Identification and expression analysis of chemosensory genes in the citrus fruit fly *Bactrocera* (*Tetradacus*) *minax*

Jun-Feng Cheng [1], **Ting Yu** [2], **Zhong-Jian Chen** [2], **Yu-Peng Chen** [1], **Lei Gao** [3], **Wen-Hu Zhang** [2], **Bo Jiang** [4], **Xue Bai** [1], **Jun Liu** [Corresp., 2], **Yong-Yue Lu** [Corresp. 1]

[1] College of Agriculture, South China Agricultural University, GuangZhou, GuangDong, China

[2] Agro-biological Gene Research Center, Guangdong Academy of Agricultural Sciences, GuangZhou, GuangDong, China

[3] Crop Research Institute, Guangdong Academy of Agricultural Sciences, Guangdong Provincial Key Laboratory of Crop Genetics and Improvement, GuangZhou, GuangDong, China

[4] Fruit Tree Research Institute, Guangdong Academy of agricultural Sciences, GuangZhou, GuangDong, China

Corresponding Authors: Jun Liu, Yong-Yue Lu
Email address: 930880613@qq.com, 249910360@qq.com

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 hosts via an efficient chemosensory system. However, knowledge regarding the molecular components of the *B. minax* chemosensory system has not yet been well established. Herein, based on our newly generated whole-genome dataset for *B. minax* and by comparison with the characterized genomes of 6 other fruit fly species, we identified, for the first time, a total of 25 putative odorant-binding receptors (OBPs), 4 single-copy chemosensory proteins (CSPs) and 53 candidate odorant receptors (ORs). To further survey the expression of these candidate genes, the transcriptomes from three developmental stages (larvae, pupae and adults) of *B. minax* and *Bactrocera dorsalis* were analyzed. We found that 1) at the adult developmental stage, there were 14 highly expressed OBPs (FPKM>100) in *B. dorsalis* and 7 highly expressed OBPs in *B. minax*; 2) the expression of CSP3 and CSP4 in adult *B. dorsalis* was higher than that in *B. minax*; and 3) most of the OR genes exhibited low expression at the three developmental stages in both species. This study on the identification of the chemosensory system of *B. minax* not only enriches the existing research on insect olfactory receptors but also provides new targets for preventative control and ecological regulation of *B. minax* in the future.

Identification and expression analysis of chemosensory genes in the citrus fruit fly *Bactrocera* (*Tetradacus*) *minax*

Jun-Feng Cheng[1#], Ting Yu[2#], Zhong-Jian Chen[2#], Yu-Peng Chen[1#], Lei Gao[3], Wen-Hu Zhang[2], Bo Jiang[4], Xue Bai[1], Jun Liu[2*], and Yong-Yue Lu[1*]

[1]College of Agriculture, South China Agricultural University, Guangzhou, Guangdong, China

[2]Agro-biological Gene Research Center, Guangdong Academy of Agricultural Sciences, Guangzhou, Guangdong, China

[3]Crop Research Institute, Guangdong Academy of Agricultural Sciences, Guangdong Provincial Key Laboratory of Crop Genetics and Improvement, Guangzhou, Guangdong, China

[4]Fruit Tree Research Institute, Guangdong Academy of agricultural Sciences, Guangzhou, Guangdong, China

[*]Corresponding authors

Yong-Yue Lu[1]

No.483 Wushan Road, Guangzhou, Guangdong, China

Email address: luyongyue@scau.edu.cn

Jun Liu[2]

No.20 Jinying Road, Guangzhou, Guangdong, China

Email address: liujun@gdaas.cn

# These authors have contributed equally to this work.

ABSTRACT

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 hosts via an efficient chemosensory system. However, knowledge regarding the molecular components of the *B. minax* chemosensory system has not yet been well established. Herein, based on our newly generated whole-genome dataset for *B. minax* and by comparison with the characterized genomes of 6 other fruit fly species, we identified, for the first time, a total of 25 putative odorant-binding receptors (OBPs), 4 single-copy chemosensory proteins (CSPs) and 53 candidate odorant receptors (ORs). To further survey the expression of these candidate genes, the transcriptomes from three developmental stages (larvae, pupae and adults) of *B. minax* and *Bactrocera dorsalis* were analyzed. We found that 1) at the adult developmental stage, there were 14 highly expressed OBPs (FPKM>100) in *B. dorsalis* and 7 highly expressed OBPs in *B. minax*; 2) the expression of CSP3 and CSP4 in adult *B. dorsalis* was higher than that in *B. minax*; and 3) most of the OR genes exhibited low expression at the three developmental stages in both species. This study on the identification of the chemosensory system of *B. minax* not only enriches the existing research on insect olfactory receptors but also provides new targets for preventative control and ecological regulation of *B. minax* in the future.

Subjects: Agricultural science, Bioinformatics, Entomology, Genomics.

Keywords: *Bactrocera (Tetradacus) minax*; Odorant-binding protein; Chemosensory proteins; Odorant receptors; *Bactrocera dorsalis.*

Introduction

Bactrocera (Tetradacus) minax is a destructive pest that damages mainly citrus fruit trees and affects citrus production (Xia et al., 2018). Severe outbreak can regularly occur once a year in all citrus-producing areas, which often leads to severe economic loss in the citrus industry (Xia et al., 2018). Therefore, the development of prevention control methods specifically targeting *B. minax* is urgently needed.

Insects recognize their hosts mainly via perception of chemical signals (Leal, 2013). Extensive studies have shown that many proteins, such as odorant-binding proteins (OBPs), chemosensory proteins (CSPs) and odorant receptors (ORs), are involved in the chemosensitivity of insects (Krieger et al., 1996; Lartigue et al., 2002; Benton et al., 2006; Vieira and Rozas, 2011; Leal, 2013; Venthur et al., 2018). Currently, the proposed model for insect chemosensitivity and recognition is as follows (Field et al., 2010; Vieira and Rozas, 2011; Leal, 2013; Venthur et al., 2018): some lipophilic odorant molecules in the environment reach the hydrophilic lymph of the insect through the micropores on the olfactory sensilla surface and then form a complex with the OBPs or CSPs in the sensillary lymph. Then, the complex passes through the sensillary lymph and binds to the ORs on the dendritic membranes. When ORs are stimulated, the membrane permeability changes, resulting 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 to fulfill physiological responses, such as foraging, searching for shelter, avoiding enemies and finding spawning sites.

