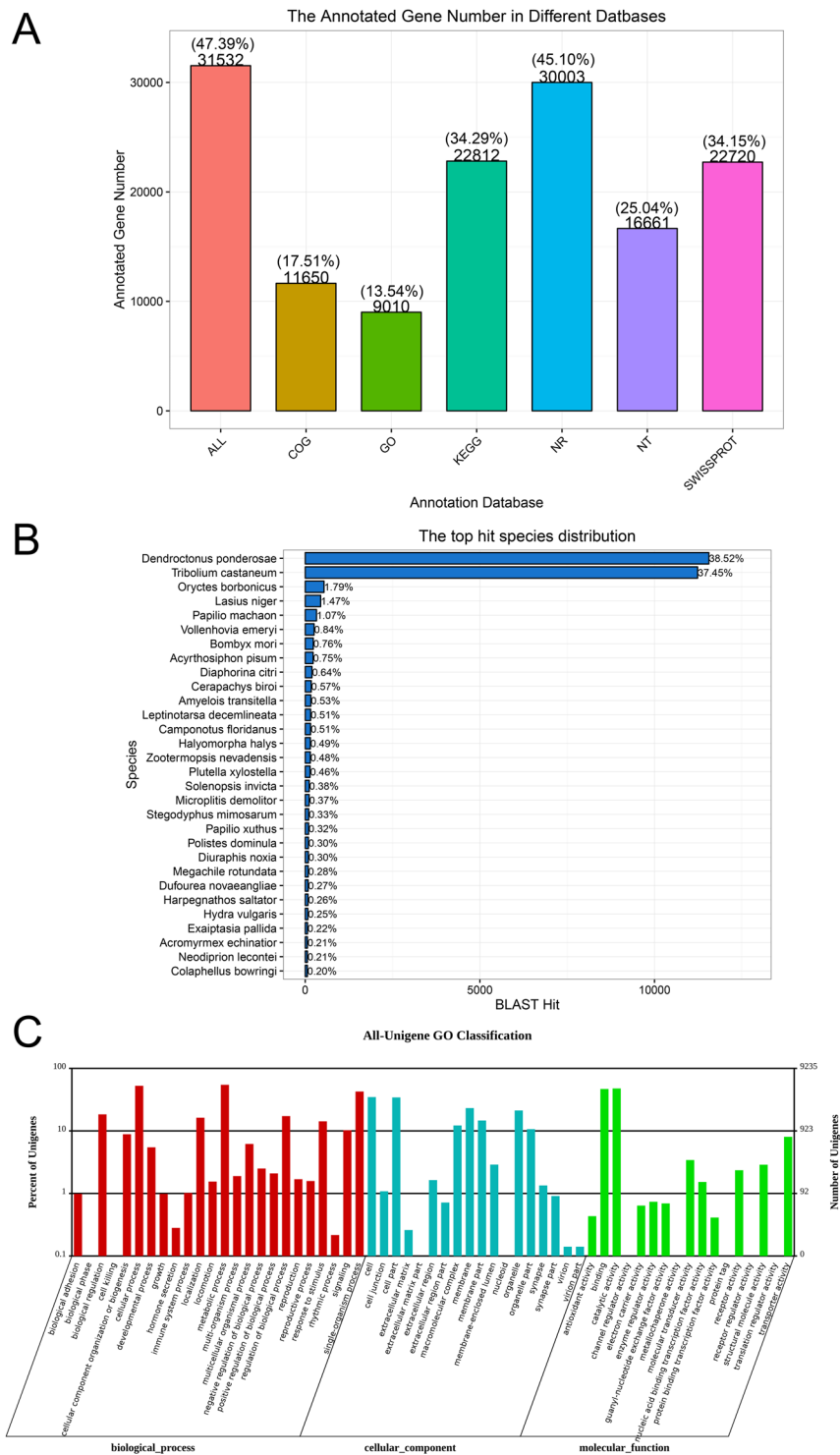
Figure 1. (**A**) Unigenes annotated through the different databases. (**B**) Percentage of homologous hits of the *C. formicarius* transcripts to other insect species. The *C. formicarius* transcripts were searched by BLASTx against the non-redundant protein database with a cutoff E-value of 10^{-5}. (**C**) Gene ontology (GO) classification of the *C. formicarius* transcripts with Blast2GO program. One unigene was annotated to more than one GO term.

was built with 11 CforGRs, and GRs from *T. castaneum*, and *D. melanogaster* (Fig. 4). The CforGRs grouped with their presumed Drosophila orthologues, which have been shown to have roles in carbon dioxide detection (GR21a and GR63a)[16, 49]. Two CforGRs clustered within known Drosophila sugar (DmelGR43a) receptors[50, 51], and the remaining CforGRs were assigned to different phylogenetic group with *T. castaneum* GRs. In addition, RPKM results showed low-level expression of all CforGRs in both male and female antennae (RPKM: 0~5.26).
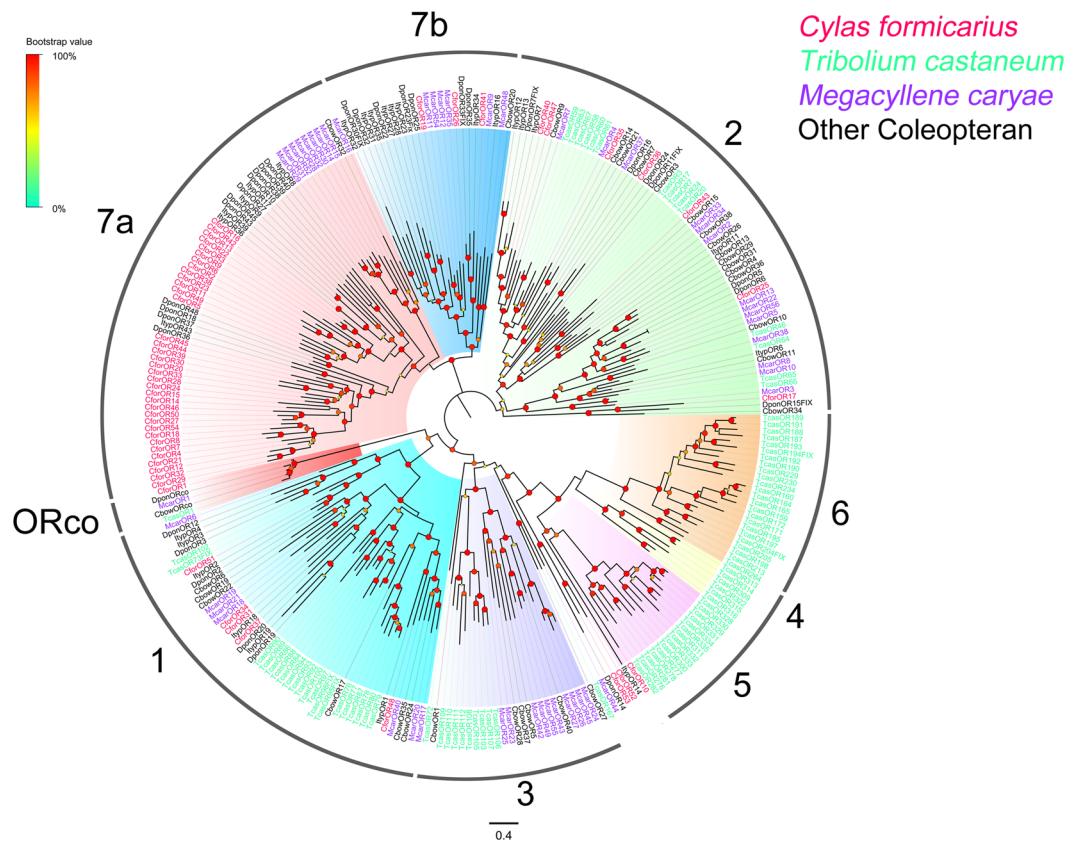
Figure 2. Phylogenetic analysis of ORs. Species abbreviations: Cfor, *Cylas formicarius*; Tcas, *Tribolium castaneum*; Mcar, *Megacyllene caryae*; Ityp, *Ips typographus*; Dpon, *Dendroctonus ponderosae*; Cbow, *Colaphellus bowringi*. Branch support (circles at the branch nodes) was estimated using an approximate likelihood ratio test based on the scale indicated at the top left. Bars indicate branch lengths in proportion to amino acid substitutions per site.

