



OPEN

Comparative genomic and transcriptomic analyses of chemosensory genes in the citrus fruit fly *Bactrocera (Tetradacus) minax*

Jun-Feng Cheng^{1,2,6}, Ting Yu^{2,6}, Zhong-Jian Chen^{2,6}, Shicheng Chen^{3,6}, Yu-Peng Chen^{1,6}, Lei Gao⁴, Wen-Hu Zhang², Bo Jiang⁵, Xue Bai¹, Edward D. Walker³, Jun Liu²✉ & Yong-Yue Lu¹✉

The citrus fruit fly *Bactrocera (Tetradacus) minax* is a major and devastating agricultural pest in Asian subtropical countries. Previous studies have shown that *B. minax* interacts with plant hosts via the efficient chemosensory system. However, the molecular components of the *B. minax* chemosensory system have not been well characterized. Herein, we identified a total of 25 putative odorant-binding receptors (OBPs), 4 single-copy chemosensory proteins (CSPs) and 53 candidate odorant receptors (ORs) using a newly generated whole-genome dataset for *B. minax*. This study significantly extended the chemosensation-related gene profiles (particularly, OBPs and ORs) in six other tephritid species. Comparative transcriptome analysis of adult *B. minax* and *Bactrocera dorsalis* showed that there were 14 highly expressed OBPs (FPKM > 100) in *B. dorsalis* and 7 highly expressed ones in *B. minax*. The expression level of *CSP3* gene and *CSP4* gene was higher in *B. dorsalis* than that in *B. minax*. Comparative genomic and transcriptomic analyses of chemosensory genes in the citrus fruit fly *B. minax* provided new insights for preventive control of this agriculture important pest and closely related species.

Bactrocera (Tetradacus) minax (Diptera: Tephritidae) is a destructive pest that damages primarily citrus tree fruits and diminishes fruit production¹. Once severe infestations occur in citrus-producing areas, severe economic losses typically follow¹. Therefore, new prevention and control methods specifically targeting *B. minax* are urgently needed.

Insects recognize their plant hosts mainly via perception of chemical signals emanating from the plants². Odorant-binding proteins (OBPs), chemosensory proteins (CSPs) and odorant receptors (ORs) are involved in insect chemosensation^{3–7}. The current model for insect chemosensation and recognition is as follows^{2,6–8}: some lipophilic odorant molecules in the environment reach the hydrophilic lymph of the insect through micropores on the olfactory sensilla surface and form a complex with the OBPs or CSPs in the sensillar lymph. Then, the complex passes through the sensillar lymph and binds to the ORs on the dendritic membranes. The membrane permeability changes when ORs are stimulated with the bindings. The above process results in the formation of an action potential and triggering cascade reactions, with the complex eventually entering the insect's central nervous system. Therefore, insects can sense exogenous odorant molecules and react accordingly with physiological and behavioral responses, such as foraging for food and oviposition sites.

The sizes of OBPs are approximately 15–17 kDa, consisting of 120–150 amino acids⁹. Specifically, the complete amino acid sequences of OBPs from *B. dorsalis* have a length ranging from 134 to 274 amino acids¹⁰. OBPs on

¹College of Agriculture, South China Agricultural University, Guangzhou, Guangdong, China. ²Agro-Biological Gene Research Center, Guangdong Academy of Agricultural Sciences, Guangzhou, Guangdong, China. ³Department of Microbiology and Molecular Genetics, Michigan State University, East Lansing, USA. ⁴Crop Research Institute, Guangdong Academy of Agricultural Sciences, Guangdong Provincial Key Laboratory of Crop Genetics and Improvement, Guangzhou, Guangdong, China. ⁵Fruit Tree Research Institute, Guangdong Academy of Agricultural Sciences, Guangzhou, Guangdong, China. ⁶These authors contributed equally: Jun-Feng Cheng, Ting Yu, Zhong-Jian Chen, Shicheng Chen and Yu-Peng Chen. ✉email: liujun@gdaas.cn; luyongyue@scau.edu.cn

the surface of insect antennae play an important role in binding odorants such as volatile substances⁷. Similar to OBPs, CSPs are also small, highly water-soluble, and acidic proteins with hydrophobic binding sites¹¹. CSPs contain about 120 amino acids (12–14 kDa). CSPs have been reported to participate in the chemosensory process by delivering hydrophobic sensory molecules to trigger neuronal responses in other insects (such as locusts, ants and *Bemisia tabaci*^{12–14}).

Insect ORs are a group of G-protein-coupled receptors with seven transmembrane domains. Different from those in vertebrate ORs⁵, the N terminal peptides of the ORs of insects are located within the cell while the C terminal domains are located outside. Insect ORs include both the conventional ORs and the atypical OR Or83b¹⁵. *Drosophila* olfactory conventional ORs and Or83b encode 370–400 amino acids and 486 amino acids, respectively¹⁶. The insect Or83b protein genes were coexpressed with conventional ORs in most olfactory neurons^{17,18}, which possibly impact on the olfactory behavior of insects¹⁹.

Application of sex pheromones to trap and kill adults is an important technique for managing *B. minax* infestation, which effectively reduces the number of pests and the rate of fruit damage¹. Thus, the identification of olfaction-related genes provides insights into the alternative strategies for specifically targeting these agriculturally important pests. However, the molecular components of olfaction system in *B. minax* have not been well studied. In this study, we focused on identification of the new chemosensory genes in *B. minax* by comparison to those from other tephritid species. Moreover, the expression patterns at three different developmental stages (larva, pupa and adult) were compared between *B. dorsalis* and *B. minax*.

Results

Identification and gene expression of OBPs in *B. minax*. A total of 25, 37, 33, 35, 30, 29, 34 OBP genes were identified in *B. minax*, *B. dorsalis*, *B. cucurbitae*, *B. latifrons*, *B. oleae*, *R. zephyria*, and *C. capitata*, respectively (Table 1). 37 of 52 OBP sequences in *D. melanogaster* showed high homology to those in the seven selected tephritid species (Table 1). Among them, the sequences of 8 OBPs (19a, 19d, 50e, 56a, 56d, 56h, 84a and 99c) were conserved (Table 1).

A phylogenetic tree revealed the OBPs' evolution relationships among the 7 tephritid species and *D. melanogaster* (Fig. 1). OBP19a showed two orthologs (OBP19a1 and OBP19a2) in Tephritidae, which were clustered with DmelOBP19a (Fig. 1). OBP19d1, d2 and d3 are pheromone binding protein related proteins (PBPRPs) in the tephritids. OBP19d1 and d2 were clustered with DmelOBP19d; instead, OBP19d3 formed a separate branch near them on the phylogenetic tree. The sequence of OBP50e showed good homology to DmelOBP50e. OBP50e proteins were encoded by multi-copy genes in *B. cucurbitae* (7 copies) and in *R. zephyria* (5 copies). OBP84a (classified as PBPRP4) had two orthologs in the tephritids, which formed a branch with DmelOBP84a (Fig. 1). Similarly, OBP99c with two orthologs in the tephritids processed the species-specific multiple copies in *B. dorsalis* (8) and *C. capitata* (6) (see Table 1 and Fig. 1).

OBPs can be classified into three types (Classic, Plus-C and Minus-C) based on the their conserved cysteines²⁰. Approximately 20 classic OBPs (with 6 conserved cysteines) were found in the tephritids, including 19a, 19b, 19c, 19d, 28a, 44a, 47a, 56b, 56c, 56d, 56e, 56g, 56h, 57c, 69a, 73a, 83a, 83b, 83g, 99a, 99b, and lush. Further, at least 12 Plus-C OBPs (more than 6 conserved cysteine residues) were identified in the tephritids, which were clustered in two clades on the phylogenetic tree: one cluster contained eight OBPs (46a, 47b, 50a, 50c, 50e, 56a, 58c, 58d) and another one had four (59a, 83cd, 83ef, 84a). Three Minus-C OBPs (less than 6 cysteines) were identified in the tephritids (8a, 99c, 99d), which formed an independent cluster on the phylogenetic tree. The above observations suggested that these OBPs possibly evolved in different routes in the tephritids.

In the larval stage, both *B. minax* and *B. dorsalis* had six OBP genes showing a high transcription level (i.e., 99c1, 56d, 50c, 83g, 19a2 and 99b in *B. minax* as well as 99c1, 56d3, 99c5, 44a, 99b and 56d2 in *B. dorsalis*, respectively). Four (99c1, 99b, 50c and 56d) and five (99c1, 99b, 44a, 50c and 83g) OBP genes with high expression levels were found in *B. minax* and *B. dorsalis* in the pupal stage, respectively. At least seven (99c2, 99b, 99c1, 19d2, 83g, 28a, and 19d1) and fourteen (99c1, 44a, 99b, 56d1, 56d3, 56d2, 50c, 83a, 83g, 19d1, 19d2, 83b, 56g, 28a) OBP genes demonstrated high transcription levels in adult *B. minax* and *B. dorsalis*, respectively (Fig. 2). Notably, OBP99c1 was expressed at the highest levels in all developmental stages (with the exception of adult *B. minax*) for both species. The expression level of OBP99c1 peaked (FPKM = 12,091) in the pupal stage of *B. dorsalis*, suggesting its important role in the metamorphosis and/or detoxification. It is very interesting that the transcriptional level of OBP99c2 was higher (FPKM = 3823) in *B. minax* adults while it was relatively lower (FPKM = 6) in *B. dorsalis* adults, indicating that the regulation of OBP99c2 was species-specific. However, more studies are warranted to elucidate its physiological functions.