OBPs are water-soluble proteins with relatively small molecular weights and consist of six conserved cysteine residues, with a signal peptide of 20 amino acids at the N terminus (Pelosi and Maida, 1995). Six alpha-helical dimers constitute the OBP secondary structure, and six cysteines form three cross-linked disulfide bonds that reinforce the OBP structure (Briand et al., 2010; Field et al., 2010). The sizes of OBPs are generally approximately 15-17 kD, consisting of 120-150 amino acids (Zheng et al., 2013). While the complete OBPs from *Bactrocera dorsalis* encode the protein length ranging from 134 to 274 amino acids (Wu et al., 2015). OBPs located in the surface of insect antennae and their main function is to bind odorants such as volatile substances (Venthur et al., 2018). The interaction between OBPs and fat-soluble odorant molecules is the first biochemical reaction in the identification of odorant compounds by insects, which is of great significance for insect perception and communication with the external environment (Leal, 2013).

Similar to OBPs, CSPs are also small, highly water-soluble, and acidic proteins with hydrophobic binding sites (Leal et al., 1999). CSPs are only approximately 120 amino acids in length (12-14 kD) and the secondary structure is also composed of six alpha-helices, but with only four conserved cysteines, forming two disulfide bonds (Lartigue et al., 2002; Tegoni et al., 2004). CSPs have been found to play an important role in the chemosensory process as a sort of odor binding proteins by delivering hydrophobic sensory molecules to trigger neuronal responses in locusts, ants and *Bemisia tabaci* (Angeli et al., 1999; Ozaki et

al., 2005; Liu et al., 2016).

Insect ORs are a group of G-protein-coupled receptors with seven transmembrane domains. The N terminus of these ORs is located within the cell, and the C terminus is located outside, which is different from the ORs of vertebrates (Benton et al., 2006). Insect ORs include conventional ORs and the atypical OR Or83b (Larsson et al., 2004). Drosophila olfactory conventional ORs and Or83b encode 370~400 amino acids and 486 amino acids, respectively (Vosshall, 2003). The insect Or83b protein has been found to be coexpressed with conventional ORs in most olfactory neurons (Robertson et al., 2003; Smith, 2007) and has a considerable impact on the olfactory behavior of insects (Jones et al., 2005).

Thus, the identification of olfaction-related genes and proteins may provide insight into novel strategies for specific targeting of particular agricultural pests. The use of sex pheromones to trap and kill adults is an important technique for predicting and controlling *B. minax* infestation, which can effectively reduce the number of adults and the rate of fruit damage (Xia et al., 2018). Hence, the recognition of hosts and pheromones by *B. minax* has become a hot topic in recent years. However, the molecular components of olfaction in *B. minax* have not been well studied. In this study, based on our newly generated whole-genome sequence of *B. minax* and by comparison with the characterized genomes of 6 other fruit fly species, the genes related to olfactory perception in *B. (Tetradacus) minax* were identified. The expression patterns at three different developmental stages (larva, pupa and adult) were compared between *B. dorsalis* and *B. minax*.

Materials and Methods

Insect Rearing and Ethics Statement

Bactrocera (Tetradacus) minax. *B. minax* insects were collected and identified by the Hunan Academy of Agricultural Sciences in Jishou City, Hunan Province, China. Because it is a major fruit fly pest in tropical and subtropical countries, this species is not included in the "List of Endangered and Protected Animals in China". The larvae were reared in tangerines and transformed to pupae in wet soil. The larvae used herein were collected at the 3rd instar stage. The pupae used herein were collected 20 days after pupation. The adults used herein was collected 12 h after emergence without any feeding.

Bactrocera dorsalis (Hendel). *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). The artificial diets for the larvae were provided by the Institute of Insect Ecology of South China Agricultural University; these diets included banana, corn flour, sucrose, yeast extract, paper, sodium benzoate, hydrochloric acid and water in appropriate proportions. Larvae were transformed to pupae in wet sand. The positive artificial diets for the adults were yeast extract and sucrose mixed at a 1:1 ratio (Jin et al., 2011). The larvae used herein were collected 6 days after incubation at the 3rd instar stage. The pupae used herein were collected 6 days after pupation. The adults used herein were collected 12 h after emergence without any feeding.

All experiments were performed in compliance with general ethical guidelines to minimize pain and discomfort to the insects.

Sample Preparation and Genome Sequencing

The body parts of only *B. minax* males were performed to obtain DNA for whole-genome sequencing. The total DNA was extracted using the SDS method from Magen

reagent kit (Guangzhou, China) according to the manufacturer's instructions and the samples were sent to the GeneDenovo company (Guangzhou, China) for libraries 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. Brief description in herein, for NGS, five gradient insert libraries were constructed, two short fragmentary libraries and three long fragmentary libraries. For third-generation sequencing, two libraries were constructed; five SMRT cells were sequenced; 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 (version1.2.1) (Nishikawa et al., 2015) was used to assemble Illumina data and GapCloser (v1.10) (Luo et al., 2012) was used to extend the contig length; 2) PBjelly (PBSuite_15.8.24) (English et al., 2012) was adopted to extend the scaffolds and fill the holes with combining the third-generation sequencing data. Finally a genome size of 340 Mb was obtained (unpublished). Then, several methods of de novo prediction, homologous prediction, and cDNA/EST prediction were all used to predict the gene structures. Three softwares including Augustus 2.7 (Stanke et al., 2008), Genscan 1.0 (Burge and Karlin, 1998), Glimmer HMM 3.0.1 (Majoros et al., 2004) were used to do the de novo prediction of gene models. Two softwares EVM r2012-06-25 (Haas et al., 2008) and MAKER 2.28 (Holt and Yandell, 2011) were used to integrate all gene sets to a non-redundant and more complete gene set. At last, a reliable gene set of 12,533 gene models were 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), *Ceratitis capitata* (Ccap), and *Rhagoletis zephyria* (Rzep) were downloaded from NCBI (ftp://ftp.ncbi.nlm.nih.gov/genome). The protein sequences from above species, which were annotated as chemosensing-related genes, were all extracted. The sequences of fifty-one OBP genes and one lush gene from *Drosophila melanogaster* obtained from a reference (Hekmatscafe et al., 2002) and those of 31 OBP genes from *B. dorsalis,* also obtained from a reference (Wu et al., 2015), were collected. The sequences of forty-seven CSP genes from 12 fruit fly species obtained from a reference (Vieira and Rozas, 2011) and those of four CSPs (CSP1-4) from *B. dorsalis*, also obtained from a reference (Wu et al., 2015), were collected. The sequences of sixty-two OR genes from *D. melanogaster* obtained from a reference (Robertson et al., 2003) and those of 23 OR genes from *B. dorsalis*, also obtained from a reference (Wu et al., 2015), were collected. The protein sequences from *B. minax* were aligned to the annotated sequences from 6 fruit fly species, and the collected 83 OBPs, 51 CSPs, 85 ORs using BLASTP with an e-value setting of 1e-5. Then, the hit protein sequences of *B. minax* were extracted and appended to the sequences set for subsequent analysis.