Ionotropic receptors.   Fifteen candidate iGluRs/IRs transcripts were identified from the antennal transcriptomes, which were predicted to encode ligand-gated cation channels (S1 and S2) with three transmembrane domains (M1, M2, and M3) or portions of domains. Among these, nine iGluRs/IRs represented full-length ORFs encoding more than 540 amino acids. All identified iGluRs/IRs in *C. formicarius* clustered with their orthologs from *T. castaneum*, *I. typographus*, *D. ponderosae* and *D. melanogaster* and were assigned to three phylogenetic groups including non-N-Methyl-D-aspartic acid (NMDA) iGluRs, IR co-receptors (IR25a/IR8a), and antennal IRs (Fig. 5A). With the antennal IRs, two IR co-receptors (CforIR25a and CforIR8a), and CforIR75c retained all key amino acids of predicted glutamate binding domains (R, T and D/E)[52], and one or more of these positions were absent in other antennal IR candidates indicating variable ligand binding properties (Fig. 5B).

The RPKM results showed that CforIR93a (male RPKM: 57.49; female RPKM: 78.79) displayed the highest expression levels, followed by CforIR76b. The qPCR results for *C. formicarius* antennal IRs revealed that all of the 11 candidates displayed predominately antenna linked expression levels. There were two (CforIR68 and 75 s) and two (CforIR64 and 93a) with significantly higher expression in the male and female antennae, respectively (Fig. 5C). These qPCR results matched the FPKM data for sex differences.

Sensory neuron membrane proteins.   Three candidate SNMP transcripts were identified that matched the CD36 family, with a full ORF and two transmembrane domains. Phylogenetic analysis showed that CforSNMP1 was clustered with the homologous SNMP1 group from other insect species, while CforSNMP2a and CforSNMP2b clustered with SNMP2 group homologs from other insects (Fig. 6A).

The SNMP1 subfamilies, CforSNMP1 with a relatively high abundance of transcripts in the antennal transcriptome of both sexes, appears to be expressed at high levels (male RPKM: 640.23; female RPKM: 457.03) in *C. formicarius* antennae. Semi-quantitative RT-PCR analysis revealed that CforSNMP1 was exclusively expressed in antennae (Fig. 6B).

Odorant binding proteins.   Analysis of the *C. formicarius* antennal transcriptomes identified 26 candidate OBP transcripts, which matched with insect pheromone/OBP domains. All but four of these CforOBP transcripts (CforOBP17, 24, 25 and 26) had a complete ORF and possessed signal peptides. Insect OBPs can be classed into subfamilies based on the presence of cysteine residues, including Classic OBPs, Plus-C OBPs, and Minus-C OBPs[53]. Multiple amino acid sequence alignments showed that 18 CforOBPs with six highly conserved
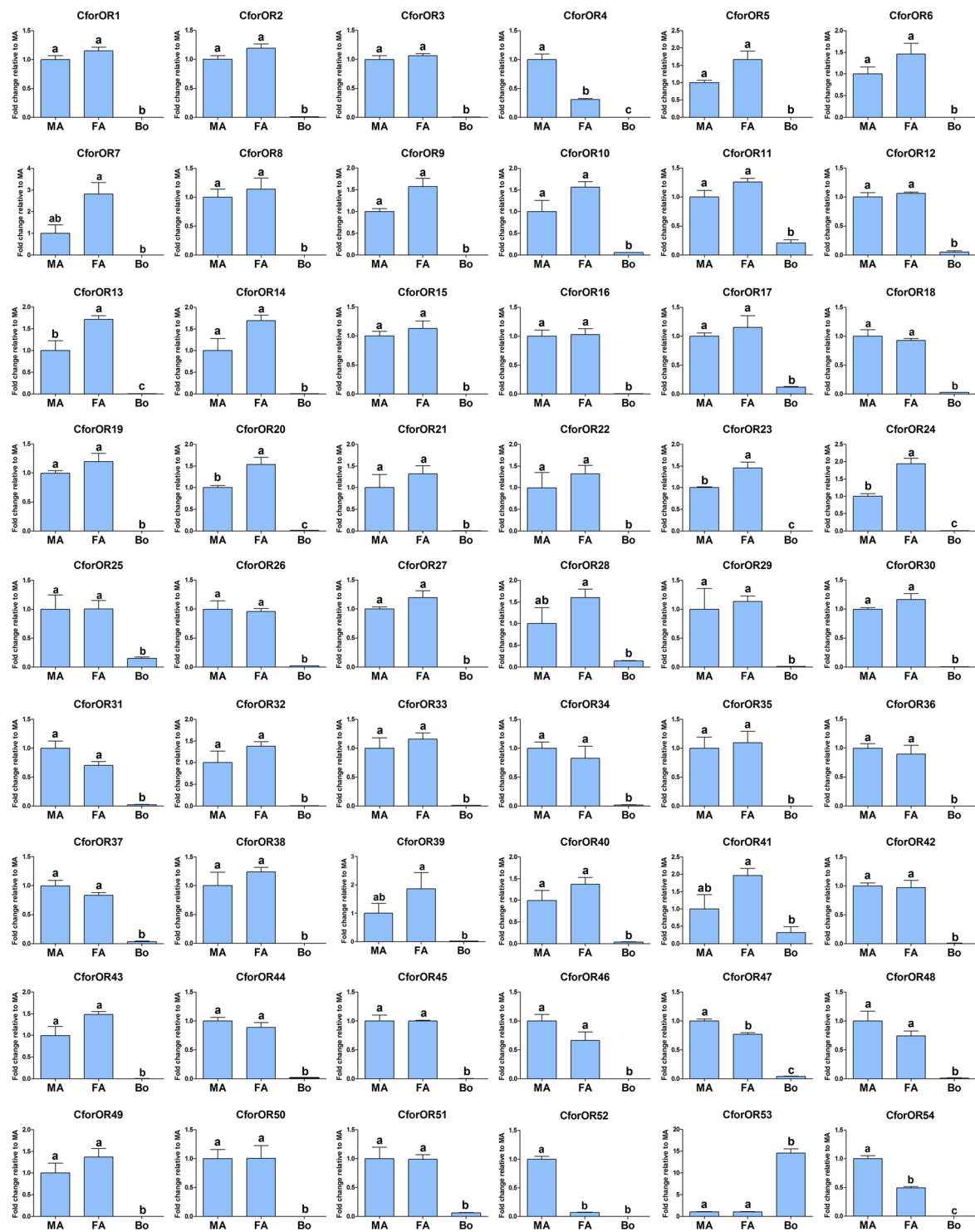
Figure 3. Relative expression levels of putative *C. formicarius* ORs in the male and female antennae, and whole insect body, using qPCR. Abbreviations: MA, female antennae; FA, female antennae; Bo, whole insect body without antennae. The relative expression level is indicated as mean ± SE (n = 3). Standard error is represented by the error bar, and different letters indicate significant differences between tissues (p < 0.05, ANOVA, HSD).

cysteine residues belonged to the Classic class, one belonged to the Plus-C class (CforOBP16), with 4–6 additional cysteines and one characteristic proline, and the remaining CforOBPs belonged to the Minus-C class, with a loss of two conserved cysteines (C2 and C5) (see Supplementary Figure S1). Phylogenetic analysis of CforOBPs with coleopteran sequences indicated that these CforOBPs segregated into the Classic, Plus-C, and Minus-C OBP sub-families (Fig. 7A).
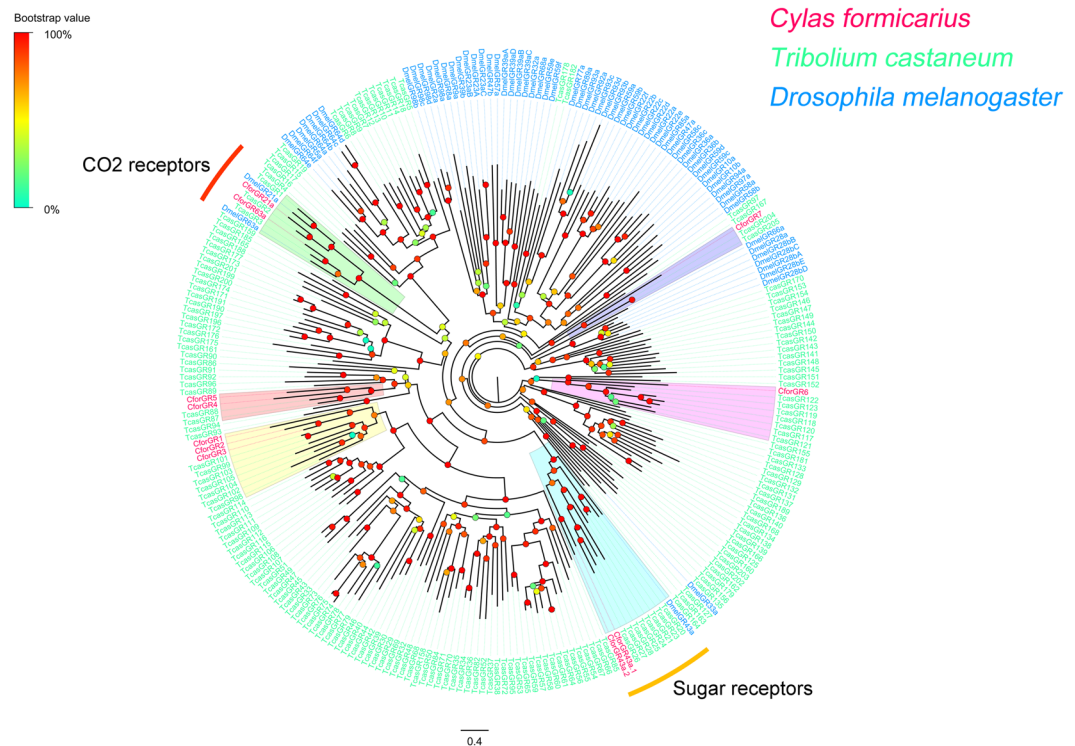
Figure 4. Phylogenetic analysis of GRs. Species abbreviations: Cfor, *Cylas formicarius*; Tcas, *Tribolium castaneum*; Dmel, *Drosophila melanogaster*. Branch support (circles at the branch nodes) was estimated using an approximate likelihood ratio test based on the scale indicated at the top left. Bars indicate branch lengths in proportion to amino acid substitutions per site.