Identification and gene expression of CSPs in *B. minax*. There were at least 4 genes encoding the CSP family proteins (CSP1, CSP2, CSP3, and CSP4) among all tested species in tephritids. The CSPs in *B. dorsalis* were closely related to those in *B. latifrons*, followed by *B. oleae*, *B. minax*, *B. cucurbitae*, *C. capitata*, *R. zephyria* and *D. melanogaster* (Fig. 3). There were four conserved cysteines in CSP1–4. The interspecies sequence similarity of CSPs among the tephritids was higher than that in *D. melanogaster*. All CSP3 genes of the 7 tephritid species had two splice variants, namely, CSP3X1 and CSP3X2. The average identities of CSP1, CSP2, CSP3X1, CSP3X2 and CSP4 among the tephritids were 91%, 91%, 82%, 80%, and 89%, respectively (Table 2; Supplementary file 7). However, the average degrees of similarity of CSP1, CSP2, CSP3X1, CSP3X2, and CSP4 between the tephritids and *D. melanogaster* were 64%, 73%, 50%, 28%, and 66%, respectively (Table 2). The gene encoding CSP1 was expressed at the highest transcriptional level in the pupae stage for both *B. minax* and *B. dorsalis*. However, the gene encoding CSP2-4 showed a higher transcriptional level than CSP1 in the adult stages. Further, the transcript level of CSP3 and CSP4 was lower in adult *B. minax* than that in *B. dorsalis* (Fig. 4).

OBPs	Tephritidae									Total
	Dmel	Bmin	Bdor	Bcuc	Blat	Bole	Rzep	Ccap	Tephritidae	
8a	1	1	1	1	1	1	1	1	+	8
18a	1	0	0	0	0	0	0	0	–	1
19a	1	2	2	1	2	2	2	2	+	14
19b	1	1	1	1	1	1	1	1	+	8
19c	1	1	1	1	1	1	1	1	+	8
19d	1	3	3	3	2	3	3	3	+	21
22a	1	0	0	0	0	0	0	0	–	1
28a	1	1	1	1	1	1	1	1	+	8
44a	1	0	1	1	1	1	1	1	+	7
46a	1	0	1	0	0	0	1	1	+	4
47a	1	1	1	1	1	1	0	1	+	7
47b	1	0	0	1	1	1	0	1	+	5
49a	1	0	0	0	0	0	0	0	–	1
50a	1	0	1	1	1	0	0	0	+	4
50b	1	0	0	0	0	0	0	0	–	1
50c	1	1	1	1	1	0	2	1	+	8
50d	1	0	0	0	0	0	0	0	–	1
50e	1	0	1	7	2	2	5	2	+	20
51a	1	0	0	0	0	0	0	0	–	1
56a	1	1	1	1	2	2	3	2	+	13
56b	1	1	1	1	1	1	1	1	+	8
56c	1	1	1	1	1	1	1	1	+	8
56d	1	1	3	1	2	1	3	1	+	13
56e	1	0	1	1	1	1	0	1	+	6
56f	1	0	0	0	0	0	0	0	–	1
56g	1	1	1	1	1	0	2	2	+	9
56h	1	2	3	3	3	2	2	2	+	18
56i	1	0	0	0	0	0	0	0	–	1
57a	1	0	0	0	0	0	0	0	–	1
57b	1	0	0	0	0	0	0	0	–	1
57c	1	0	1	1	1	1	1	1	+	7
57d	1	0	0	0	0	0	0	0	–	1
57e	1	0	0	0	0	0	0	0	–	1
58b	1	0	0	0	0	0	0	0	–	1
58c	1	1	1	1	1	0	0	1	+	6
58d	1	0	1	1	1	0	0	1	+	5
59a	1	0	1	0	1	0	0	1	+	4
69a	1	0	1	0	0	1	1	1	+	5
73a	1	1	1	1	1	1	1	1	+	8
83a	1	1	1	1	1	1	1	1	+	8
83b	1	1	1	1	1	1	2	1	+	9
83cd	1	1	1	0	1	1	1	1	+	7
83ef	1	0	1	1	1	1	1	0	+	6
83g	1	1	1	1	1	1	0	1	+	7
84a	1	1	2	2	2	2	2	2	+	14
85a	1	0	0	0	0	0	0	0	–	1
93a	1	0	0	0	0	0	0	0	–	1
99a	1	0	1	1	1	1	1	0	+	6
99b	1	1	1	1	1	1	1	1	+	8
99c	1	2	8	2	4	3	1	6	+	27
99d	1	1	1	1	1	1	1	1	+	8
lush	1	1	1	1	1	1	1	1	+	8
Sequences	52	30	51	45	46	39	45	47		355
Families	52	25	37	33	35	30	29	34	37	

Table 1. Identification of OBP genes in Tephritidae*. *The abbreviation used for the species are: *B. minax* (Bmin), *B. dorsalis* (Bdor), *B. cucurbitae* (Bcuc), *B. latifrons* (Blat), *B. oleae* (Bole), *C. capitata* (Ccap), *R. zephyria* (Rzep), *D. melanogaster* (Dmel).

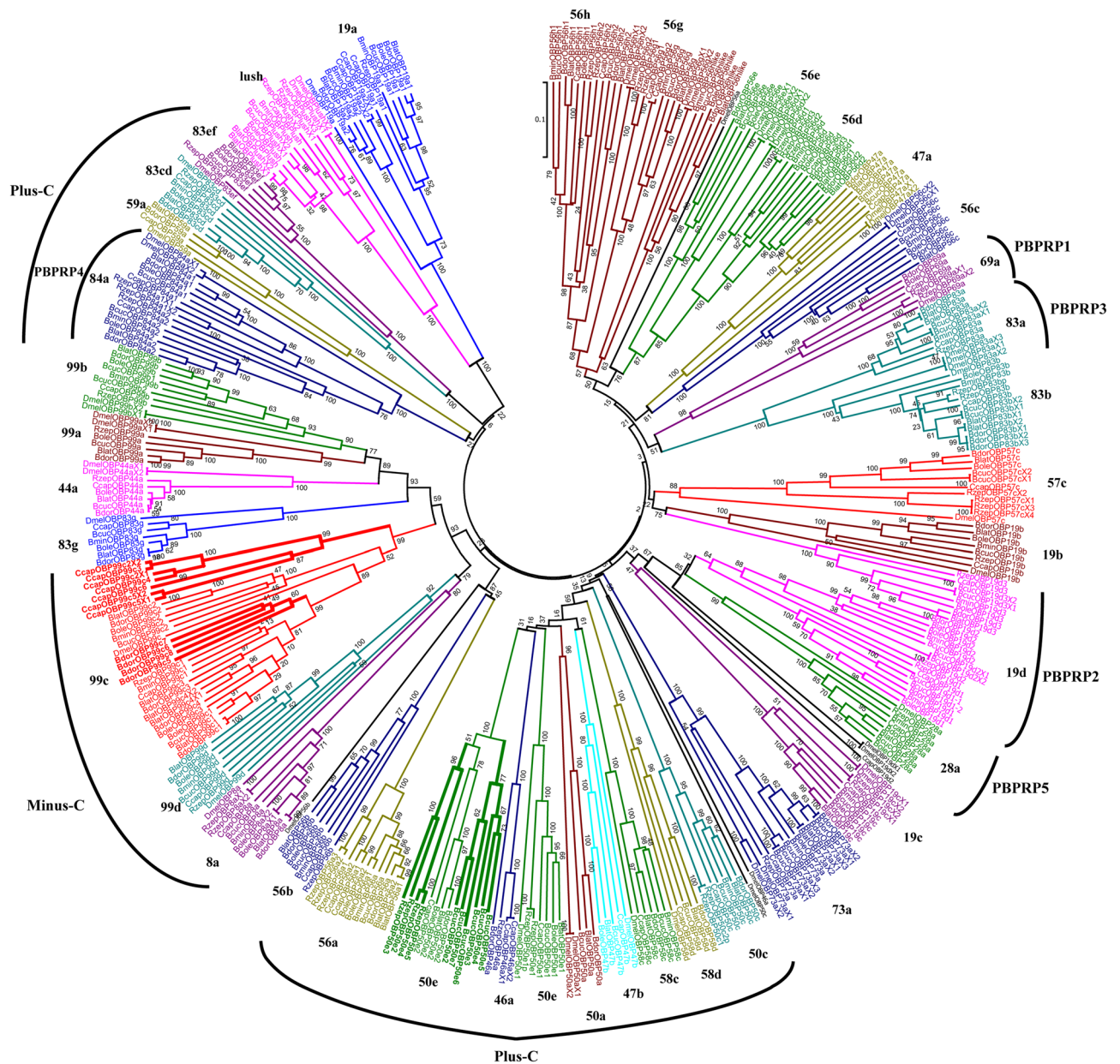


Figure 1. Phylogenetic relationships of OBP proteins in the selected Tephritid (NJ method). Bootstrap values greater than 50% (1000 replications) were displayed. The numbers of OBP genes present in *D. melanogaster* were previously reported⁴⁵.