Sequence Comparison and Phylogenetic Analysis

ClustalX (Larkin et al., 2007) was used to perform multiple sequence alignment with default parameters, and GeneDoc (Nicholas et al., 1997) was used to visualize the alignment information. Highly divergent sequences were removed. Then, the sequence alignment was input into MEGA6 (Tamura et al., 2013) to construct a phylogenetic tree by the neighbor-

joining (NJ) method. The pairwise deletion option was selected under the JTT empirical amino acid substitution model. Branch support was assessed by bootstrap analysis with 500 replicates. The gene names for *B. minax* and six other fruit fly species were modified according to the phylogenetic clusters.

Sequence Information

The OBP, CSP and OR family sequences from *B. minax* have been submitted the GenBank database for their accession numbers MH937211 - MH937240, MH937207 - MH937210, MH937241 - MH937290. The protein sequences and alignments of each OBP/CSP/OR member with those from other species are presented in supplemental file 1, supplemental file 2 and supplemental file 3, respectively.

RNA Sequencing and Transcript Analysis

Total RNA was extracted from the larvae, pupae and adults of *B. minax* and *B. dorsalis* using TRIzol reagent (Invitrogen, California, USA) following the manufacturer's instructions. The adult samples were all obtained from females and males in a 1:1 ratio. Two duplicates per developmental stage were used to construct the cDNA libraries by adopting the Illumina TruSeq RNA Sample Preparation kit and the mixed libraries were sequenced on the Illumina Hiseq 4000 platform with paired-end 150-bp reads. Clean reads obtained by trimming low-quality bases and removing short reads were mapped to the *B. minax* and *B. dorsalis* genomes using the Bowtie2 (Langmead et al., 2009) and tophat2 (Kim et al., 2013) programs with default parameter values. The Cufflinks program (Mortazavi et al., 2008) was used to calculate the FPKM (fragments per kilobase of exon model per million mapped reads) values for determination of gene expression levels. EdgeR (Robinson et al., 2010)

199 was used to identify differentially expressed genes. Genes were deemed significantly
200 differentially expressed when the p-values were ≤0.05 by Student's *t*-test and the relative
201 change threshold was 2-fold.

202 ## Results

203 ### 3.1 Identification and Gene Expression of OBPs in *Bactrocera minax*

204 To identify OBP genes in *B. minax*, the predicted gene sequences of *B. minax* were
205 analyzed by comparison with those of 6 other fruit fly species. As shown in Table 1, a total
206 of 25, 37, 33, 35, 30, 29, 34 OBP genes were identified in *B. minax* (Bmin), *B. dorsalis*
207 (Bdor), *B. cucurbitae* (Bcuc), *B. latifrons* (Blat), *B. oleae* (Bole), *R. zephyria* (Rzep) and
208 *C. capitata* (Ccap), respectively. Bdor contained the most OBPs, followed by Ccap, while
209 Bmin contained the fewest OBPs. Among the 52 OBPs in *D. melanogaster*, 37 were
210 identified in Tephritidae, namely,8a, 19a, 19b, 19c, 19d, 28a, 44a, 46a, 47a, 47b, 50a, 50c,
211 50e, 56a, 56b, 56c, 56d, 56e, 56g, 56h, 57c, 58c, 58d, 59a, 69a, 73a, 83a, 83b, 83cd, 83ef,
212 83g, 84a, 99a, 99b, 99c, 99d, and lush (Table 1).

213 A phylogenetic tree was generated to show the relationships among the identified
214 OBPs of *B. minax* and those of the above mentioned Dipteran species. Eight OBPs, namely,
215 19a, 19d, 50e, 56a, 56d, 56h, 84a, and 99c, could be amplified from Tephritidae, among
216 which, OBP 50e exhibited species-specific amplification from Bcuc (7 copies) and Rzep (5
217 copies), while OBP 99c exhibited species-specific amplification from Bdor (8 copies) and
218 Ccap (6 copies) (Table 1; Figure 1).

219 Based on the conserved cysteine profiles present, OBPs can be classified into three types:
220 the first type is classic OBPs, with the six conserved cysteine residues characteristic of insect

OBPs. Approximately 20 classic OBPs were identified in Tephritidae, including 19a, 19b, 19c, 19d, 28a, 44a, 47a, 56b, 56c, 56d, 56e, 56g, 56h, 57c, 69a, 73a, 83a, 83b, 83g, 99a, 99b, and lush. The second type is plus-c OBPs, with more than six conversed cysteines. Twelve plus-c OBPs were identified in Tephritidae, and these OBPs were clustered in two classes on the phylogenetic tree, with one cluster containing eight OBPs (46a, 47b, 50a, 50c, 50e, 56a, 58c, 58d) and the other cluster containing four OBPs (59a, 83cd, 83ef, 84a). The third type is "Minus-C" OBPs, with less than six conversed cysteines, usually missing one or two cysteines. Three "Minus-C" OBPs (8a, 99c, 99d) were identified in Tephritidae and were clustered together on the phylogenetic tree, suggesting that these OBPs share different evolutionary relationships compared to the classic OBPs.