The CforOBP transcripts displayed different patterns of tissue distribution and abundance. Only three transcripts (CforOBP2, 11 and 14) were almost exclusively transcribed in the antennae, while three transcripts (CforOBP8, 12 and 15) were abundant in the main insect olfactory and gustatory organs, i.e. antennae, mouthparts or tarsi. OBPs transcripts appear to be abundant in all or several body parts (Fig. 7B). RPKM data showed that four OBPs (CforOBP5, 10, 11 and 14) were expressed significantly higher in the male antennae and five OBPs (CforOBP15, 16, 17, 19 and 20) were significantly higher expressed in female antennae.

Chemosensory proteins.   In total, 12 CSP encoding candidates were identified in the antennal transcriptomes, which were matched up with the OS-D domains. Among these CforCSPs, 8 contained a complete ORF and a signal peptide. Alignments of the amino acid sequences of *C. formicarius* CSP revealed the presence of a highly conserved four-cysteine profile (see Supplementary Figure S2). The phylogenetic tree of CforCSPs formed two lineage-specific clades, including clade 1 (2 CSPs) and clade 2 (10 CSPs) (Fig. 8A). Semi-quantitative RT-PCR analysis revealed that all CforCSPs were present in all or several body parts, and no candidate was enriched in the antenna (Fig. 8B).

Discussion

Investigations on the molecular mechanisms of olfaction in *C. formicarius* provide insight into chemoreception and could ultimately lead to the identification of new targets for olfactory disruption and development of safer pest control strategies. We sequenced and analyzed the transcriptome of male and female *C. formicarius* antennae and identified and analyzed expression patterns of a set of novel chemosensory genes including 54 ORs, 11 GRs, 15 IRs, 3 SNMPs, 26 OBPs and 12 CSPs. Our results will be useful for functional characterization of olfaction genes in *C. formicarius*.

Based on phylogenetic tree analysis, the *C. formicarius* Orco (CforOR1) grouped with other coleopteran Orcos, but revealed a large species-specific expansion of *C. formicarius* ORs distributed in the previously defined coleopteran OR subfamily 7a that differ from other coleopterans[22]. This may reflect that *C. formicarius* inhabits a different ecological niche than the other coleopterans. Generally, sexually dimorphic expression of ORs in the antennae, indicate possible pheromone receptors contributing to facilitation of sexual behaviors. Typically, Lepidoptera sex pheromones are produced by females and they affect males. Several moth sex pheromone ORs have been functionally characterized, and most are expressed at higher levels in the male antennae[54–56]. *C. formicarius* females release the sex pheromone, (Z)-3-dodecen-1-ol (E)-2-butenoate, and elicit strong male attraction[39, 40]. Here, we found that four male-biased CforORs (CforOR4, 47, 52, and 54) reside in different clades, suggesting that the sex pheromone receptors might be screened from these candidates, according to the lock-and-key mechanism as they are activated by single pheromone components[57]. Additionally, ORs expressed biasedly in the
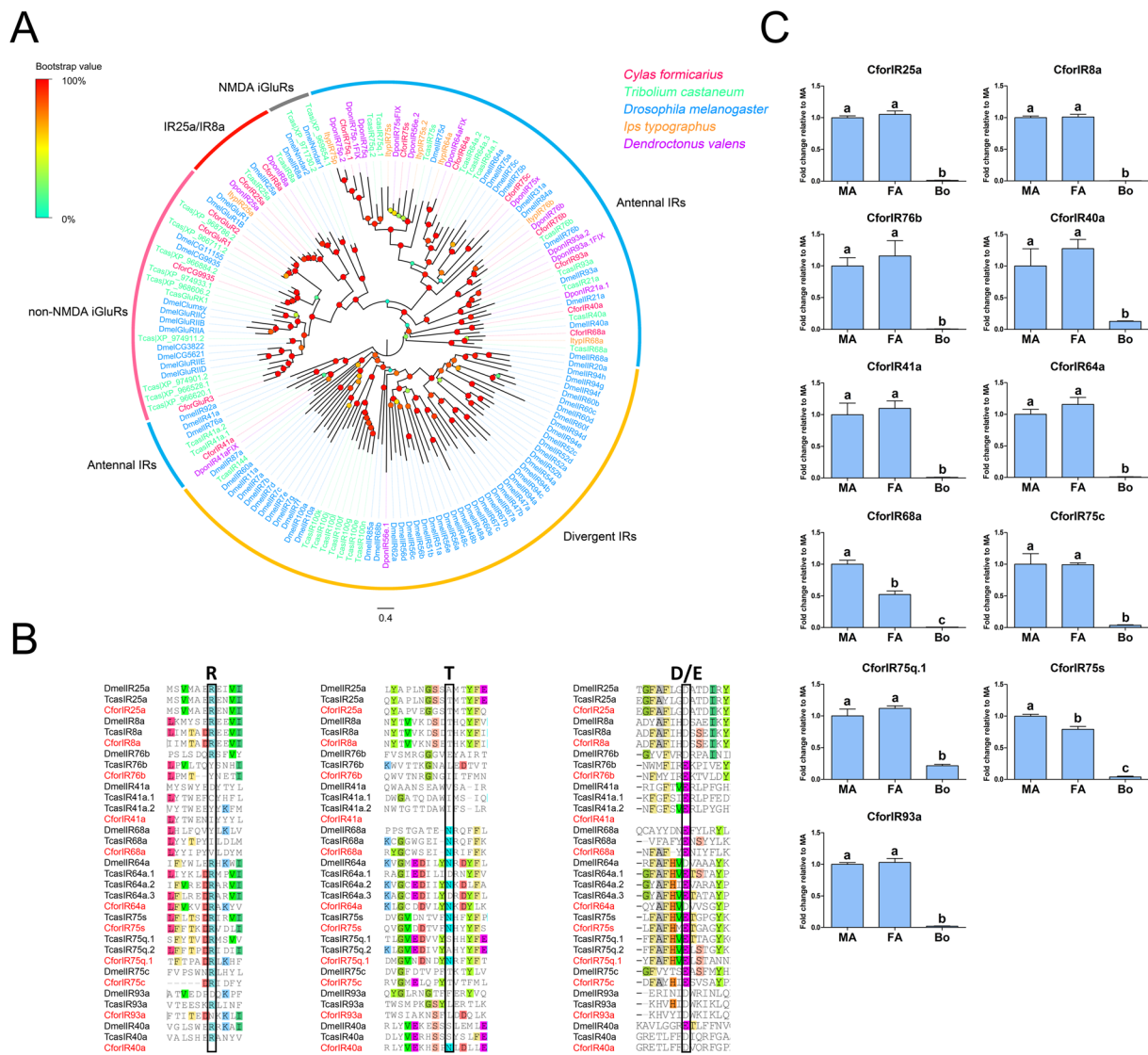
Figure 5. (**A**) Phylogenetic analysis of iGluRs/IRs. Species abbreviations: Cfor, *Cylas formicarius*; Tcas, *Tribolium castaneum*; Dmel, *Drosophila melanogaster*; Ityp, *Ips typographus*; Dpon, *Dendroctonus ponderosae*. Branch support (circles at the branch nodes) was estimated using an approximate likelihood ratio test based on the scale indicated at the top left. Bars indicate branch lengths in proportion to amino acid substitutions per site. (**B**) Excerpts from the amino acids alignment showing the predicted iGluRs/IRs binding domains. (**C**) Relative expression levels of putative *C. formicarius* antennal IRs in the male and female antennae, and whole insect body. Abbreviations: MA, female antennae; FA, female antennae; Bo, whole insect body without antennae. The relative expression level is indicated as mean ± SE (n = 3). Standard error is represented by the error bar, and different letters indicate a significant difference between tissues (p < 0.05, ANOVA, HSD).

female antennae are predicted to function in egg-laying-related odorant detection; ORs expressed evenly in the male and female antennae are predicted to function in general odorant perception[58, 59]. Therefore, we hypothesize that some or all of female-biased CforORs (CforOR13, 20, 23 and 24) appear to be involved in female specific behaviors i.e. finding plant hosts for oviposition, while remaining ORs with equal in the female and male antennae might be dedicated to general odorant detection or others.