Identification and gene expression of ORs in *B. minax*. To search for candidate OR genes in *B. minax*, the sequence similarity analysis and the phylogenetic tree construction were performed. As shown in Table 3, 53, 70, 61, 58, 59, 64, and 61 OR genes were identified in *B. minax*, *B. dorsalis*, *B. cucurbitae*, *B. latifrons*, *B. oleae*, *R. zephyria*, and *C. capitata*, respectively. 35 of 62 ORs from *D. melanogaster* were present in these tephritids (Table 3). 14 ORs (7a, 33abc, 45a, 59a, 63a, 67d, 69aA/B, 74a, 83a, 85bc, 85d, 85e, 88a and 94ab) showed gene duplications in the tephritids (Table 3). For example, OR85e and OR88a had 3 copies in *Bcuc* and in *Bole*, respectively (Fig. 5). In addition, nine OR genes specific to the tephritids were identified, which were individually named OR1, OR2, OR3, OR4, OR5, OR6, OR7, OR8 and OR9. Except OR9, these ORs all presented gene duplications; for example, OR3 had 8 copies in *B. dorsalis*.

The copy numbers of OR families were much higher than those of OBP and CSP families⁷. We further identified 21 putative ORs in the tephritids with their ancestral nodes (here named as 2a, 7a, 10a, 13a, 22c, 24a, 33abc, 43a, 45a, 46a, 47b, 49a, 49b, 59a, 67c, 67d, 69a, 71a, 82a, 85e, and 88a). Six genes in the tephritids were identified with two orthologs, which were distributed in two distinct clades together with *D. melanogaster* OR genes. This observation suggested that these ORs maybe have appeared earlier in the tephritids than those in *D. melanogaster* (63a, 74a, 83a, 85bc, 85d and 94ab). Nine OR genes (OR1-9) were conserved in *D. melanogaster* and in the tephritids. These results showed that the OR gene family possibly underwent rapid evolution with a

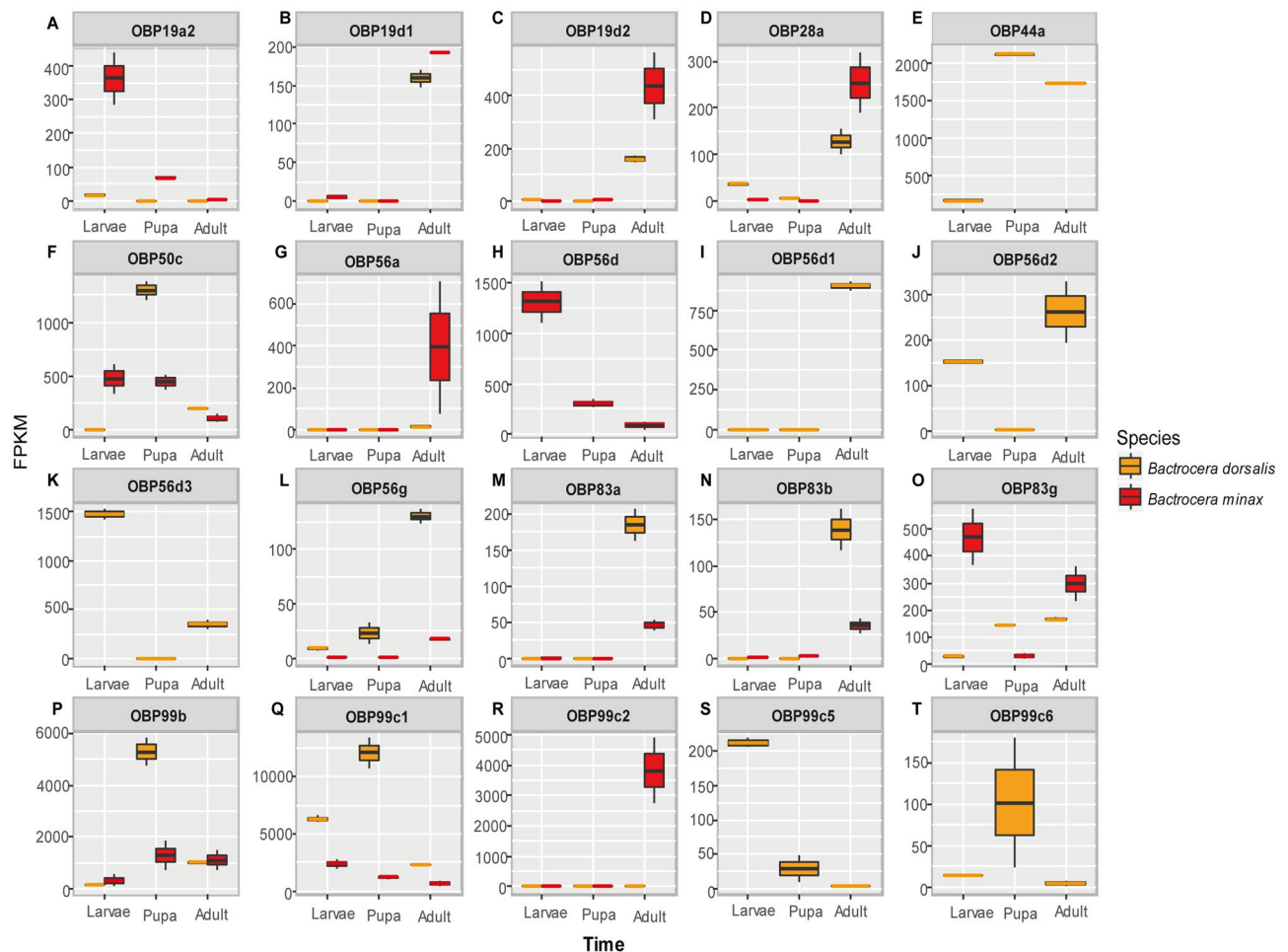


Figure 2. Highly expressed OBPs in *B. minax* and *B. dorsalis*. Twenty selected OBPs were shown in Fig A-T. These OBPs were chosen because of their highly expressed level (FPKM ≥ 100) in *B. minax* or *B. dorsalis*.

large amount of variations. The exception of the OR46a1 gene exhibited moderate expression between FPKM 8.1 and 13.4 in the larval stage of *B. minax*. However, most OR genes showed a relative lower level at the three developmental stages in both species (Supplementary Table S5).

Validation of RNA-Seq data by qRT-PCR analysis. The expression of five selected genes including *OBP99c1*, *OBP99c2*, *OBPlush*, *CSP1*, and *ORCO* were validated using qRT-PCR from *B. minax* and *B. dorsalis* in the larvae and adult stages (Fig. 6). Coefficient analysis showed that qRT-PCR data of the 5 selected genes were significantly correlated with the RNA-Seq results ($r=0.91$ in *B. minax* and $r=0.95$ in *B. dorsalis*; Supplementary file 8), which indicated that the RNA-Seq data in the present study were reliable and could support the transcriptomic analysis presented above. For example, the results from qRT-PCR also showed that the transcriptional level of *OBP99c1* was high at the larvae stage but relatively low in the adult stage in *B. minax*. *OBP99c2* was highly expressed in *B. minax* adults, indicating that *OBP99c2* may be an important gene for sensing chemicals in adult *B. minax*.

Discussion

Insects recognize their hosts mainly via the insect chemosensory systems (i.e. OBPs, CSPs and ORs)⁶. For example, *B. minax* Enderlein has a restricted host range and almost always lays eggs in *Citrus* fruits¹; however, *B. dorsalis* Hendel infests over 200 different fruits and vegetables²¹. Identification and characterization of the insect chemosensory system will contribute to the development of novel biocontrol methods for targeting the specific agricultural pests such as *B. minax*. Wu et al.¹⁰ investigated the transcriptome profiles in *B. dorsalis*. At least 31 OBPs, 4 CSPs and 23 ORs were found in *B. dorsalis*¹⁰. However, Liu et al. (2016) reported that there were 20 OBPs, 5 CSPs and 35 ORs in male and female adults of *B. dorsalis*¹³. Both studies likely underestimate the insect chemosensory systems due to the limitation of the RNA-seq technology (transcriptome coverage, low abundance of the genes, or the chosen studied sites).

We first identified up to 25 OBP, 4 CSP and 53 OR genes in the genome of *B. minax* using comparative genomic approaches. Several chemosensory proteins identified in this study were not previously reported. Remarkably, we discovered at least 37 OBP, 4 CSP, and 70 OR genes in the genome of *B. dorsalis*. Moreover, the