To investigate the developmental stage-specific expression of the candidate OBP genes, the transcriptomes of *B. minax* and *B. dorsalis* were analyzed, and the genes with FPKM values greater than 100 were considered to be highly expressed genes. As shown in Figure 2, six (from high to low: **99c1,** 56d, 50c, 83g, 19a2, 99b) and six (**99c1,** 56d3, 99c5, 44a, 99b, 56d2) OBP genes were highly expressed in *B. minax* and *B. dorsalis* larvae, respectively, four (**99c1,** 99b, 50c, 56d) and five (**99c1,** 99b, 44a, 50c, 83g) in the pupae, and seven (99c2, 99b, **99c1**, 19d2, 83g, 28a, 19d1) and fourteen (**99c1,** 44a, 99b, 56d1, 56d3, 56d2, 50c, 83a, 83g, 19d1, 19d2, 83b, 56g, 28a) in the adults (Figure 2). Therefore, during the adult developmental stage, there were more highly expressed OBP genes in *B. dorsalis* than in *B. minax*, implying that *B. dorsalis* may have stronger odorant-binding activity than *B. minax.* Among the OBP genes, OBP99c1 was the expressed at the highest levels at all the indicated developmental stages in both species except the adult stage of *B. minax*,

indicating that OBP99c1 is a key gene in *Bactrocera*. The expression level of OBP99c1 peaked (FPKM=12,091) at the pupal stage of *B. dorsalis* and was lowest (FPKM=672) at the adult stage of *B. minax*, suggesting that OBP99c1 not only plays a role in adults by affecting odorant-binding protein function but also may play important roles associated with the processes of metamorphosis, detoxification and resistance in the pupae. The transcriptional level of OBP99c2 was highest (FPKM=3,823) in *B. minax* adults and lowest (FPKM=6) in *B. dorsalis* adults, indicating that OBP99c2 is an important, species-specific gene in *B. minax.*

3.2 Identification and Gene Expression of CSPs in *Bactrocera minax*

Compared to the 6 reference genome datasets, we found that, similar to *D. melanogaster*, among all tested species of Tephritidae, the CSP family included four single-copy genes. These 4 CSP genes had different origins and were named CSP1, CSP2, CSP3, CSP4. As shown in Figure 3, CSP1 and CSP2 had the same ancestor; CSP4 appeared later than the other CSPs. The CSPs in Bdor were most closely related to the CSPs in Blat, followed by Bmin, Bcuc, Bole, Ccap, Rzep and fruit fly (Dmel). To date, the known genome of Diptera carries less than 10 CSP genes, of which, *D. melanogaster* and *B. dorsalis* carry only 4 each.

There were four conserved cysteines in CSP1-4. The interspecies sequence homology of CSPs among Tephritidae was higher than that with fruit fly. All the CSP3 genes of the 7 Tephritidae species had two splice variants, namely, CSP3X1 and CSP3X2. The average identities of CSP1, CSP2, CSP3X1, CSP3X2, CSP4 among Tephritidae were 91%, 91%, 82%, 80%, and 89%, respectively (Table 2; Supplement File), while the average degrees of homology of CSP1, CSP2, CSP3X1, CSP3X2, and CSP4 between Tephritidae and fruit fly

were 64%, 73%, 50%, 28%, and 66%, respectively (Table 2).

From the transcript data, CSP1 was expressed at the highest levels in the pupae of both *B. minax* and *B. dorsalis*, while CSP2-4 was expressed at higher levels than CSP1 at the adult stages in both species, where the expression level of CSP2 was higher than those of CSP3 and CSP4. However, the expression of CSP3 and CSP4 was lower in adult *B. minax* than in *B. dorsalis* (Figure 4).

3.3 Identification and Gene Expression of ORs in *Bactrocera minax*

To search for candidate OR genes in *B. minax*, sequence homology comparison and phylogenetic tree analysis were performed with known fruit fly genomes. As shown in Table 3, a total of 53, 70, 61, 58, 59, 64, and 61 OR genes were identified in *B. minax* (Bmin), *B. dorsalis* (Bdor), *B. cucurbitae* (Bcuc), *B. latifrons* (Blat), *B. oleae* (Bole), *R. zephyria* (Rzep) and *C. capitata* (Ccap), respectively. The Bdor species contained the most ORs, while Bmin contained the fewest. Thirty-five of the 62 ORs from *D. melanogaster* were present in Tephritidae, namely, ORCR, 2a, 7a, 10a, 13a, 22c, 24a, 33abc (3 ORs in Dmel), 35a, 43a, 45a, 46aA/B (2), 47b, 49a, 49b, 59a, 63a, 67c, 67d, 69aA/B(2), 71a, 74a, 82a, 83a, 85bc (2), 85d, 85e, 88a, and 94ab (2) (Table 3). Fourteen ORs, namely, 7a, 33abc, 45a, 59a, 63a, 67d, 69aA/B, 74a, 83a, 85bc, 85d, 85e, 88a, and 94ab, exhibited gene amplification in Tephritidae; OR85e exhibited species-specific amplification (3 copies) in Bcuc, while OR88a exhibited species-specific amplification (3 copies) in Bole (Figure 5). In addition, nine Tephritidae-specific OR genes were identified, which were individually named OR1, OR2, OR3, OR4, OR5, OR6, OR7, OR8 and OR9. All

of these ORs except OR9 exhibited gene amplification in Tephritidae, and OR3 exhibited species-specific amplification (8 copies) in Bdor.

The number of OR families was greater than that of OBP and CSP families. Based on the OR phylogenetic, we determined that twenty-one OR genes in Tephritidae originated from OR genes of *D. melanogaster*, namely, 2a, 7a, 10a, 13a, 22c, 24a, 33abc, 43a, 45a, 46a, 47b, 49a, 49b, 59a, 67c, 67d, 69a, 71a, 82a, 85e, 88a; six OR genes in Tephritidae were older than the OR genes of *D. melanogaster*, namely, 63a, 74a, 83a, 85bc, 85d, and 94ab; and nine OR genes in Tephritidae evolved with the OR genes of *D. melanogaster*, such as OR1-9. These results showed that the OR gene family underwent rapid evolution with a large amount of variation.

Based on the transcriptional data, with the exception of the OR46a1 gene, which exhibited moderate expression between 8.1 and 13.4 at the larval stage of *B. minax*, most OR genes exhibited lower expression or even no expression at the three developmental stages in both species (Supplement Table 3). Therefore, the physiological functions of these genes should be verified in future studies.