The GR family of insect chemoreceptors includes receptors for CO_2, D fructose, sucrose, bitter, and other receptors[60]. Gustatory receptors perceive essential nutrients whose chemical structures remain constant such as sugars and CO_2 receptors, thus CO_2 and sugar receptor genes are highly conserved among insects[61–63]. A total of 11 GR transcripts were identified in the *C. formicarius* antennal transcriptome dataset. The GR family in *C. formicarius* includes two putative CO_2 receptors (CforGR21a and 63a), which were the orthologous genes of the GR21a/GR63a CO_2 receptor, and two sugar receptors (CforGR43a.1 and 43a.2), share the same clade with DmelGR43a, a receptor of fructose as the nutrient sensor in the *Drosophila* brain[50]. We suggest that these GRs potentially have similar functions.
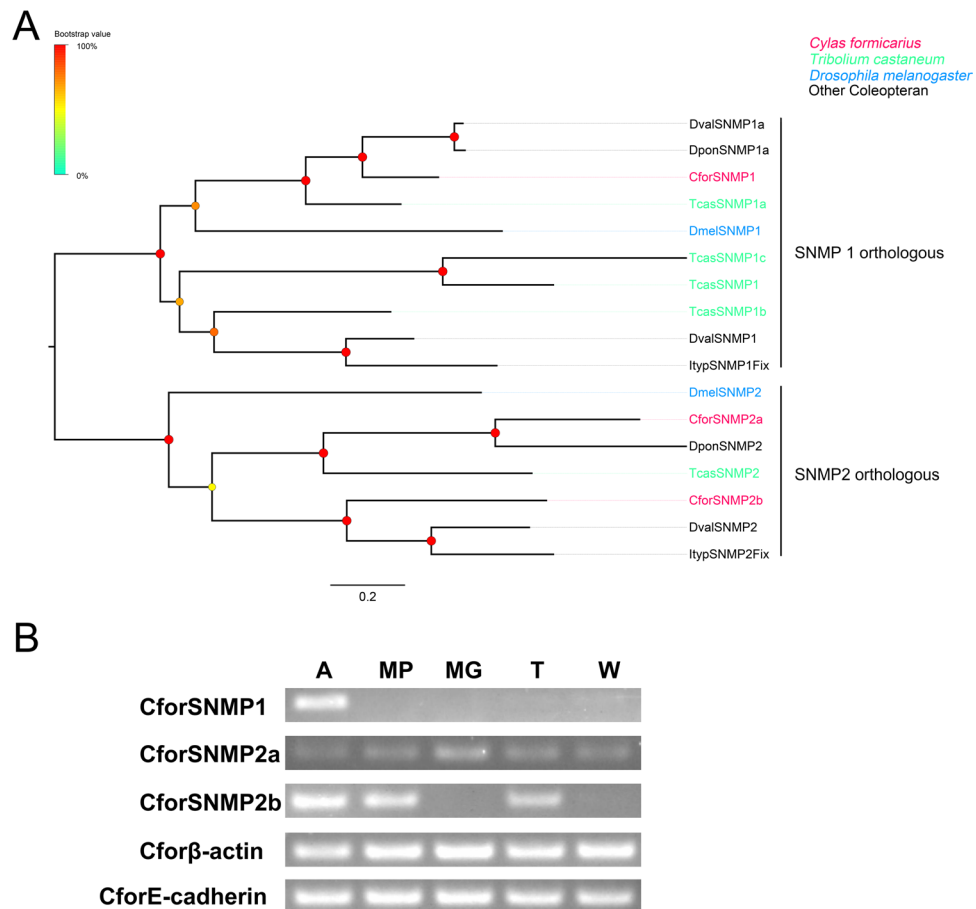
Figure 6. (**A**) Phylogenetic analysis of SNMPs. Species abbreviations: Cfor, *Cylas formicarius*; Tcas, *Tribolium castaneum*; Dmel, *Drosophila melanogaster*. Ityp, *Ips typographus*; Dpon, *Dendroctonus ponderosae*; Dval, *Dendroctonus valens*. Branch support (circles at the branch nodes) was estimated using an approximate likelihood ratio test based on the scale indicated at the top left. Bars indicate branch lengths in proportion to amino acid substitutions per site. (**B**) Transcriptional profiles of putative *C. formicarius* SNMPs in different body parts as determined using semi-quantitative RT-PCR. Two housekeeping genes, β-actin (Cforβ-actin) and E-cadherin (CforE-cadherin), were used as internal references to test the integrity of each cDNA template. Abbreviations: A: antenna; MP, mouthparts; MG, midgut; T, tarsus; W, wing.

IR is another chemosensory receptor family that has been characterized in *D. melanogaster*[52]. IRs are a conserved family and function as chemoreceptors for the detection of a variety of chemical molecules[9–11, 14, 64, 65]. Phylogenetic analysis indicated two co-receptors (IR25a and IR8a), five antennal IRs (IR40a, IR64a, IR68a, IR76b, and IR93a) in *C. formicarius* that have orthologs in both *D. melanogaster* and *T. castaneum*. The antennal IRs in *D. melanogaster* have been confirmed by functional studies. IR40a is required for response to the insect repellent DEET[10], and IR64a is acid sensitive[14]. IR76b is co-expressed with IR41a to mediate long-range attraction to odor[9], and IR93a acts with different combinations of other IRs to mediate physiological and behavioral responses to both temperature and moisture cues[66]. These IRs orthologs in *C. formicarius* might have similar sensory functions. But the function(s) of IR68a remain uncharacterized. We found CforIR68a and CforIR75s had male-biased expression in antennae and are likely to play a role in the perception of female pheromones.

In *D. melanogaster*, SNMP1 is expressed specifically in pheromone-sensitive ORNs, and is responsible for the sensitivity of these neurons to cVA stimulation[7, 8, 67]. In the present study, two SNMP subfamilies (both SNMP1 and SNMP2) were obtained from *C. formicarius* antennal transcriptomes. As expected, SNMP1 homologs (CforSNMP1) presented a clear antenna-predominant expression, while SNMP2 homologs (CforSNMP2a and 2b) were not restricted to the antennae. Additionally, RPKM results showed that CforSNMP1 is the only SNMP gene to have high expression in the antennae compared to the others suggesting that CforSNMP1 may play a role similar to its homolog in *D. melanogaster*.