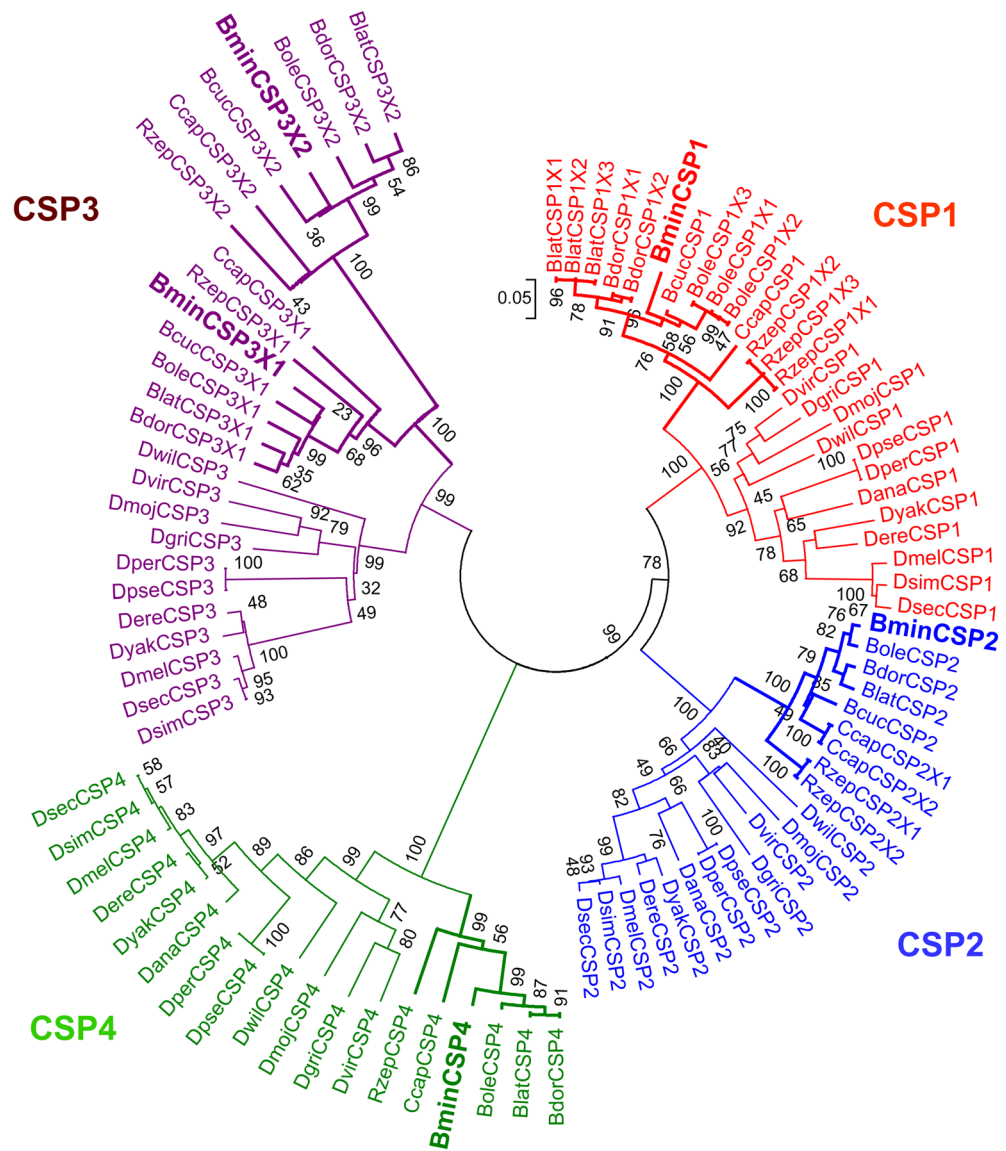


Figure 3. Neighbor joining (NJ) phylogenetic tree of the CSP family. The abbreviation used for the species are: *B. minax* (Bmin), *B. dorsalis* (Bdor), *B. cucurbitae* (Bcuc), *B. latifrons* (Blat), *B. oleae* (Bole), *C. capitata* (Ccap), *R. zephyria* (Rzep), *D. melanogaster* (Dmel), *D. simulans* (Dsim), *D. sechellia* (Dsec), *D. erecta* (Dere), *D. yakuba* (Dyak), *D. ananassae* (Dana), *D. pseudoobscura* (Dpse), *D. persimilis* (Dper), *D. willistoni* (Dwil), *D. mojavensis* (Dmoj), *D. virilis* (Dvir) and *D. grimshawi* (Dgri). The number of CSP genes present in *Drosophila* has been previously reported⁶.

	CSP1 (%)	CSP2 (%)	CSP3X1 (%)	CSP3X2 (%)	CSP4 (%)
Intra-Tephritidae	91	91	82	80	89
Tephritidae and <i>D. melanogaster</i>	64	73	50	28	66

Table 2. Sequence similarity of CSPs in Tephritidae.

comparative genomic analyses allowed us to identify at least 33 OBPs, 4 CSPs and 61 ORs in the genome of *B. cucurbitae* (Supplementary files 3–6). Only 17 OBPs were previously reported in the Mediterranean fruit fly, *Ceratitis capitata*, from the EST libraries of adult heads, embryos, male accessory glands and testes²². Instead, we found 34 OBP, 4 CSP and 61 OR genes in the genome of *C. capitata* (Supplementary files 3–6). Collectively, our discoveries significantly extend the chemosensing-related gene profiles (particularly, OBPs and ORs) in these agriculture important pests.

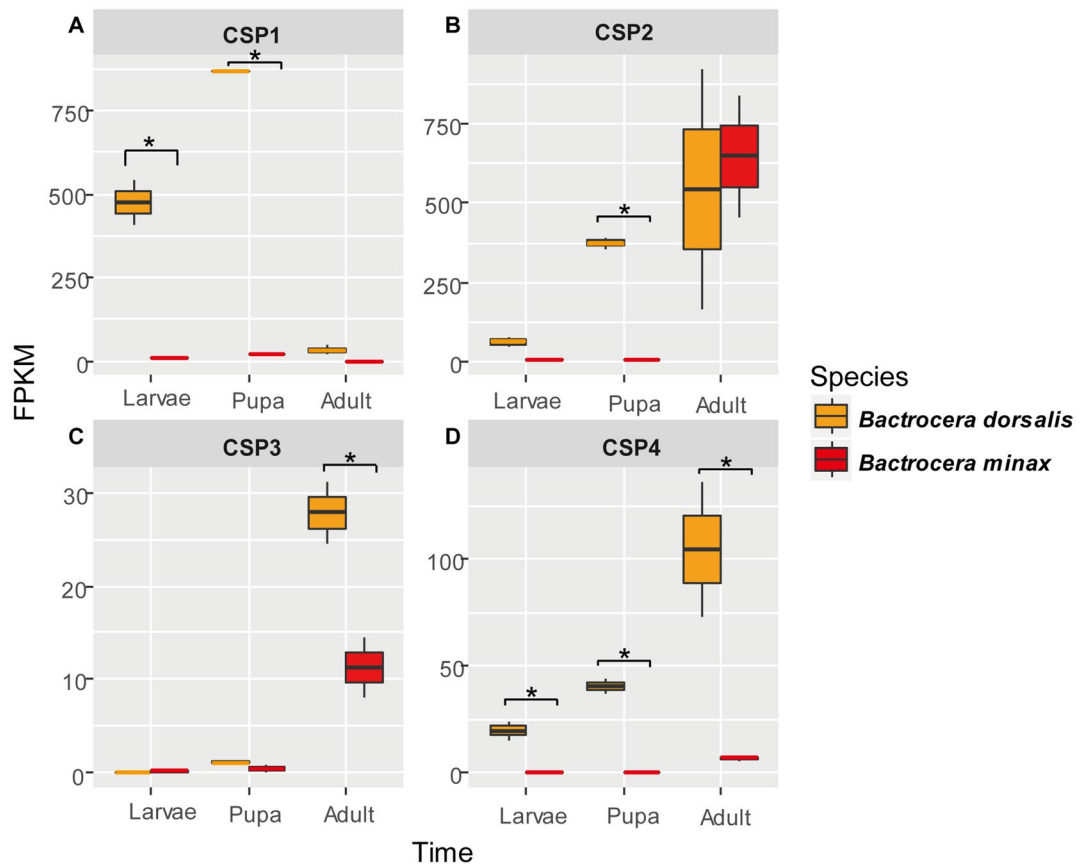


Figure 4. Gene expression of CSPs in *B. minax* and *B. dorsalis*. The transcriptional expression values of CSP1–4 at the three developmental stages in *B. minax* and *B. dorsalis* were shown in (A–D).

Compared to those in *B. dorsalis*, the less OBPs and ORs in *B. minax* possibly related to its narrower plant host ranges²³. However, this hypothesis needs to be further tested within the context of the generalized or specific binding properties of OBPs⁷. At the adult developmental stage, 14 OBP genes were highly expressed (FPKM > 100) in *B. dorsalis* while only 7 OBP genes were highly expressed in *B. minax*. *B. dorsalis* may utilize more OBPs to perform the odorant and pheromone binding functions than those in *B. minax*. Among them, it is worthy of highlighting that gene *OBP99c1* was highly expressed in all the developmental stages in both species. However, *OBP99c1* exhibited a gradually declining trend during *B. minax* development, suggesting that it may play a crucial role in the odorant-binding in the citrus fruit fly. The gene *OBP99c1* in *B. dorsalis* (named *BdorOBP10*) was highly expressed in females, especially in the abdomen, where the reproductive organs are located²⁴. Interestingly, the OBP10 protein of two sibling Lepidopteran species, *Helicoverpa armigera* and *Helicoverpa assulta*, was detected first in the male reproductive system, then in females during mating, and eventually in eggs²⁵. In addition, OBP10 exhibited binding to an insect repellent, indicating that this protein may be a carrier for some semiochemicals²⁵. Therefore, *OBP99c1* may be a good target for the design of novel pest control agents. On the other hand, *OBP99c2* was found to be highly expressed in *B. minax* adults, but it was barely expressed in *B. dorsalis*. *OBP99c2* may play a species-specific role in the development of odorant-binding function. However, further study is warranted to elucidate the physiological functions of *OBP99c2*.

The number of conventional ORs varies dramatically among different insect species. For example, the genome of honey bee *Apis mellifera* contains 170 OR genes. However, *D. melanogaster* and *Anopheles gambiae* only carry 62 and 79 OR genes, respectively²⁶. Our study showed that the numbers of ORs in 6 selected Tephritidae ranged from 53 to 70. Unlike the relatively high expression levels of OBPs and CSPs in *B. minax* and *B. dorsalis*, the expression level in OR genes was considerably low during the development progress, which was consistent with the previous observations¹⁰. However, Liu et al. (2016) reported that two OR genes (*Bdoror 13* and *Bdoror 14*) in *B. dorsalis* were specifically expressed at a high level in the male antennae²⁷. The discrepancy may be caused by the different selection of study sites (whole body vs tissues) or different investigation approaches (qPCR vs RNAseq)¹⁰.