Discussion

B. minax Enderlein almost always targets the fruit from the *Citrus* genus of the Rutaceae family (Xia et al., 2018). On the other hand, *B. dorsalis Hendel* affects a wide range of host species, including more than 40 families and 200 different types of fruits and vegetables (Clarke et al., 2005). Insects recognize their hosts mainly via the insect chemosensory system, the major components of which include OBPs, CSPs and ORs, to locate host plants (Vieira et al., 2011). Here, by using comparative genomic technology,

we identified 37 OBPs, 4 CSPs, and 70 ORs in the genome of *B. dorsalis*, surpassing a previous report of 31 OBPs, 4 CSPs and 23 ORs in the transcriptomes of four *B. dorsalis* developmental stages (egg, larva, pupa and adult chemosensory tissues) (Wu et al., 2015) and 20 OBPs, 5 CSPs and 35 ORs in the transcriptomes of the antennae of male and female adults of *B. dorsalis* (Liu et al., 2016). The chemosensing-related genes of *B. minax* have not been characterized previously, and our study identified 25 OBPs, 4 CSPs and 53 ORs in the genome of *B. minax*. Thus, there were fewer chemosensing-related genes in *B. minax* than in *B. dorsalis.* At the adult stage, OBPs can bind various odorant molecules in the environment and transport these molecules to the ORs to induce the olfactory signal transduction system (Leal, 2013; Venthur et al., 2018). Furthermore, our results showed that at the adult developmental stage, 14 OBPs were highly expressed (FPKM>100) in *B. dorsalis,* while 7 OBPs were highly expressed in *B. minax*, suggesting that *B. dorsalis* may exhibit odorant binding. CSPs play similar roles as OBPs in chemical communication in insects, binding small molecules, such as odorants and pheromones (Pelosi et al., 2005). The expression of CSP3 and CSP4 was higher in adult *B. dorsalis* than in *B. minax*, indicating that CSP transcription may be stronger in *B. dorsalis* than in *B. minax*. Therefore, the higher the total number of olfaction-related genes, the greater the number of highly expressed OBPs at the adult stage, and the high expression of CSPs in *B. dorsalis* may account for the fact that *B. dorsalis* has a wider range of host species than *B. minax*.

In addition, we also identified 33 OBPs, 4 CSPs and 61 ORs in the genome of *B. cucurbitae* (in the Supplement Figure File 1-3), which is more than the 13 OBPs and 1 OR

previously identified in the transcriptome of adult *B. cucurbitae* melon fly (Elfekih et al., 2016). Siciliano reported 17 OBPs in Mediterranean fruit fly, *C. capitata* (medfly), from the EST libraries of adult heads, embryos, male accessory glands and testes (Siciliano et al., 2014). Herein, we identified 34 OBPs, 4 CSPs and 61 ORs in the genome of *C. capitata* (to see the Supplement Figure File 1-3).

From transcriptomic data, we obtained expression profiles of candidate OBPs, CSPs and ORs in *B. minax* and *B. dorsalis.* OBP99c1 was highly expressed in all the indicated developmental stages in both species and exhibited a gradually declining trend during *B. minax* development, suggesting that *OBP99c1* may play a crucial role in the development of the odorant-binding system in the citrus fruit fly. The gene corresponding to OBP99c1 in *B. dorsalis* (*BdorOBP10*) was highly expressed in females, especially in the abdomen, where the reproductive organs are located (Zheng et al., 2013), suggesting that this gene, if expressed, may be associated with oviposition. 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 eggs (Sun et al., 2012). In addition, OBP10 exhibited binding to an insect repellent, indicating that this protein may be a carrier for some semiochemicals (Sun et al., 2012). Therefore, OBP99c1 may be a good target for the design of a wide range of environmentally friendly pest control agents. Meanwhile, OBP99c2 was found to be highly expressed in *B. minax* adults, but was barely expressed in *B. dorsalis*, indicating that OBP99c2 may play a species-specific role in odorant system development.

CSPs are relatively conservative and stable in the evolution of insects. Not only do

CSPs of the same species maintain high homology, but also the homology among insects of different genera or different orders is high. For example, CSPs of *Schistocerca gregaria* and *Locusta migratoria* share 50-60% homology, and the homology between Orthoptera and Lepidoptera is around 37-50% (Jacquinjoly et al., 2001). Similarly, we found that the homology of CSPs among Tephritidae were higher than between Tephritidae and *D. melanogaster*. CSPs have been proposed to play a key role in olfactory perception in insects. Therefore, CSPs have been used to screen potential bioactive compounds for agricultural applications (Duan et al., 2018). The antenna-biased CSP CmedCSP33 from *Cnaphalocrocis medinalis*, a major rice pest in Asia, has been successfully matched with two behaviorally active compounds. Notably, insecticides have been shown to significantly upregulate adult-specific CSP1 gene expression in the sweet potato whitefly *Bemisia tabaci* (Liu et al., 2016). Additionally, in whitefly, due to the ligand binding specificity, CSP1 may be responsible for regulation of the insect immune response by fatty acids, while CSP2 and CSP3 facilitate insect communication with the surrounding environment via favorable or unpleasant odors (Liu et al., 2016). A recent study has 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 insect (Zhao et al., 2018). In fact, the honeybee *Apis mellifera* exhibits abnormal head development upon loss of CSP5 function and cannot progress to the larval stage (Maleszka et al., 2007). Therefore, further investigation of the tissue distribution of the CSPs identified in this study may facilitate the functional analysis of these genes.

The number of conventional ORs varis dramatically among different insect species.

For example, the genome of honey bee *Apis mellisfera* contains 170 OR genes with 7 pseudogenes, while *Drosophila melanogaster* and mosquito *Anopheles gambiae* carry 62 OR genes and 79 OR genes respectively (Robertson and Wanner, 2006). In our study, the number of ORs in Tephritidae (based on 6 fruit fly species) ranged from 53 to 70. Unlike the abundant expression profiles of OBPs and CSPs in *B. minax* and *B. dorsalis*, puzzlingly, the identified OR genes exhibited considerably low abundances or even undetectable expression during development. Previously, by means of qPCR analysis, two OR genes in *B. dorsalis*, namely, *Bdoror 13* and *Bdoror 14*, were identified as being highly and specifically expressed in the male antennae (Liu et al., 2016). The discrepancy between our study and the previous report may be due to the low expression of some transcripts in the whole insect and differences in examination approaches, which collectively led to decreased sensitivity of detection of the expression of the conserved ORs typically observed in insects (Wu et al., 2015). This hypothesis was supported by a previous study, which showed that ORs were expressed exclusively in insect antennae at low levels, such as in *B. dorsalis* (Wu et al., 2015).

CONCLUSION

In conclusion, we first 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 study compared the genetic relationships of candidate genes among 7 species. 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 elucidated and compared, demonstrating that *B.*

dorsalis had more OBPs and ORs than *B. minax*. In addition, the data showed that almost all the OBPs and the CSPs were more highly expressed in *B. dorsalis* than in *B. minax*. These findings indicated that *B. dorsalis* exhibited stronger odorant and pheromone binding properties than *B. minax*, which could explain why *B. dorsalis* exhibits a wider range of host species than *B. minax*. The chemosensory system of citrus fruit flies could be modified at a molecular level by using genome editing or gene knockout techniques for further exploration of ecologically friendly and highly efficient strategies for pest control.