OBP are commonly regarded as solubilizers and carriers of odorants and sex pheromones[68]. Additionally, OBPs may contribute to the sensitivity of the olfactory system[69–72]. They are not restricted to chemosensory tissues and may participate in other non-sensory functions[68]. We found that most CforOBPs were distributed in all examined body parts, and only three classic OBPs (OBP2, 11 and 14) were present exclusively in the antennae, suggesting that these participate in olfactory sensory functions. In the antennal-specific OBPs, CforOBP14 displayed male-biased expression that may play a role in sex pheromone perception. In some cases, CSPs as well as OBPs act as carriers of odorant molecules[73–76]. However, in our RT-PCR analysis, these CSPs were abundant in
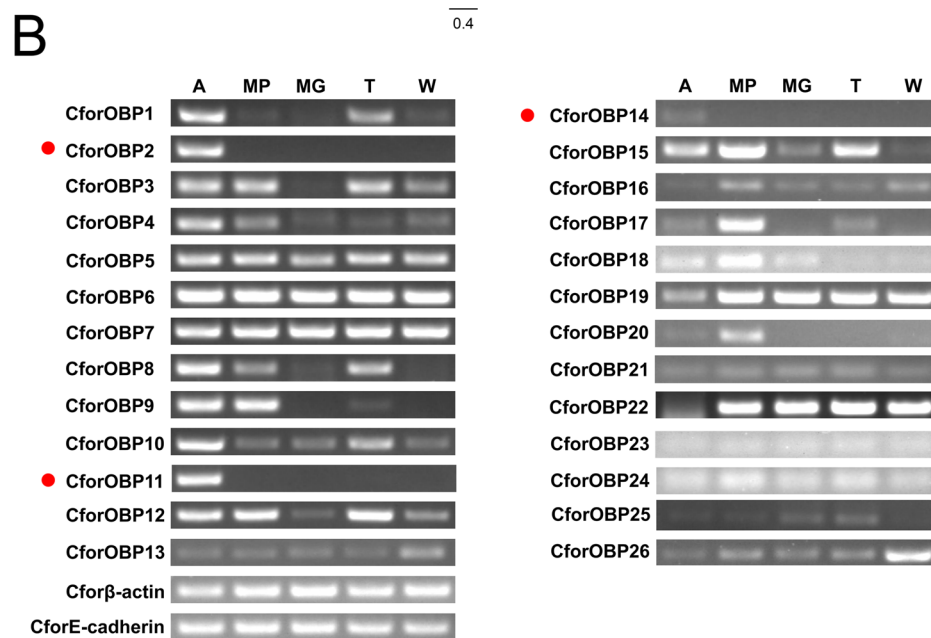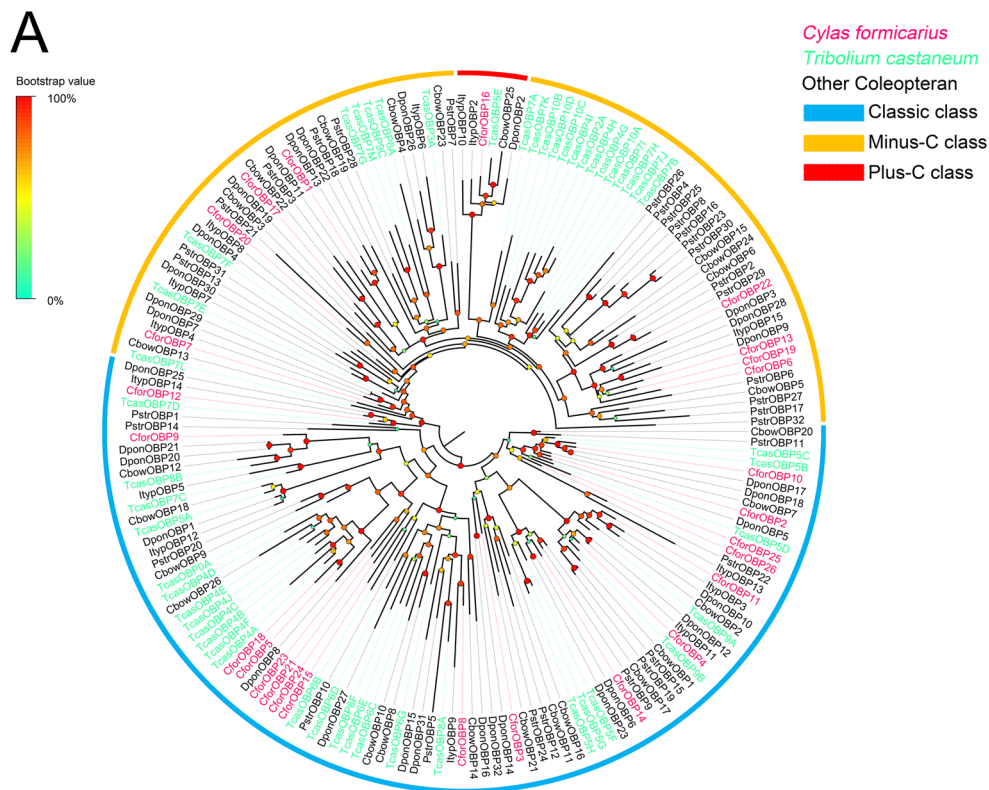
Figure 7. (**A**) Phylogenetic analysis of OBPs. Classic OBPs from *C. formicarius* and other coleopteran species form a clade labeled in red. Labeled in orange are Minus-C OBPs from *C. formicarius* and other coleopterans, and labeled in blue are Plus-C OBPs. Species abbreviations: Cfor, *Cylas formicarius*; Tcas, *Tribolium castaneum*; Ityp, *Ips typographus*; Dpon, *Dendroctonus ponderosae*; Cbow, *Colaphellus bowringi*; Pstr, *Phyllotreta striolata*. Branch support (circles at the branch nodes) was estimated using an approximate likelihood ratio test based on the scale indicated at the top left. Bars indicate branch lengths in proportion to amino acid substitutions per site. (**B**) Transcriptional profiles of putative *C. formicarius* OBPs in different body parts determined using semi-quantitative RT-PCR. The OBPs expressed specifically in antennae are labeled with red dots. Two reference genes, β-actin (Cforβ-actin) and E-cadherin (CforE-cadherin), were used as internal references to test the integrity of each cDNA template. Abbreviations: A: antenna; MP, mouthparts; MG, midgut; T, tarsus; W, wing.
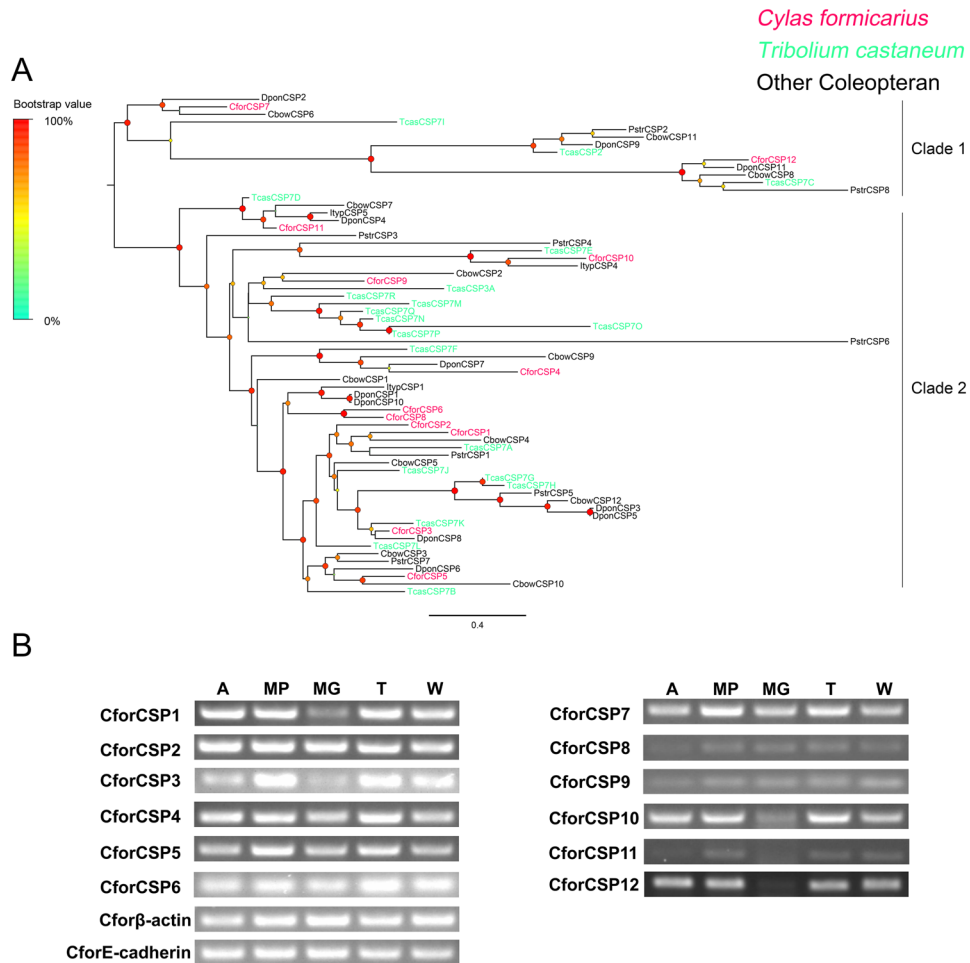
Figure 8. (**A**) Phylogenetic analysis of CSPs. Species abbreviations: Cfor, *Cylas formicarius*; Tcas, *Tribolium castaneum*; Ityp, *Ips typographus*; Dpon, *Dendroctonus ponderosae*; Cbow, *Colaphellus bowringi*; Pstr, *Phyllotreta striolata*. Branch support (circles at the branch nodes) was estimated using an approximate likelihood ratio test based on the scale indicated at the top left. Bars indicate branch lengths in proportion to amino acid substitutions per site. (**B**) Transcriptional profiles of putative *C. formicarius* CSPs in different body parts determined using semi-quantitative RT-PCR. Two housekeeping genes, β-actin (Cforβ-actin) and E-cadherin (CforE-cadherin), were used as internal references to test the integrity of each cDNA template. Abbreviations: A: antenna; MP, mouthparts; MG, midgut; T, tarsus; W, wing.

all body parts examined, suggesting that, in *C. formicarius*, these CSPs could be also be involved in non-sensory functions.