The protein sequence similarity among CSPs among the selected insects was high. For example, CSPs of *Schistocerca gregaria* and *Locusta migratoria* share 50–60% sequence identity²⁸. We found that the sequence similarities of CSPs among Tephritidae were higher than those between Tephritidae and *D. melanogaster*. CSPs were known to play a key role in olfactory perception. Therefore, CSPs have been used to screen potential bioactive compounds for pest management²⁹. Notably, insecticides have been shown to significantly up-regulated adult-specific CSP1 gene expression in the sweet potato whitefly *Bemisia tabaci*¹³ (Liu et al., 2016). Additionally,

	Dmel	Bmin	Bdor	Bcuc	Blat	Bole	Rzep	Ccap	Total
Nonclassical olfactory receptor									
83b (ORCO)	1	1	1	1	1	1	1	1	8
Typical olfactory receptor ORs									
2a	1	1	1	1	1	1	1	1	8
7a	1	2	3	3	3	3	6	4	25
10a	1	1	1	1	1	1	1	1	8
13a	1	0	1	1	1	1	1	1	7
22c	1	1	1	1	1	1	1	1	8
24a	1	1	1	1	1	1	1	1	8
33abc	3	3	1	4	0	2	5	2	20
35a	1	1	1	1	1	1	1	1	8
43a	1	1	1	1	1	1	1	1	8
45a	1	2	3	2	2	3	3	3	19
46aA/B	2	1	1	0	1	1	1	1	8
47b	1	1	1	1	1	1	1	1	6
49a	1	1	1	1	1	1	1	1	8
49b	1	0	1	1	1	1	1	1	7
59a	1	2	2	3	2	2	1	1	14
63a	1	3	4	1	3	4	2	2	20
67c	1	1	1	1	1	1	1	1	8
67d	1	3	2	2	2	2	4	2	18
69aA/B	2	0	2	2	1	2	2	2	13
71a	1	1	1	1	1	1	2	1	9
74a	1	2	2	2	2	2	2	2	15
82a	1	1	1	1	1	1	1	1	8
83a	1	2	2	3	1	1	1	2	14
85bc	2	1	2	1	2	2	1	2	13
85d	1	2	2	2	2	1	2	0	12
85e	1	1	1	3	1	1	1	1	12
88a	1	1	1	1	1	3	1	1	10
94a/b	2	2	3	1	3	3	1	2	17
Common Diptera ORs	35	39	45	44	40	46	48	41	339
Tephritidae-specific ORs									
OR1	0	1	2	2	2	1	0	2	10
OR2	0	1	2	3	2	2	2	1	13
OR3	0	2	8	3	3	3	2	3	24
OR4	0	1	2	2	2	1	1	3	12
OR5	0	2	1	1	1	1	2	1	9
OR6	0	1	2	2	2	2	7	3	19
OR7	0	1	1	1	1	0	1	2	7
OR8	0	4	6	2	4	2	0	4	22
OR9	0	1	1	1	1	1	1	1	7
Total Diptera ORs	62	53	70	61	58	59	64	61	426

Table 3. Identification of OR genes in Tephritidae*. *The abbreviation used for the species are: *B. minax* (Bmin), *B. dorsalis* (Bdor), *B. cucurbitae* (Bcuc), *B. latifrons* (Blat), *B. oleae* (Bole), *C. capitata* (Ccap), *R. zephyria* (Rzep), *D. melanogaster* (Dmel).

in whitefly, due to the ligand binding specificity, CSP1 may be responsible for regulation of the insect immune response mediated by fatty acids, while CSP2 and CSP3 facilitate insect communication with the surrounding environment via favorable or unpleasant odors¹³. A recent study indicated that, in *Bradysia odoriphaga* (Diptera: Sciaridae), tissue-specific enrichment of CSP4 (in both the antennae and heads) and CSP1/CSP2 (in the legs and heads) may be involved in other crucial physiological functions of this insect³⁰. In fact, the honeybee (*Apis mellifera*) exhibited abnormal head development upon loss of CSP5 function and could not turn into the larval stage³¹. Therefore, further investigation of the tissue distribution of the CSPs identified in this study may facilitate the functional analysis of these genes.

In conclusion, we identified the chemosensing-related genes of the citrus fruit fly *B. minax* based on genome data and identified 82 candidate chemosensing-related genes, including 25 OBPs, 4 CSPs and 53 ORs. Our

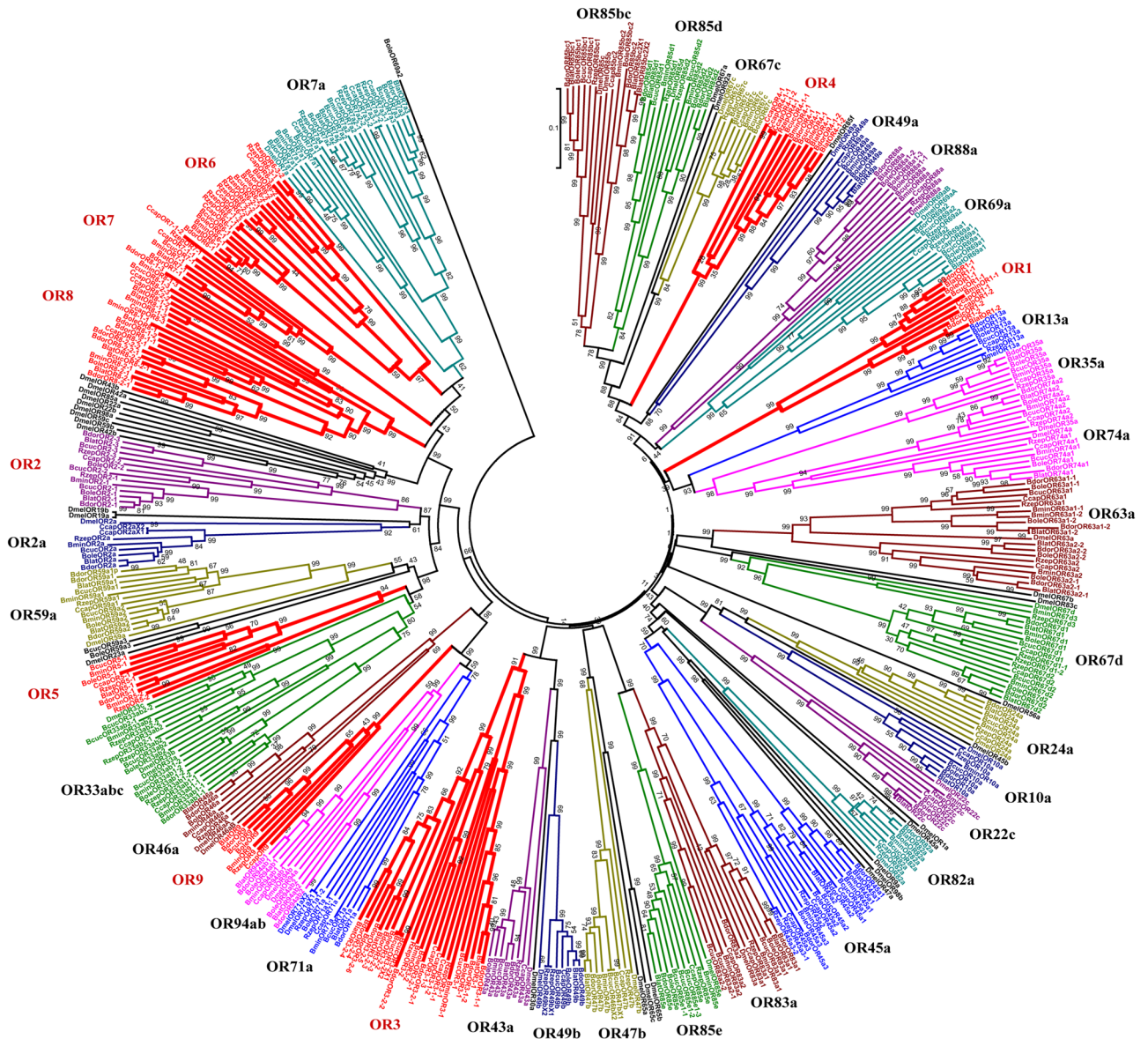


Figure 5. NJ tree of the OR family in Tephritidae. NOTE: The number of OR genes present in *D. melanogaster* has been previously reported²⁰.

study compared the genetic relationships of candidate genes among 7 species and showed that *B. minax* had the least numbers of OBPs and ORs. Based on the transcriptomes of three developmental stages (larvae, pupae and adults) of *B. minax* and *B. dorsalis*, the expression profiles of candidate OBPs, CSPs and ORs were compared and demonstrated that almost all the OBPs and the CSPs presented more highly transcriptional expression values in *B. dorsalis* than in *B. minax*. These findings suggested that *B. dorsalis* may exhibit more powerful odorant and pheromone binding properties than *B. minax*, which could be related to the fact that *B. dorsalis* targets a wider range of host species than *B. minax*.