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.2017A070702005, 2017B020203003). 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.

Author Contributions

- Yong-Yue Lu conceived and designed the research.
- Jun Liu supervised the research.
- Jun-Feng Cheng, Yu-Peng Chen and Xue Bai performed the experiments.
- Ting Yu and Jun-Feng Cheng analyzed the data.
- Ting Yu, Zhong-Jian Chen, and Jun-Feng Cheng wrote the draft of this paper.
- Zhong-Jian Chen reviewed this paper.
- Yong-Yue Lu and Yu-Peng Chen provided the method of the insect rearing, genome

and transcriptome sequencing.

- Ting Yu prepared figures and tables.
- Lei Gao, Wen-Hu Zhang and Bo Jiang contributed reagents/materials/analysis tools.

REFERENCES

Angeli S, Ceron F, Scaloni A, Monti M, Monteforti G, Minnocci A, Petacchi R, Pelosi P. 1999. Purification, structural characterization, cloningand immunocytochemical localization of chemoreception proteins from Schistocerca gregaria. *European Journal of Biochemistry* **262:** 745–754 DOI: 10.1046/j.1432-1327.1999.00438.x.

Benton R, Sachse S, Michnick SW, Vosshall LB. 2006. Atypical Membrane Topology and Heteromeric Function of Drosophila Odorant Receptors In Vivo. *Plos Biology* **4(2):** e20 DOI:10.1371/journal.pbio.0040020.

Briand L, Nespoulous C, Huet JC, Pernollet JC. 2010. Disulfide pairing and secondary structure of ASP1, an olfactory-binding protein from honeybee (Apis mellifera L). *Chemical Biology & Drug Design* **58:**540-545 DOI:10.1034/j.1399-3011.2001.00949.x.

Burge CB, Karlin S. 1998. Finding the genes in genomic DNA. *Current Opinion in Structural Biology* **8:**346-354 DOI: 10.1016/S0959-440X(98)80069-9.

Clarke AR, Armstrong KF, Carmichael AE, Milne JR, Raghu S, Roderick GK, Yeates DK. 2005. Invasive phytophagous pests arising through a recent tropical evolutionary radiation: The Bactrocera dorsalis complex of fruit flies. *Annual Reviews of Entomology* **50:**293–319 DOI: 10.1146/annurev.ento.50.071803.130428.

Duan GG, Li DZ, Wang MQ. 2018. Chemosensory protein used as target for screening

behaviourally active compounds in rice pest Cnaphalocrocis medinalis (Lepidoptera: Pyralidae). *Insect Molecular Biology* DOI:10.1111/imb.12532.

English AC, Richards S, Han Y, Wang M, Vee V, Qu JX, Qin X, Muzny DM, Reid JG, Worley KC, Gibbs RA. 2012. Mind the Gap: Upgrading Genomes with Pacific Biosciences RS Long-Read Sequencing Technology. *PLOS ONE* **7(11):** e47768 DOI:10.1371/journal.pone.0047768.

Elfekih S, Chen CY, Hsu JC, Belcaid M, Haymer D.2016. Identification and preliminary characterization of chemosensory perception-associated proteins in the melon fly Bactrocera cucurbitae using RNA-seq. *Scientific Reports* **6:**19112 DOI: 10.1038/srep19112.

Field LM, Pickett JA, Wadhams LJ. 2010. Molecular studies in insect olfaction. *Insect Molecular Biology* **9:**545-551 DOI: abs/10.1046/j.1365-2583.2000.00221.x.

Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, White W, Buell CB, Wortman JR. 2008. Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biology* **9:**R7 DOI:10.1186/gb-2008-9-1-r7.

Hekmatscafe DS, Scafe CR, Mckinney AJ, Tanouye MA. 2002. Genome-Wide Analysis of the Odorant-Binding Protein Gene Family in Drosophila melanogaster. *Genome Research* **12:**1357-1369 DOI: 10.1101/gr.239402.

Holt C, Yandell M. 2011. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics* **12:**491 DOI:10.1186/1471-2105-12-491.

462 **Jacquinjoly E, Vogt RG, François MC, Nagnanle MP. 2001.** Functional and expression
463 pattern analysis of chemosensory proteins expressed in antennae and pheromonal gland
464 of Mamestra brassicae. *Chemical Senses* ***26*****:** 833-844 DOI:10.1093/chemse/26.7.833.
465 **Jin T, Zeng L, Lin YY, Lu YY, Liang G. 2011.** Insecticide resistance of the oriental fruit
466 fly, Bactrocera dorsalis (Hendel) (Diptera: Tephritidae), in mainland China. *Pest*
467 *Management Science* **67:** 370–376 DOI:10.1002/ps.2076.
468 **Jones WD, Nguyen TA, Kloss B, Lee KJ, Vosshall LB. 2005.** Functional conservation of
469 an insect odorant receptor gene across 250 million years of evolution. *Current Biology*
470 *Cb* **15:**R119-R121 DOI: 10.1016/j.cub.2005.02.007.
471 **Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. 2013.** TopHat2:
472 accurate alignment of transcriptomes in the presence of insertions, deletions and gene
473 fusions. *Genome Biology* **14:**R36 DOI: 10.1186/gb-2013-14-4-r36.
474 **Krieger J, Nickischrosenegk EV, Mameli M, Pelosi P, Breer H. 1996.** Binding proteins
475 from the antennae of Bombyx mori. *Insect Biochemistry & Molecular Biology* **26:** 297
476 DOI:10.1016/0965-1748(95)00096-8.
477 **Langmead B, Trapnell C, Pop M, Salzberg SL. 2009.** Ultrafast and memory-efficient
478 alignment of short DNA sequences to the human genome. *Genome Biology and Evolution*
479 **10:**R25 DOI: 10.1186/gb-2009-10-3-r25.
480 **Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H,**
481 **Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG.**
482 **2007.** Clustal W and Clustal X version 2.0. *Bioinformatics* **23:** 2947-2948 DOI:
483 10.1093/bioinformatics/btm404.

Larsson MC, Domingos AI, Jones WD, Chiappe ME, Amrein H, Vosshall LB. 2004. Or83b encodes a broadly expressed odorant receptor essential for Drosophila olfaction. *Neuron* **43:**703-714 DOI: 10.1016/j.neuron.2004.08.019.