Conclusion

We obtained substantial molecular information on the combined male and female antennal transcriptome of *C. formicarius* using high-throughput sequencing technology. The goal was to identify gene families involved in odorant detection. Based on the transcriptomic analysis, a repertoire of 54 ORs, 11 GRs, 15 IRs, 3 SNMPs, 26 OBPs and 12 CSPs were identified and further analyzed for their expression profiles. Our results directly provide a foundation for advanced functional studies of these olfactory genes in *C. formicarius*.

Methods

Insect rearing and collection.   Sweetpotato weevils were originally collected in infested sweet potato fields in Xinhui County, Jiangmen City, Guangdong Province, China (E113°13′, N22°25′). The collected weevils were reared at the Institute for Management of Invasive Alien Species, Zhongkai University of Agriculture and Engineering. They were fed commercial sweet potato and kept in a climatic chamber at 28 °C ± 1 °C, 80% relative humidity, and a 16 h: 8 h light dark photoperiod. After 5–6 generations, 3–4 week-old adults, prior to reproduction, were used for experiments.

RNA isolation and sequencing.   Antennae of 200 males and 200 females were hand-dissected, flash frozen in liquid nitrogen, and then crushed with a hand mortar. Total RNA was isolated using TRIzol reagent (Invitrogen, USA) from male antennae and female antennae individually according to manufacturer instructions. RNA samples were treated with DNase (Qiagen, Germany) and then purified with RNeasy Mini Kit (Qiagen,

Germany). The quality and concentration of total RNA were examined using Qubit@2.0 Fluorometer (Invitrogen, Life Technologies), and RNA integrity was further confirmed using an Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA, USA).

Paired-end cDNA libraries were generated from purified RNA (0.5 μg of each sample) using TruSeq RNA Sample Preparation Kit v2 (Illumina Inc., San Diego, CA) according to Illumina instructions and sequenced on the Illumina HiSeq. 4000 platform. The raw sequence transcriptome data from the female and male antennae libraries were deposited in the NCBI Short Read Archive (SRA) database as BioProject Accession Number SRP067907.

De novo transcriptome assembly and gene annotation. Within the combined female and male antennae libraries, raw data (raw reads) of fastq format were first processed to remove unknown (poly-N) or low-quality sequences and adaptor sequences using FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/), and then assembled into unigenes using the Trinity pipeline (ver. r2013-02-25) with the default assembly parameters. All unigenes were annotated using BLASTx search against with a cutoff E-value of 10^{-5} the following databases: the non-redundant protein sequence (Nr), non-redundant nucleotide (Nt), Pfam, Clusters of Orthologous Groups (KOG/COG), Swiss-Prot, Kyoto Encyclopedia of Genes and Genomes (KEGG) and Gene Ontology (GO) databases, according to the highest sequence similarity.

Chemosensory gene identification. Identification of putative *C. formicarius* chemosensory gene families using both BLASTx searches in Nr and tBLASTn searches using known sequences as queries. *C. formicarius* chemosensory genes were also used as queries to identify additional genes (tBLASTx and BLASTp). Repetitions were completed until no new candidates were identified. The open reading frames (ORFs) of putative chemosensory genes transcripts were identified using the ORF finder tool (http://www.ncbi.nlm.nih.gov/gorf/gorf.html), and further confirmed using the SMART BLAST searching result according to the best matches with well-studied reference species. Definitive protein domains (e.g. transmembrane domains, signal peptides, secondary structures, etc.) in chemosensory genes were predicted by queries against InterPro using the InterProScan tool plug-in in Geneious (ver. 9.1.3.)[77].

Alignment and phylogenetic analyses. Amino acid sequences were aligned using the MAFFT alignment tool plug-in in Geneious (ver. 9.1.3.) (E-INS-I parameter set)[78]. Phylogenetic relationship was deduced using maximum likelihood analysis with FastTree2 (JTT substitution model, 1000 bootstrap replications)[79, 80] and subsequently viewed and graphically edited in FigTree v1.4.2 (http://tree.bio.ed.ac.uk/software/figtree). Incomplete transcripts lacking sufficient overlap in alignments and transcripts less than 180 amino acids in length (except for the OBPs where full-length transcripts are generally shorter than 200 amino acids) were excluded from phylogenetic analyses to ensure that the analysed transcripts corresponded to individual genes and to maintain greater accuracy in the analyses. The respective chemosensory genes families used for constructing phylogenetic trees are listed in Supplementary Table S2.

Abundance estimation and differential expression analyses. Gene expression levels were assessed using the RSEM v1.2.8[81] separately for the filtered reads from female and male antennae libraries, and their converted FPKM values (fragments per kilobase per million reads)[82]. Raw read counts data generated via RSEM were normalized using the Trimmed Mean of M-value normalization method[83] and were used for differential expression analyses between female and male antenna using edgeR package (v3.4.2)[84].

Tissue expression analyses. Semi-quantitative RT-PCR was employed to investigate and compare the expression of SNMPs, OBPs and CSPs in different tissues including antennae, mouthparts, midgut, foreleg tarsus, and wings, to define the antenna-predominant candidates. Total RNA from the analyzed tissues was isolated using an RNeasy Mini kit (Qiagen, Germany), and cDNA was synthesized using PrimeScript RT reagent Kit (Takara, China). An equal amount of cDNA (100 ng) was used as the Semi-quantitative RT-PCR templates. Primers were designed using Primer3web (ver. 4.0.0) (http://primer3.ut.ee/) and are listed in Supplementary Table S3. PCR was performed under the following conditions: 95 °C for 2 min, followed by 35 cycles of 95 °C for 30 sec, 56 °C for 30 sec, 72 °C for 1 min, and a final extension for 10 min at 72 °C. To reach reproducibility, each semi-quantitative RT-PCR was repeated three times with two independently isolated RNA samples. The semi-quantitative PCR products were analyzed on 1.5% agarose gel electrophoresis. According to a previous study[85], two housekeeping genes, β-actin (Cforβ-actin) and E-cadherin (CforE-cadherin) from *C. formicarius* antennal transcriptomes were used as the controls.

The expression profiles of ORs and antennal IRs were analyzed using quantitative real-time PCR (qPCR). First, RNA isolation and cDNA synthesis were performed on samples including 20 male or female antennae each and 10 whole insect bodies without antennae (male and female, ratio 1:1). qPCR was performed on a LightCycler 480 system (Roche Applied Science) in a reaction volume of 10 μl SYBR Green I Master mix (Roche Applied Science), 1 μl of each primer (0.5 μM), 2 μl (approximately 2.0 ng) of sample cDNA and 6 μl sterilized ultrapure H$_2$O. The cycling parameters were as follows: denaturation at 95 °C for 5 min, followed by 45 cycles of 95 °C for 10 sec, and 60 °C for 20 sec. A melting curve analysis was then performed at 95 °C for 20 sec, 60 °C for 30 sec, and 95 °C for 30 sec in order to determine the specificity of primers. Negative controls without template were included in each reaction. For each gene, three biological replications (3 separate RNA extractions from samples) were performed with each biological replication measured in three technical replications. The results were analyzed using LightCycler 480 Gene Scanning Software. The comparative $2^{-\Delta\Delta CT}$ method was used to calculate the relative expression levels of each gene[86], with two housekeeping genes β-actin and E-cadherin as the reference genes. The comparative analyses of each target gene among different tissues were determined using a one-way nested analysis of variance (ANOVA) followed by Tukey's honest significance difference (HSD) test using Prism 6.0 (GraphPad Software, CA). Values are presented as mean ± SE.