Materials and methods

Insect rearing and ethics statement. *B. minax.* *B. minax* insects were collected and identified by personnel of the Hunan Academy of Agricultural Sciences in Jishou City, Hunan Province, China. Larvae were reared in tangerines and pupae were buried in wet soil. Larvae were collected at the 3rd instar, and pupae at 20 days after pupation. The adults were collected 12 h after emergence without any feeding or sexually matured.

B. dorsalis. *B. dorsalis* insects were reared in the laboratory at 25 ± 1 °C under a 16:8 h light:dark photoperiod and 70–80% relative humidity (RH). Artificial diets for the larvae were provided by the Institute of Insect Ecology of South China Agricultural University, and consisted of banana, corn flour, sucrose, yeast extract, paper, sodium benzoate, hydrochloric acid and water in appropriate proportions^{32,33}. Larvae developed into pupae in wet sand. Artificial diet for adults was a mixture (1:1) of yeast extract and sucrose³⁴. Larvae were collected 6 days

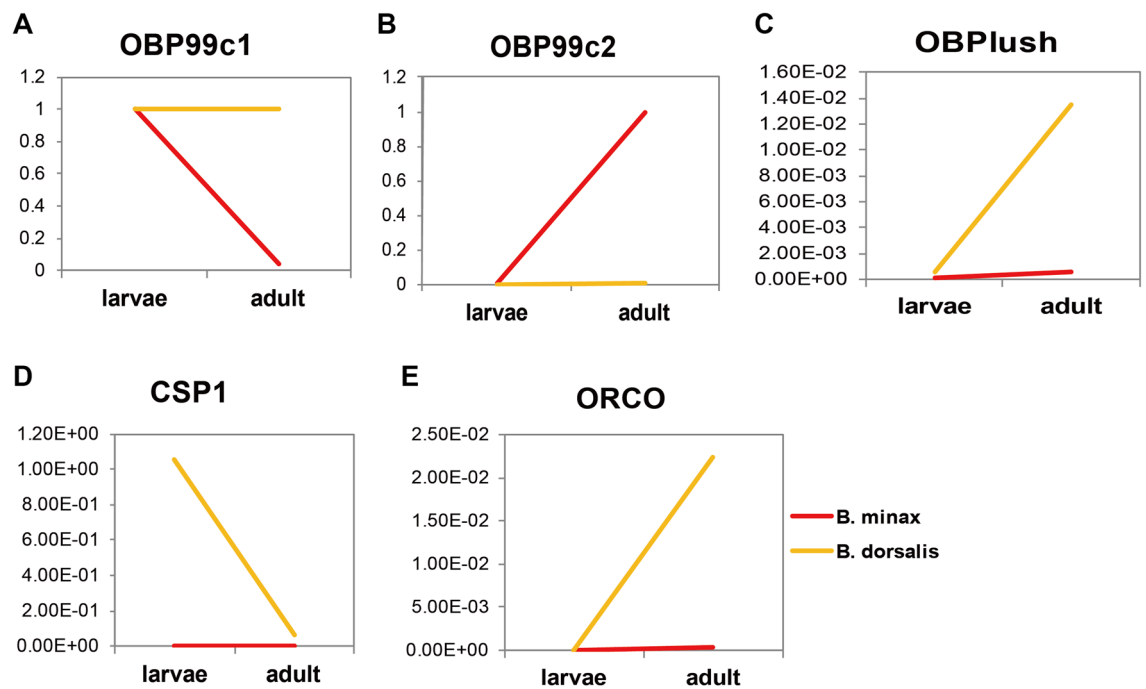


Figure 6. The qRT-PCR verification of gene expression. Five genes, including OBP99c1, OBP99c2, OBPlush, CSP1 and ORCO, were selected and their gene expressions were validated using qRT-PCR from *B. minax* and *B. dorsalis* in the larvae and adult stages.

after incubation at the 3rd instar. Pupae were collected 6 days after pupation. Adults were collected 12 h after emergence without any feeding and before sexual maturation.

Sample preparation and genome sequencing. Body parts of single *B. minax* males were dissected to obtain DNA for whole-genome sequencing. Total DNA was extracted using the Magen HiPure Insect DNA Kits (D3129-02) (Guangzhou, China) according to the manufacturer's instructions. Samples were sent to the GeneDenovo company (Guangzhou, China) (<https://www.genedenovo.com/>) for library construction and genome sequencing. Genome sequencing was carried out with the combination of Next-generation sequencing (NGS) using the HiSeq 2500 and third-generation sequencing adopting the Pacbio RSII. Briefly, five libraries were constructed, making two duplicated short fragment libraries (450 bp + 800 bp) and three long mate-pair libraries (2 kb + 5 kb + 10 kb), which produced in total 102G nucleotide bases (Supplementary file 1, Table S1-1) and covered an estimated 300× of the genome size (Supplementary file 1, Figure S1-1). The genome size was estimated at about 331 Mb based on Kmer (k=17) short fragmentary libraries analysis (Supplementary file 1, Figure S1-2). For third-generation sequencing, two libraries were constructed, five SMRT cells were sequenced, total 5G raw data were obtained, and the genome coverage reached 15× (Supplementary file 1, Table S1-2); and SMRT analysis software (version 2.3.0) (<https://www.pacb.com/products-and-services/analytical-software/smrt-analysis/>) provided from Pacbio was used for the sequencing quality control. The genome assembly was divided into two steps: (1) Platanus (version 1.2.1)³⁵ was used to assemble Illumina data and GapCloser (v1.10)³⁶ was utilized to extend the contig length; (2) PBjelly (PBSuite_15.8.24)³⁷ was used to extend the scaffolds and fill the gaps by combining the third-generation sequencing data. Totally, we obtained a genome size of 340 Mb, which is close to the estimated genome size. The numbers of contigs/scaffolds were 38,509/8019 and the contig/scaffold N50 reached 23 kb/1.6 Mb (Supplementary file 1, Table S1-3). The genome sequences have been submitted to NCBI (SAYV000000000). The repetitive sequences account for 20.97% of the genome with using de novo prediction (RepeatModeler³⁸ and LTR-FINDER³⁹), RepBase⁴⁰-based homology prediction (RepeatMasker⁴¹ and RepeatProteinMask⁴²), and tandem repeats finder (TRF⁴³) (Supplementary file 1, Table S1-4 and Table S1-5). Several methods of de novo prediction, homology-based gene prediction, and cDNA/EST prediction were used to predict gene structure after excluding the repetitive sequences (Supplementary file 1, Figure S1-3). Softwares from three systems (Augustus 2.7⁴⁴, Genscan 1.0⁴⁵, and Glimmer HMM 3.0.1⁴⁶) were used for de novo prediction of gene models. MAKER 2.28 software⁴⁷ was applied to integrate all gene sets into a non-redundant and more complete one. Lastly, the BUSCO method⁴⁸ was used to estimate the reliable degree of gene models, and the complete single-copy BUSCO scores reached 98.5% in the assembled gene set and 94.9% in the annotated gene set, respectively (Supplementary file 1, Table S1-6). After all these analyses, a set of 12,533 gene models was obtained and used for identification of chemosensory genes.

Identification of chemosensory genes. The protein sequences from *B. dorsalis* (Bdor), *Bactrocera cucurbitae* (Bcuc), *Bactrocera latifrons* (Blat), *Bactrocera oleae* (Bole), *Ceratitidis capitata* (Ccap), and *Rhagoletis zephyria* (Rzep), all Diptera, family Tephritidae, were downloaded from NCBI (<ftp://ftp.ncbi.nlm.nih.gov/>)

genome). Protein sequences from the above species that were annotated as chemosensation-related genes were extracted. The protein sequences from *B. minax* were aligned to the protein sequences annotated as chemosensation-related genes from other 6 species by using BLASTP with an e-value setting of 1e-5. From this analysis, the possible chemosensation-related genes of *B. minax* were obtained. The sequences of fifty-one OBP genes and one lush gene from *Drosophila melanogaster*⁴⁹ and those of 31 OBP genes from *B. dorsalis* were obtained. The sequences of forty-seven CSP genes from 12 *Drosophila* species⁶ and those of four CSPs (CSP1-4) from *B. dorsalis* were acquired. The sequences of sixty-two OR genes from *D. melanogaster*¹⁷ and those of 23 OR genes from *B. dorsalis*¹² were obtained. Then the potential chemosensation-related genes of *B. minax* and the protein sequences annotated as chemosensation-related genes from the other 6 tephritid species were aligned to the collected gene sets (83 OBPs, 51 CSPs and 85 ORs) by using BLASTP with an e-value setting of 1e-5. According to sequence similarity of more than 40% and manual inspection: (1) multiple consistent hits to directly classify this one group; (2) multiple inconsistent hits to take the one with high similarity, the potential or annotated chemosensation-related genes were assigned to each known chemosensory gene. Other genes that were difficult to assign were classified as indeterminate genes, including the genes annotated as chemosensation-related genes but their sequence similarity lower than 40% to the collected gene sets.

Sequence comparison and phylogenetic analysis. The tool ClustalX⁵⁰ was used to execute multiple sequence alignment with default parameters, and GeneDoc⁵¹ was used to visualize the alignment. Highly divergent sequences which by visual inspection had no common sites with others were moved to the indeterminate gene category. Then, the sequence alignment was input into MEGA6⁵²; the NJ (neighbor-joining) method with the pairwise deletion option and the maximum likelihood method with partial deletion option were selected to construct phylogenetic trees (to see supplementary file 2). Robustness of the branches was assessed with 1000 bootstrap pseudo-replicates. In this manner, the determinate chemosensory genes in the 7 species were obtained. Lastly, all identified chemosensory genes and the indeterminate genes were combined to carry out sequence comparison and phylogenetic analysis. Some highly divergent or short sequences, reported by MEGA6 to affect the phylogenetic tree construction, were deleted. The gene names for *B. minax* and the six other species were modified according to the phylogenetic clusters.