Lartigue A, Campanacci V, Roussel A, Larsson AM, Jones TA, Tegoni M, Cambillau C. 2002. X-ray Structure and Ligand Binding Study of a Moth Chemosensory Protein. *Journal of Biological Chemistry* 277:32094-32098 DOI: 10.1074/jbc.M204371200.

Leal WS, Nikonova L, Peng G. 1999. Disulfide structure of the pheromone binding protein from the silkworm moth, Bombyx mori.*FEBS Lett* **464:**85–90 DOI:10.1016/S0014-5793(99)01683-X.

Leal WS. 2013. Odorant reception in insects: roles of receptors, binding proteins, and degrading enzymes. *Annual Review of Entomology* ***58*:**373-391 DOI 10.1146/annurev-ento-120811-153635.

Liu GX, Ma HM, Xie HY, Xuan N, Guo X, Fan ZX, Rajashekar B, Arnaud P, Offmann B, Picimbon JF. 2016. 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 DOI:10.1371/journal.pone.0154706.

Liu Z, Smagghe G, Lei Z, Wang JJ. 2016. Identification of male- and female-specific olfaction genes in antennae of the oriental fruit fly (Bactrocera dorsalis). *PLOS ONE* **11(2):**e0147783 DOI 10.1371/journal.pone.0147783.

Liu G, Ma H, Xie H, Xuan N, Guo X, Fan Z, Rajashekar B, Arnaud P, Offmann B, Picimbon JF. 2016. 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 DOI:10.1371/journal.pone.0154706.

Luo RB, Liu BH, Xie YL, Li ZY, Huang WH, Yuan JY, He GZ, Chen YX, Pan Q, Liu YJ, Tang JB, Wu GX, Zhang H, Shi YJ, Liu Y, Yu C, Wang B, Lu Y, Han CL, Cheung DW, Yiu SM, Peng SL, Zhu XQ, Liu GM, Liao XK, Li YR, Yang HM, Wang J, Lam TW, Wang J. 2012. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *GigaScience* **1:**18 DOI:10.1186/2047-217x-1-18.

Majoros WH, Pertea M, Salzberg SL. 2004. TigrScan and GlimmerHMM: two open-source ab initio eukaryotic gene-finders. *Bioinformatics* **20**:2878-2879 DOI:10.1093/bioinformatics/bth315.

Maleszka J, Forêt S, Saint R, Maleszka R. 2007. RNAi-induced phenotypes suggest a novel role for a chemosensory protein CSP5 in the development of embryonic integument in the honeybee (Apis mellifera). *Development Genes and Evolution* **217(3):**189-96 DOI: 10.1007/s00427-006-0127-y.

Mortazavi A, Williams BA, Mccue K, Schaeffer L, Wold B. 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods* **5:**621-628. DOI: 10.1038/nmeth.1226.

Nicholas KB, Nicholas Jr HB, Deerfield II DW. 1997. GeneDoc: Analysis and visualization of genetic variation. *Embnew News* **4:**14.

Nishikawa H, Iijima T, Kajitani R, Yamaguchi J, Ando T, Suzuki Y, Sugano S, Fujiyama A, Kosugi S, Hirakawa H, Tabata S, Ozaki K, Morimoto H, Ihara K, Obara M, Hori H, Itoh T, Fujiwara H. 2015. A genetic mechanism

528 for female-limited Batesian mimicry in Papilio butterfly. *Nature Genetick* **47(4):**405-9
529 DOI: 10.1038/ng.3241.

530 **Ozaki M, Wada-Katsumata A, Fujikawa K, Iwasaki M, Yokohari F, Satoji Y, Nisimura**
531 **T, Yamaoka R. 2005.** Ant nestmate and nonnestmate discrimination by a chemosensory
532 sensillum. *Science* **309:** 311–314 DOI: 10.1126/science.1105244.

533 **Pelosi P, Maida R. 1995.** Odorant-binding proteins in insects. *Comparative Biochemistry*
534 *and Physiology Part B: Biochemistry and Molecular Biology* **111:**503
535 DOI:10.1016/0305-0491(95)00019-5.

536 **Pelos P, Calvello M, Ban LP. 2005.** Diversity of Odorant-binding Proteins and
537 Chemosensory Proteins in Insects. Chemical Senses 30: i291–i292
538 DOI:10.1093/chemse/bjh229.

539 **Robertson HM, Wanner KW. 2006.** The chemoreceptor superfamily in the honey bee,
540 Apis mellifera: expansion of the odorant, but not gustatory, receptor family. *Genome*
541 *Research* **16:**1395 DOI: 10.1101/gr.5057506.

542 **Robertson HM, Warr CG, Carlson, JR. 2003.** Molecular evolution of the insect
543 chemoreceptor gene superfamily in Drosophila melanogaster. *Proceedings of the*
544 *National Academy of Sciences of the United States of America* **100 Suppl 2:**14537 DOI:
545 10.1073/pnas.2335847100.

546 **Robinson MD, McCarthy DJ, Smyth GK. 2010.** edgeR: a Bioconductor package for
547 differential expression analysis of digital gene expression data. *Bioinformatics* **26:**139-
548 140 DOI: 10.1093/bioinformatics/btp616.

549 **Siciliano P, Scolari F, Gomulski LM, Falchetto M, Manni M, Gabrieli P, Field LM,**

Zhou JJ, Gasperi GL, Malacrida AR. 2014. Sniffing out chemosensory genes from the mediterranean fruit fly, Ceratitis capitata. *PLOS ONE* **9(1):** e85523 DOI 10.1371/journal.pone.0085523.

Smith DP. 2007. Odor and pheromone detection in Drosophila melanogaster. *European Journal of Physiology* **454:**749-758 DOI: 10.1007/s00424-006-0190-2.

Stanke M, Diekhans M, Baertsch R, Haussler D. 2008. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics* **24 (5):** 637-644 DOI: 10.1093/bioinformatics/btn013.

Stephens A, Kriticos D, Leriche A. 2007. The current and future potential geographical distribution of the oriental fruit fly, Bactrocera dorsalis (Diptera: Tephritidae). Bulletin of Entomological Research **97:**369–378 DOI: 10.1017/S0007485307005044.

Sun YL, Huang LQ, Pelosi P, Wang CZ. 2012. Expression in antennae and reproductive organs suggests a dual role of an odorant-binding protein in two sibling Helicoverpa species. *Plos One* **7:**e30040 DOI: 10.1371/journal.pone.0030040.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Molecular Biology and Evolution* **30:**2725-2729 DOI:10.1093/molbev/mst197.