References

1. Leal, W. S. Odorant Reception in Insects: Roles of Receptors, Binding Proteins, and Degrading Enzymes. *Annu. Rev. Entomol.* **58**, 373–391 (2013).
2. Suh, E., Bohbot, J. & Zwiebel, L. J. Peripheral olfactory signaling in insects. *Curr. Opin. Insect Sci.* **6**, 86–92 (2014).
3. Su, C. Y., Menuz, K. & Carlson, J. R. Olfactory perception: receptors, cells, and circuits. *Cell* **139**, 45–59 (2009).
4. Hildebrand, J. G. Analysis of chemical signals by nervous systems. *Proc. Natl. Acad. Sci.* **92**, 67–74 (1995).
5. Mustaparta, H. Chemical information processing in the olfactory system of insects. *Physiol. Rev.* **70**, 199–245 (1990).
6. Gomez-Diaz, C. *et al.* A CD36 ectodomain mediates insect pheromone detection via a putative tunnelling mechanism. *Nat. Commun.* **7**, 11866 (2016).
7. Jin, X., Ha, T. S. & Smith, D. P. SNMP is a signaling component required for pheromone sensitivity in Drosophila. *Proc. Natl. Acad. Sci.* **105**, 10996–1001 (2008).
8. Benton, R., Vannice, K. S. & Vosshall, L. B. An essential role for a CD36-related receptor in pheromone detection in Drosophila. *Nature* **450**, 289 (2007).
9. Hussain, A. *et al.* Ionotropic Chemosensory Receptors Mediate the Taste and Smell of Polyamines. *PLoS Biol.* **14**, e1002454 (2006).
10. Kain, P. *et al.* Odour receptors and neurons for DEET and new insect repellents. *Nature* **502**, 507–512 (2013).
11. Ai, M. *et al.* Ionotropic glutamate receptors IR64a and IR8a form a functional odorant receptor complex *in vivo* in Drosophila. *J. Neurosci.* **33**, 10741–10749 (2013).
12. Silbering, A. F. *et al.* Complementary Function and Integrated Wiring of the Evolutionarily Distinct Drosophila Olfactory Subsystems. *J. Neurosci.* **31**, 13357–13375 (2011).
13. Grosjean, Y. *et al.* An olfactory receptor for food-derived odours promotes male courtship in Drosophila. *Nature* **478**, 236–240 (2011).
14. Ai, M. *et al.* Acid sensing by the Drosophila olfactory system. *Nature* **468**, 691–695 (2010).
15. Turner, S. L. & Ray, A. Modification of CO2 avoidance behaviour in Drosophila by inhibitory odorants. *Nature* **461**, 159–277 (2009).
16. Jones, W. D., Cayirlioglu, P., Kadow, I. G. & Vosshall, L. B. Two chemosensory receptors together mediate carbon dioxide detection in Drosophila. *Nature* **445**, 86–90 (2008).
17. Holdcraft, R., Rodriguez-Saona, C. & Stelinski, L. L. Pheromone Autodetection: Evidence and Implications. *Insects* **7**, 17 (2016).
18. Hanks, L. M. & Millar, J. G. Sex and Aggregation-Sex Pheromones of Cerambycid Beetles: Basic Science and Practical Applications. *J. Chem. Ecol.* **42**, 631–654 (2016).
19. Engsontia, P. *et al.* The red flour beetle's large nose: an expanded odorant receptor gene family in *Tribolium castaneum*. *Insect Biochem. Mol. Biol.* **38**, 387–397 (2008).
20. Richards, S. *et al.* The genome of the model beetle and pest *Tribolium castaneum*. *Nature* **452**, 949–955 (2008).
21. Mitchell, R. F. *et al.* Sequencing and characterizing odorant receptors of the cerambycid beetle *Megacyllene caryae*. *Insect Biochem. Mol. Biol.* **42**, 499–505 (2012).
22. Andersson, M. N. *et al.* Antennal transcriptome analysis of the chemosensory gene families in the tree killing bark beetles, *Ips typographus* and *Dendroctonus ponderosae* (Coleoptera: Curculionidae: Scolytinae). *BMC Genomics* **14**, 198 (2013).
23. Mamidala, P. *et al.* Identification of Odor-Processing Genes in the Emerald Ash Borer, *Agrilus planipennis*. *PloS one* **8**, e56555 (2013).
24. Li, H., Zhang, A., Chen, L. Z., Zhang, G. & Wang, M. Q. Construction and analysis of cDNA libraries from the antennae of *Batocera horsfieldi* and expression pattern of putative odorant binding proteins. *J. Insect Sci.* **14**, 57 (2014).
25. Chen, H., Lin, L., Xie, M., Zhang, G. & Su, W. De novo sequencing, assembly and characterization of antennal transcriptome of *Anomala corpulenta* Motschulsky (Coleoptera: Rutelidae). *PLoS One* **9**, e114238 (2014).
26. Li, X. *et al.* Chemosensory Gene Families in Adult Antennae of *Anomala corpulenta* Motschulsky (Coleoptera: Scarabaeidae: Rutelinae). *PLoS One* **10**, e0121504 (2015).
27. Gu, X. C., Zhang, Y. N., Kang, K., Dong, S. L. & Zhang, L. W. Antennal Transcriptome Analysis of Odorant Reception Genes in the Red Turpentine Beetle (RTB). *Dendroctonus valens. PLoS One.* **10**, e0125159 (2015).
28. Liu, S. *et al.* Identification of candidate chemosensory genes in the antennal transcriptome of *Tenebrio molitor* (Coleoptera: Tenebrionidae). *Comp. Biochem. Physiol. Part D Genomics Proteomics* **13**, 44–51 (2015).
29. Li, X. M. *et al.* Candidate chemosensory genes identified in *Colaphellus bowringi* by antennal transcriptome analysis. *BMC Genomics* **16**, 1028 (2015).
30. Wang, Y., Chen, Q., Zhao, H. & Ren, B. Identification and Comparison of Candidate Olfactory Genes in the Olfactory and Non-Olfactory Organs of Elm Pest *Ambrostoma quadriimpressum* (Coleoptera: Chrysomelidae) Based on Transcriptome Analysis. *PLoS One* **11**, e0147144 (2016).
31. Antony, B. *et al.* Identification of the genes involved in odorant reception and detection in the palm weevil *Rhynchophorus ferrugineus*, an important quarantine pest, by antennal transcriptome analysis. *BMC Genomics* **17**, 69 (2016).
32. Wu, Z. *et al.* Differential Expression Analysis of Chemoreception Genes in the Striped Flea Beetle *Phyllotreta striolata* Using a Transcriptomic Approach. *PLoS One* **11**, e0153067 (2016).
33. Hu, P., Wang, J., Cui, M., Tao, J. & Luo, Y. Antennal transcriptome analysis of the Asian longhorned beetle *Anoplophora glabripennis*. *Sci. Rep.* **6**, 26652 (2016).
34. Bin, S. Y. *et al.* Antennal and abdominal transcriptomes reveal chemosensory gene families in the coconut hispine beetle, *Brontispa longissima*. *Sci. Rep.* **7**, 2809 (2017).
35. Chalfant, R. B., Jansson, R. K., Seal, D. R. & Schalk, J. M. Ecology and management of sweet potato insects. *Entomology* **35**, 157–180 (1990).
36. Sherman, M. & Tamashiro, M. The sweetpotato weevils in Hawaii, their biology and control. *Hawaii Agricultural Experiment Station Technical Bulletin* (1954).
37. Leng, P. H. & Reddy, G. V. P. Bioactivity of Selected Eco-Friendly Pesticides Against *Cylas formicarius* (Coleoptera: Brentidae). *Fla. Entomol.* **2014**(95), 1040–1047 (2014).
38. Pickett, J. A., Woodcock, C. M., Midega, C. A. & Khan, Z. R. Push-pull farming systems. *Curr. Opin. Biotechnol.* **26**, 125–132 (2014).
39. Coffelt, J. A., Vick, K. W., Sower, L. L. & Mcclellan, W. T. M. Sex pheromone of the sweetpotato weevil, *Cylas formicarius* elegantulus laboratory bioassay and evidence for a multiple component system. *Environ. Entomol.* **7**, 756–758 (1978).
40. Heath, R. R. *et al.* Identification of sex pheromone produced by female sweetpotato weevil, *Cylas formicarius* elegantulus (Summers). *J. Chem. Ecol.* **12**, 1489–1503 (1986).
41. Reddy, G. V., Gadi, N. & Taianao, A. J. Efficient sex pheromone trapping: catching the sweetpotato weevil. *Cylas formicarius. J. Chem. Ecol.* **38**, 846–853 (2012).
42. Reddy, G. V., Wu, S., Mendi, R. C. & Miller, R. H. Efficacy of pheromone trapping of the sweetpotato weevil (Coleoptera: Brentidae): based on dose, septum age, attractive radius, and mass trapping. *Environ. Entomol.* **43**, 767–773 (2014).
43. Liu, Z., Smagghe, G., Lei, Z. & Wang, J. J. Identification of Male- and Female-Specific Olfaction Genes in Antennae of the Oriental Fruit Fly (*Bactrocera dorsalis*). *PLoS One* **11**, e147783 (2016).
44. Li, X. M. *et al.* Molecular Characterization and Sex Distribution of Chemosensory Receptor Gene Family Based on Transcriptome Analysis of *Scaeva pyrastri*. *PLoS One* **11**, e155323 (2016).
45. Leitch, O., Papanicolaou, A., Lennard, C., Kirkbride, K. P. & Anderson, A. Chemosensory genes identified in the antennal transcriptome of the blowfly *Calliphora stygia*. *BMC Genomics* **16**, 255 (2015).