Sequence information. The OBP, CSP and OR family sequences from *B. minax* have been submitted to GenBank with accession numbers MH937211–MH937240, MH937207–MH937210, MH937241–MH937290. The NCBI reference sequence and GenBank accessions of OBPs, CSPs and ORs in all 7 tephritid species are provided in the supplementary file 3. Protein sequences and alignments of each OBP/CSP/OR member with those from other species are presented in supplementary files 4, 5 and 6, respectively.

RNA sequencing and transcript analysis. Total RNA was extracted from two *B. minax* individuals and four *B. dorsalis* individuals at the development of larvae, pupae and adult using TRIzol reagent (Invitrogen, California, USA). The adult samples were equal numbers of females and males. Two biological replicates per developmental stage were used to construct the cDNA libraries using the Illumina TruSeq RNA Sample Preparation kit and the mixed libraries were sequenced on one lane of Illumina HiSeq 4000 platform with paired-end 150-bp reads. Raw paired-end reads from all samples were submitted to the Sequence Read Archive of the NCBI (*B. minax*: SRP193917 and *B. dorsalis*: SRP193924). Clean reads were obtained by Perl scripts: (1) the adapters were removed and the reads, which were shorter than 50 nt and including adapters, were also removed; (2) the reads with a percentage of low-quality bases (lower than Q20) more than 40% were filtered; (3) the reads containing more than 10% N were excluded. The number of clean reads in all samples ranged from 22 to 32 million (Supplementary Table S1). Clean reads from different species were mapped to the genome sequences of the corresponding species using Bowtie2⁵³ and Tophat2⁵⁴ programs with default parameter values, which generated an average of 72% mapped ratio in all samples (Supplementary Table S1). The Cufflinks program⁵⁵ was used to calculate the FPKM values (fragments per kilobase of exon model per million mapped reads) for determination of gene expression levels. The expression dataset including 12,533 predicted gene from *B. minax* and 13,121 predicted gene from *B. dorsalis* was listed in Supplementary Table S2. Among them, the transcript information of these identified OBPs, CSPs and ORs was specifically extracted to evaluate their gene expression level. The Student's *t* test method without having to rely on the total input genes was suitable for cross-species comparisons of expression of OBPs, CSPs and ORs. Compared *B. minax* to *B. dorsalis* at the same development stage, genes were deemed significantly differentially expressed after setting corrected *p*-value ≤ 0.05 and the relative change threshold ≥ 2 -fold (to see Supplementary Table S3–S5). The genes with FPKM values greater than 100 were considered to be highly expressed.

Validation of gene expression by qRT-PCR. The expression of 5 genes from *B. minax* and *B. dorsalis* at the development of larvae and adult were validated using qRT-PCR, including *OBP99c1* (Bmi011415 and gene11468/BdorOBP10_hz2013/AKI29023.1/AGC82131.1/AGS08192.1/XP_011210420.1), *OBP99c2* (Bmi011416 and gene11474/AKI29026.1/AGC82130.1/XP_011210427.1), *OBPlush* (Bmi007112 and gene9962/AKI28996.1/XP_011208040.1), *CSP1* (Bmi004355 and gene876 /AKI28975.1/XP_011204433.1), *ORCO* (Bmi011214 and gene7268/AKI29027.1/JAC55447.1/XP_011203778.1). Two genes, GAPDH and A-tub, were selected from non-DEGs in the RNA-Seq dataset respectively as candidate internal reference genes of *B. minax*⁵⁶ and *B. dorsalis*⁵⁷. Primers were designed on the online website NCBI Primer-BLAST (<https://www.ncbi.nlm.nih.gov/tools/primer-blast/>) according to the coding sequences (CDS) of each gene and their corresponding transcriptome. Two pairs of primers per gene recommended by NCBI were selected and synthesized. To be consistent with RNA-Seq information, two *B. minax* individuals and four *B. dorsalis* individuals at the development

stage of larvae and adult (male:female = 1:1) were respectively used to extract total RNA as a biological replicate. Three biological replicates per sample were performed. The ages of the samples from *B. minax* and *B. dorsalis* were the same as that of the samples used for RNA-Seq, except the samples from pupa stage were not picked due to seasonal reasons. Complementary DNA was synthesized using the FirstKing cDNA Strand Synthesis Kit (KR180123) (Beijing TianGen biological co., LTD., China). Through gel imaging observation, one pair of better primers were selected for each target gene. Information on the primers is listed in Supplementary Table S6. qRT-PCR was performed with three technical replicates on a CFX-96_CFX Connect Real-Time PCR System (Bio-Rad, American). The relative expression levels of the 4 selected DEGs normalized to the expression level of the internal reference control were calculated using the $2^{-\Delta\Delta Ct}$ method⁵⁸, and coefficient analysis was carried out to evaluate the correlation between the qRT-PCR and the RNA-Seq results.

Data availability

All data generated or analysed during this study are included in this published article (and its Supplementary Information files).

Received: 5 August 2019; Accepted: 16 September 2020

Published online: 22 October 2020

References

- Xia, Y. L., Ma, X. L., Hou, B. H. & Ouyang, G. C. A Review of *Bactrocera minax* (Diptera:Tephritidae) in China for the purpose of safeguarding. *Adv. Entomol.* **6**, 35–61 (2018).
- Leal, W. S. Odorant reception in insects: roles of receptors, binding proteins, and degrading enzymes. *Annu. Rev. Entomol.* **58**, 373–391 (2013).
- Krieger, J., Nickischroenegk, E. V., Mameli, M., Pelosi, P. & Breer, H. Binding proteins from the antennae of *Bombyx mori*. *Insect Biochem. Mol. Biol.* **26**, 297 (1996).
- Lartigue, A. *et al.* X-ray structure and ligand binding study of a moth chemosensory protein. *J. Biol. Chem.* **277**, 32094–32098 (2002).
- Benton, R., Sachse, S., Michnick, S. W. & Vosshall, L. B. Atypical membrane topology and heteromeric function of *Drosophila* odorant receptors in vivo. *PLoS Biol.* **4**, 240–257 (2006).
- Vieira, F. G. & 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.* **3**, 476–490. <https://doi.org/10.1093/gbe/evr033> (2011).
- Venthur, H. & Zhou, J. J. Odorant receptors and odorant-binding proteins as insect pest control targets: a comparative analysis. *Front. Physiol.* **9**, 1163. <https://doi.org/10.3389/fphys.2018.01163> (2018).
- Field, L. M., Pickett, J. A. & Wadhams, L. J. Molecular studies in insect olfaction. *Insect Mol. Biol.* **9**, 545–551. <https://doi.org/10.1046/j.1365-2583.2000.00221.x> (2010).
- Zheng, W. W. *et al.* Identification and expression profile analysis of odorant binding proteins in the oriental fruit fly *Bactrocera dorsalis*. *Int. J. Mol. Sci.* **14**, 14936–14949. <https://doi.org/10.3390/ijms140714936> (2013).
- Wu, Z. Z. *et al.* Discovery of chemosensory genes in the oriental fruit fly, *Bactrocera dorsalis*. *PLoS ONE* **10**(6), e0129794. <https://doi.org/10.1371/journal.pone.0129794> (2015).
- Leal, W. S., Nikonova, L. & Peng, G. Disulfide structure of the pheromone binding protein from the silkworm moth, *Bombyx mori*. *FEBS Lett.* **464**, 85–90. [https://doi.org/10.1016/S0014-5793\(99\)01683-X](https://doi.org/10.1016/S0014-5793(99)01683-X) (1999).
- Angeli, S. *et al.* Purification, structural characterization, cloning and immunocytochemical localization of chemoreception proteins from *Schistocerca gregaria*. *Eur. J. Biochem.* **262**, 745–754. <https://doi.org/10.1046/j.1432-1327.1999.00438.x> (1999).
- Liu, G. X. *et al.* Biotype characterization, developmental profiling, insecticide response and binding property of *Bemisia tabaci* chemosensory proteins: role of CSP in insect defense. *PLoS ONE* **11**(5), e0154706. <https://doi.org/10.1371/journal.pone.0154706> (2016).
- Ozaki, M. *et al.* Ant nestmate and nonnestmate discrimination by a chemosensory sensillum. *Science* **309**, 311–314. <https://doi.org/10.1126/science.1105244> (2005).
- Larsson, M. C. *et al.* Or83b encodes a broadly expressed odorant receptor essential for *Drosophila* olfaction. *Neuron* **43**, 703–714 (2004).
- Vosshall, L. B. Diversity and expression of odorant receptors in *Drosophila*. *Insect Pheromone Biochem. Mol. Biol.* **18**, 567–591 (2003).
- Robertson, H. M., Warr, C. G. & Carlson, J. R. Molecular evolution of the insect chemoreceptor gene superfamily in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. U.S.A.* **100**(Suppl 2), 14537 (2003).
- Smith, D. P. Odor and pheromone detection in *Drosophila melanogaster*. *Pflügers Arch. Eur. J. Physiol.* **454**, 749–758 (2007).
- Jones, W. D., Nguyen, T. A., Kloss, B., Lee, K. J. & Vosshall, L. B. Functional conservation of an insect odorant receptor gene across 250 million years of evolution. *Curr. Biol. Cb* **15**, R119–R121 (2005).
- Venthur, H., Mutis, A., Zhou, J.-J. & Quiroz, A. Ligand binding and homology modelling of insect odorant-binding proteins. *Physiol. Entomol.* **39**, 183–198. <https://doi.org/10.1111/phen.12066> (2014).
- Clarke, A. R. *et al.* Invasive phytophagous pests arising through a recent tropical evolutionary radiation: the *Bactrocera dorsalis* complex of fruit flies. *Annu. Rev. Entomol.* **50**, 293–319. <https://doi.org/10.1146/annurev.ento.50.071803.130428> (2005).
- Siciliano, P. *et al.* Sniffing out chemosensory genes from the Mediterranean fruit fly, *Ceratitis capitata*. *PLoS ONE* **9**(1), e85523. <https://doi.org/10.1371/journal.pone.0085523> (2014).
- Missbach, C. *et al.* Evolution of insect olfactory receptors. *Elife* **3**, e02115. <https://doi.org/10.7554/eLife.02115> (2014).
- Zheng, W. *et al.* Identification and expression profile analysis of odorant binding proteins in the oriental fruit fly *Bactrocera dorsalis*. *Int. J. Mol. Sci.* **14**, 14936–14949 (2013).
- Sun, Y. L., Huang, L. Q., Pelosi, P. & Wang, C. Z. Expression in antennae and reproductive organs suggests a dual role of an odorant-binding protein in two sibling *Helicoverpa* species. *PLoS ONE* **7**, e30040 (2012).
- Robertson, H. M. & Wanner, K. W. The chemoreceptor superfamily in the honey bee, *Apis mellifera*: expansion of the odorant, but not gustatory, receptor family. *Genome Res.* **16**, 1395 (2006).
- 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**(2), e0147783. <https://doi.org/10.1371/journal.pone.0147783> (2016).
- Jacquinjoly, E., Vogt, R. G., François, M. C. & Nagnanle, M. P. Functional and expression pattern analysis of chemosensory proteins expressed in antennae and pheromonal gland of *Mamestra brassicae*. *Chem. Senses* **26**, 833–844. <https://doi.org/10.1093/chemes/26.7.833> (2001).

29. Duan, G. G., Li, D.-Z. & Wang, M. Q. Chemosensory protein used as target for screening behaviourally active compounds in rice pest *Cnaphalocrocis medinalis* (Lepidoptera: Pyralidae). *Insect Mol. Biol.* <https://doi.org/10.1111/imb.12532> (2018).
30. Zhao, Y. *et al.* Sex- and tissue-specific expression profiles of odorant binding protein and chemosensory protein genes in *Bradyaodoriphaga* (Diptera: Sciaridae). *Front. Physiol.* **9**, 107 (2018).
31. Maleszka, J., Foret, S., Saint, R. & Maleszka, R. RNAi-induced phenotypes suggest a novel role for a chemosensory protein CSP5 in the development of embryonic integument in the honeybee (*Apis mellifera*). *Dev. Genes Evol.* **217**(3), 189–196. <https://doi.org/10.1007/s00427-006-0127-y> (2007).
32. Chang, C. L., Vargas, R. I., Caceres, C., Jang, E. & Cho, I. K. Development and assessment of a liquid larval diet for *Bactrocera dorsalis* (Diptera: Tephritidae). *Ann. Entomol. Soc. Am.* **99**, 1191–1198. [https://doi.org/10.1603/0013-8746\(2006\)99\[1191:daaaolj2.0.co;2](https://doi.org/10.1603/0013-8746(2006)99[1191:daaaolj2.0.co;2) (2006).
33. Liu, H. *et al.* BdorOBP2 plays an indispensable role in the perception of methyl eugenol by mature males of *Bactrocera dorsalis* (Hendel). *Sci. Rep.* **7**, 15894. <https://doi.org/10.1038/s41598-017-15893-6> (2017).
34. Jin, T., Zeng, L., Lin, Y. Y., Lu, Y. Y. & Liang, G. Insecticide resistance of the oriental fruit fly, *Bactrocera dorsalis* (Hendel) (Diptera: Tephritidae), in mainland China. *Pest Manag. Sci.* **67**, 370–376. <https://doi.org/10.1002/ps.2076> (2011).
35. Nishikawa, H. *et al.* A genetic mechanism for female-limited Batesian mimicry in Papilio butterfly. *Nat. Genet.* **47**(4), 405–409. <https://doi.org/10.1038/ng.3241> (2015).
36. Luo, R. B. *et al.* SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *GigaScience* **1**, 18. <https://doi.org/10.1186/2047-217x-1-18> (2012).
37. English, A. C. *et al.* Mind the gap: upgrading genomes with Pacific Biosciences RS long-read sequencing technology. *PLoS ONE* **7**(11), e47768. <https://doi.org/10.1371/journal.pone.0047768> (2012).
38. RepeatModeler. <https://www.repeatmasker.org/RepeatModeler/>.
39. Xu, Z. & Wang, H. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res.* **35**, W265–W268 (2007).
40. Bao, W., Kojima, K. K. & Kohany, O. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mobile DNA* **6**, 11 (2015).
41. 41RepeatMasker. <https://www.repeatmasker.org/>.
42. Campbell, M. S. *et al.* MAKER-P: a tool-kit for the rapid creation, management, and quality control of plant genome annotations. *Plant Physiol.* **164**, 513–524 (2014).
43. Benson, G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* **27**, 573–580 (1999).
44. Stanke, M., Diekhans, M., Baertsch, R. & Haussler, D. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics* **24**(5), 637–644. <https://doi.org/10.1093/bioinformatics/btn013> (2008).
45. Burge, C. B. & Karlin, S. Finding the genes in genomic DNA. *Curr. Opin. Struct. Biol.* **8**, 346–354. [https://doi.org/10.1016/S0959-440X\(98\)80069-9](https://doi.org/10.1016/S0959-440X(98)80069-9) (1998).
46. Majoros, W. H., Pertea, M. & Salzberg, S. L. TigrScan and GlimmerHMM: two open-source ab initio eukaryotic gene-finders. *Bioinformatics* **20**, 2878–2879. <https://doi.org/10.1093/bioinformatics/bth315> (2004).
47. Holt, C. & Yandell, M. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinform.* **12**, 491. <https://doi.org/10.1186/1471-2105-12-491> (2011).
48. Robert, M. W. *et al.* BUSCO applications from quality assessments to gene prediction and phylogenomics. *Mol. Biol. Evol.* **35**(3), 543–548. <https://doi.org/10.1093/molbev/msx319> (2017).
49. Hekmatscafe, D. S., Scafe, C. R., Mckinney, A. J. & Tanouye, M. A. Genome-wide analysis of the odorant-binding protein gene family in *Drosophilamelanogaster*. *Genome Res.* **12**, 1357–1369 (2002).
50. Larkin, M. A. *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–2948. <https://doi.org/10.1093/bioinformatics/btm404> (2007).
51. Nicholas, K. B., Nicholas, H. B. & Deerfield, D. W. GeneDoc: analysis and visualization of genetic variation. *Embnew News* **4**, 14 (1997).
52. Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* **30**, 2725–2729. <https://doi.org/10.1093/molbev/mst197> (2013).
53. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol. Evol.* **10**, R25 (2009).
54. Kim, D. *et al.* TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36 (2013).
55. 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).
56. Lü, Z.-C. *et al.* Evaluation of endogenous reference genes of *Bactrocera* (Tetradacus) minaxby gene expression profiling under various experimental conditions. *Fla. Entomol.* **97**, 597–604. <https://doi.org/10.1653/024.097.0235> (2014).
57. Shen, G.-M., Huang, Y., Jiang, X.-Z., Dou, W. & Wang, J.-J. Effect of β -cypermethrin exposure on the stability of nine housekeeping genes in *Bactrocera dorsalis* (Diptera Tephritidae). *Fla. Entomol.* **96**(2), 442–450 (2013).
58. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. *Methods* **25**, 402–408 (2001).

Acknowledgements

This research was supported by National key research and development project of China (No. 2016YFC1201200), Science and Technology Project of Guangdong Province, China (No. 2016B030303007, 2017A070702005, 2017B020203003), Science and Technology Program of Guangzhou, China (No. 201904010432). We thank Master students Zijun Guo (Yunnan Agricultural University, China), Fulian Xiao and Biya Gong (Hunan Horticultural Research Institute, China) for their help in collecting insects. We thank professor Reinaldo A. De Brito (Universidade Federal de Sao Carlos, Brazil) for his help in revising this manuscript.

Author contributions

Y.Y.L. conceived and designed the research. J.L. supervised the project. J.F.C., Y.P.C. and X.B. performed the experiments. T.Y. and J.F.C. analyzed the data. T.Y., Z.J.C., S.C.C. and J.F.C. wrote the draft of this paper. Z.J.C., E.D.W. and S.C.C. edited this paper. S.C.C. submitted this paper. Y.Y.L. and Y.P.C. provided the method of the insect rearing, genome and transcriptome sequencing. T.Y. and Y.P.C. provided the results of qRT-PCR. L.G., W.H.Z. and B.J. contributed materials/analysis tools.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41598-020-74803-5>.

Correspondence and requests for materials should be addressed to J.L. or Y.-Y.L.

Reprints and permissions information is available at www.nature.com/reprints.

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



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020