46. Ma, L. *et al*. Identification and Comparative Study of Chemosensory Genes Related to Host Selection by Legs Transcriptome Analysis in the Tea Geometrid *Ectropis obliqua*. *PLoS One* **11**, e149591 (2016).

47. Zeng, F. F. *et al*. Identification and Comparative Expression Profiles of Chemoreception Genes Revealed from Major Chemoreception Organs of the Rice Leaf Folder, *Cnaphalocrocis medinalis* (Lepidoptera: Pyralidae). *PLoS One* **10**, e144267 (2015).

48. Gu, S. H. *et al*. Identification and comparative expression analysis of odorant binding protein genes in the tobacco cutworm *Spodoptera litura*. *Sci. Rep.* **5**, 13800 (2015).

49. Kwon, J. Y., Dahanukar, A., Weiss, L. A. & Carlson, J. R. The molecular basis of CO2 reception in Drosophila. *Proc. Natl. Acad. Sci.* **104**, 3574–3578 (2007).

50. Miyamoto, T., Slone, J., Song, X. & Amrein, H. A Fructose Receptor Functions as a Nutrient Sensor in the Drosophila Brain. *Cell* **151**, 1113–1125 (2012).

51. Miyamoto, T., Chen, Y., Slone, J. & Amrein, H. Identification of a Drosophila Glucose Receptor Using Ca2+ Imaging of Single Chemosensory Neurons. *PLoS One* **8**, e56304 (2013).

52. Benton, R., Vannice, K. S., Gomez-Diaz, C. & Vosshall, L. B. Variant Ionotropic Glutamate Receptors as Chemosensory Receptors in Drosophila. *Cell* **136**, 149–162 (2009).

53. Hekmat-Scafe, D. S., Scafe, C. R., McKinney, A. J. & Tanouye, M. A. Genome-wide analysis of the odorant-binding protein gene family in *Drosophila melanogaster*. *Genome Res.* **12**, 1357–1369 (2002).

54. Zhang, D. D. Functional evolution of a multigene family: orthologous and paralogous pheromone receptor genes in the turnip moth, *Agrotis segetum*. *PLoS One* **8**, 1002–1004 (2013).

55. Wanner, K. W. *et al*. Sex Pheromone Receptor Specificity in the European Corn Borer Moth, *Ostrinia nubilalis*. *PLoS One* **5**, e8685 (2010).

56. Krieger, J. *et al*. Genes encoding candidate pheromone receptors in a moth (Heliothis virescens). *Proc. Natl. Acad. Sci.* **101**, 11845–11850 (2004).

57. Bohbot, J. D. & Pitts, R. J. The narrowing olfactory landscape of insect odorant receptors. *Front. Ecol. Evol.* **3**, 39 (2015).

58. Pelletier, J., Hughes, D. T., Luetje, C. W. & Leal, W. S. An odorant receptor from the Southern house mosquito *Culex pipiens quinquefasciatus* sensitive to oviposition attractants. *PLoS One* **5**, 948 (2010).

59. Yan, S. W., Zhang, J., Liu, Y., Li, G. Q. & Wang, G. R. An olfactory receptor from *Apolygus lucorum* (Meyer-Dur) mainly tuned to volatiles from flowering host plants. *J. Insect Physiol.* **79**, 36–41 (2015).

60. Sanchez-Gracia, A., Vieira, F. G. & Rozas, J. Molecular evolution of the major chemosensory gene families in insects. *Heredity* **103**, 208–216 (2009).

61. Kent, L. B. & Robertson, H. M. Evolution of the sugar receptors in insects. *BMC Evol. Bio.* **9**, 41 (2009).

62. Agnihotri, A. R., Roy, A. A. & Joshi, R. S. Gustatory receptors in Lepidoptera: chemosensation and beyond. *Insect Mol. Biol.* **25**, 519–29 (2016).

63. Montell, C. A taste of the Drosophila gustatory receptors. *Curr. Opin. Neurobiol.* **19**, 345–353 (2009).

64. Min, S., Ai, M., Shin, S. A. & Suh, G. S. Dedicated olfactory neurons mediating attraction behavior to ammonia and amines in Drosophila. *Proc. Natl. Acad. Sci.* **110**, E1321–E1329 (2013).

65. Abuin, L. *et al*. Functional Architecture of Olfactory Ionotropic Glutamate Receptors. *Neuron* **69**, 44–60 (2011).

66. Knecht, Z. A. *et al*. Distinct combinations of variant ionotropic glutamate receptors mediate thermosensation and hygrosensation in Drosophila. *Elife* **22**, 5 (2016).

67. Li, Z., Ni, J. D., Huang, J. & Montell, C. Requirement for Drosophila SNMP1 for rapid activation and termination of pheromone-induced activity. *PLoS Genet.* **10**, e1004600 (2014).

68. Pelosi, P., Iovinella, I., Felicioli, A. & Dani, F. R. Soluble proteins of chemical communication: an overview across arthropods. *Front. Physiol.* **5**, 320 (2014).

69. Xu, P. X., Atkinson, R., Jones, D. & Smith, D. P. Drosophila OBP LUSH is required for activity of pheromone-sensitive neurons. *Neuron* **45**, 193–200 (2005).

70. Gomez-Diaz, C., Reina, J. H., Cambillau, C. & Benton, R. Ligands for pheromone-sensing neurons are not conformationally activated odorant binding proteins. *PLoS Biol.* **11**, e1001546 (2013).

71. Wu, Z., Lin, J., Zhang, H. & Zeng, X. BdorOBP83a-2 Mediates Responses of the Oriental Fruit Fly to Semiochemicals. *Front. Physiol.* **7**, 452 (2016).

72. Chang, H. *et al*. Pheromone binding proteins enhance the sensitivity of olfactory receptors to sex pheromones in *Chilo suppressalis*. *Sci. Rep.* **5**, 13093 (2015).

73. Zhang, Y. N., Ye, Z. F., Yang, K. & Dong, S. L. Antenna-predominant and male-biased CSP19 of *Sesamia inferensis* able to bind the female sex pheromones and host plant volatiles. *Gene* **536**, 279–86 (2014).

74. Yi, X. *et al*. BdorCSP2 Is Important for Antifeed and Oviposition-Deterring Activities Induced by Rhodojaponin-III against *Bactrocera dorsalis*. *PLoS One* **8**, e77295 (2013).

75. Zhang, T. T. *et al*. Functional Characteristics of a Novel Chemosensory Protein in the Cotton Bollworm *Helicoverpa armigera* (Hvbner). *J. Integr. Agr.* **12**, 853–861 (2013).

76. Liu, R. *et al*. Expression of chemosensory proteins in the tsetse fly *Glossina morsitans morsitans* is related to female host-seeking behaviour. *Insect Mol. Biol.* **21**, 41–48 (2012).

77. Quevillon, E. *et al*. InterProScan: protein domains identifier. *Nucleic. Acids. Res.* **33**, W116–120 (2005).

78. Katoh, K., Kuma, K., Toh, H. & Miyata, T. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic. Acids. Res.* **33**, 511–518 (2005).

79. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2–approximately maximum–likelihood trees for large alignments. *PLoS One* **5**, e9490 (2010).

80. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol. Biol. Evol.* **26**, 1641–1650 (2009).

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

82. Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA–Seq. *Nat. Methods* **5**, 621–628 (2008).

83. Robinson, M. D. & Oshlack, A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* **11**, R25 (2010).

84. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).

85. Toutges, M. J., Hartzer, K., Lord, J. & Oppert, B. Evaluation of reference genes for quantitative polymerase chain reaction across life cycle stages and tissue types of *Tribolium castaneum*. *J. Agric. Food Chem.* **58**, 8948–8951 (2010).

86. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real–time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* **25**, 402–408 (2001).

Acknowledgements

Author Contributions

J.-T.L. and Z.-Z.W. conceived the study and initiated this project. Z.-Z.W. performed bioinformatics. Z.-Z.W., M.-Q.Q. and X.-H.P. preformed the experiments. Z.-Z.W., M.-Q.Q. and S.-Y.B. analyzed and interpreted the data. S.-Y.B., Z.Z.W. and J.-T.L. wrote the manuscript. All authors read and approved the final manuscript.

Additional Information

Supplementary information accompanies this paper at doi:10.1038/s41598-017-11456-x

Competing Interests: The authors declare that they have no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.