Tegoni M, Campanacci V, Cambillau C. 2004. Structural aspects of sexual attraction and chemical communication in insects. *Trends in Biochemical Sciences* **29:**257–264 DOI: 10.1016/j.tibs.2004.03.003.

Venthur H, Zhou JJ. 2018. Odorant receptors and odorant-binding proteins as insect pest control targets: a comparative analysis. *Frontiers in Physiology* **9:**1163 DOI

10.3389/fphys.2018.01163.

Vieira FG, Rozas J. 2011. Comparative Genomics of the Odorant-Binding and Chemosensory Protein Gene Families across the Arthropoda: Origin and Evolutionary History of the Chemosensory System. Genome Biology and Evolution **3:**476–490 DOI 10.1093/gbe/evr033.

Vogt RG, Callahan FE, Rogers ME, Dickens JC. 1999. Odorant binding protein diversity and distribution among the insect orders, as indicated by LAP, an OBP-related protein of the true bug Lygus lineolaris (Hemiptera, Heteroptera). *Chemical Senses* **24:** 481-495 DOI: 10.1093/chemse/24.5.481.

Vosshall LB. 2003. Diversity and expression of odorant receptors in Drosophila. *Insect Pheromone Biochemistry and Molecular Biology* **18:**567-591 DOI: 10.1016/B978-012107151-6/50021-9.

Wu ZZ, Zhang H, Wang ZB, Bin SY, He HH, Lin JT. 2015. Discovery of Chemosensory Genes in the Oriental Fruit Fly, Bactrocera dorsalis. *Plos One* **10(6):**e0129794 DOI 10.1371/journal.pone.0129794.

Xia YL, Ma XL, Hou BH, Ouyang GC. 2018. A Review of Bactrocera minax (Diptera:Tephritidae) in China for the Purpose of Safeguarding. Advances in Entomology **6:**35-61 DOI 10.4236/ae.2018.62005.

Zhao Y, Ding J, Zhang Z, Liu F, Zhou C, Mu W. 2018. Sex- and Tissue-Specific Expression Profiles of Odorant Binding Protein and Chemosensory Protein Genes in Bradysia odoriphaga (Diptera: Sciaridae). *Frontiers in Physiology* **9:**107 DOI: 10.3389/fphys.2018.00107.

Zheng WW, Peng W, Zhu CP, Zhang Q, Saccone G, Zhang HY. 2013. Identification and Expression Profile Analysis of Odorant Binding Proteins in the Oriental Fruit Fly Bactrocera dorsalis. *International Journal of Molecular Sciences* **14:**14936-14949 DOI: 10.3390/ijms140714936.

603 **Table 1 Identification of OBP genes in Tephritidae.**

604

605 **Table 2 Sequence homology of CSPs in Tephritidae.**

606

607 **Table 3 Identification of OR genes in Tephritidae.**

608

609 **Figure 1 Phylogenetic tree of OBPs in Tephritidae.** NOTE: The four-letter codes used for the
610 species are *Bactrocera minax* (Bmin), *Bactrocera dorsalis* (Bdor), *Bactrocera cucurbitae* (Bcuc),
611 *Bactrocera latifrons* (Blat), *Bactrocera oleae* (Bole), *Ceratitis capitata* (Ccap), *Rhagoletis zephyria*
612 (Rzep), and *Drosophila melanogaster* (Dmel). A total of 37 OBP genes were identified in Tephritidae,
613 namely, 8a, 19a, 19b, 19c, 19d, 28a, 44a, 46a, 47a, 47b, 50a, 50c, 50e, 56a, 56b, 56c, 56d, 56e, 56g, 56h,
614 57c, 58c, 58d, 59a, 69a, 73a, 83a, 83b, 83cd, 83ef, 83g, 84a, 99a, 99b, 99c, 99d, and lush. The number
615 of OBP genes present in *Drosophila melanogaster* has been previously reported (Hekmatscafe et al.,
616 2002).

617

618

619 **Figure 2 Highly expressed OBPs in *B. minax* and *B. dorsalis*.**

620

621 **Figure 3 NJ tree of the CSP family.** NOTE: The four-letter codes used for the species are
622 *Bactrocera minax* (Bmin), *Bactrocera dorsalis* (Bdor), *Bactrocera cucurbitae* (Bcuc), *Bactrocera*
623 *latifrons* (Blat), *Bactrocera oleae* (Bole), *Ceratitis capitata* (Ccap), *Rhagoletis zephyria* (Rzep),
624 *Drosophila melanogaster* (Dmel), *Drosophila simulans* (Dsim), *Drosophila sechellia* (Dsec), *Drosophila*

erecta (Dere), *Drosophila yakuba* (Dyak), *Drosophila ananassae* (Dana), *Drosophila pseudoobscura* (Dpse), *Drosophila persimilis* (Dper), *Drosophila willistoni* (Dwil), *Drosophila mojavensis* (Dmoj), *Drosophila virilis* (Dvir) and *Drosophila grimshawi* (Dgri). The number of CSP genes present in *Drosophila* has been previously reported (Vieira and Rozas, 2011).

Figure 4 Gene expression of CSPs in *B. minax* and *B. dorsalis*.

Figure 5 NJ tree of the OR family in Tephritidae. NOTE: The four-letter codes used for the species are *Bactrocera minax* (Bmin), *Bactrocera dorsalis* (Bdor), *Bactrocera cucurbitae* (Bcuc), *Bactrocera latifrons* (Blat), *Bactrocera oleae* (Bole), *Ceratitis capitata* (Ccap), *Rhagoletis zephyria* (Rzep), and *Drosophila melanogaster* (Dmel). A total of 38 OR genes were identified in Tephritidae, namely, OR2a, OR7a, OR10a, OR13a, OR22c, OR24a, OR33abc, OR35a, OR43a, OR45a, OR46aA/B, OR47b, OR49a, OR49b, OR59a, OR63a, OR67c, OR67d, OR69aA/B, OR71a, OR74a, OR82a, OR83a, OR85bc, OR85d, OR85e, OR88a, OR94ab, ORCR, OR1, OR2, OR3, OR4, OR5, OR6, OR7, OR8, and OR9. The number of OR genes present in *Drosophila melanogaster* has been previously reported (Robertson et al., 2003).

Supplemental file 1 The protein sequences and alignments of OBPs.

Supplemental file 2 The protein sequences and alignments of CSPs.

Supplemental file 3 The protein sequences and alignments of ORs.

Supplemental file 4 The sequence homology of CSP family.

Figure 1

NJ tree of the CSP family

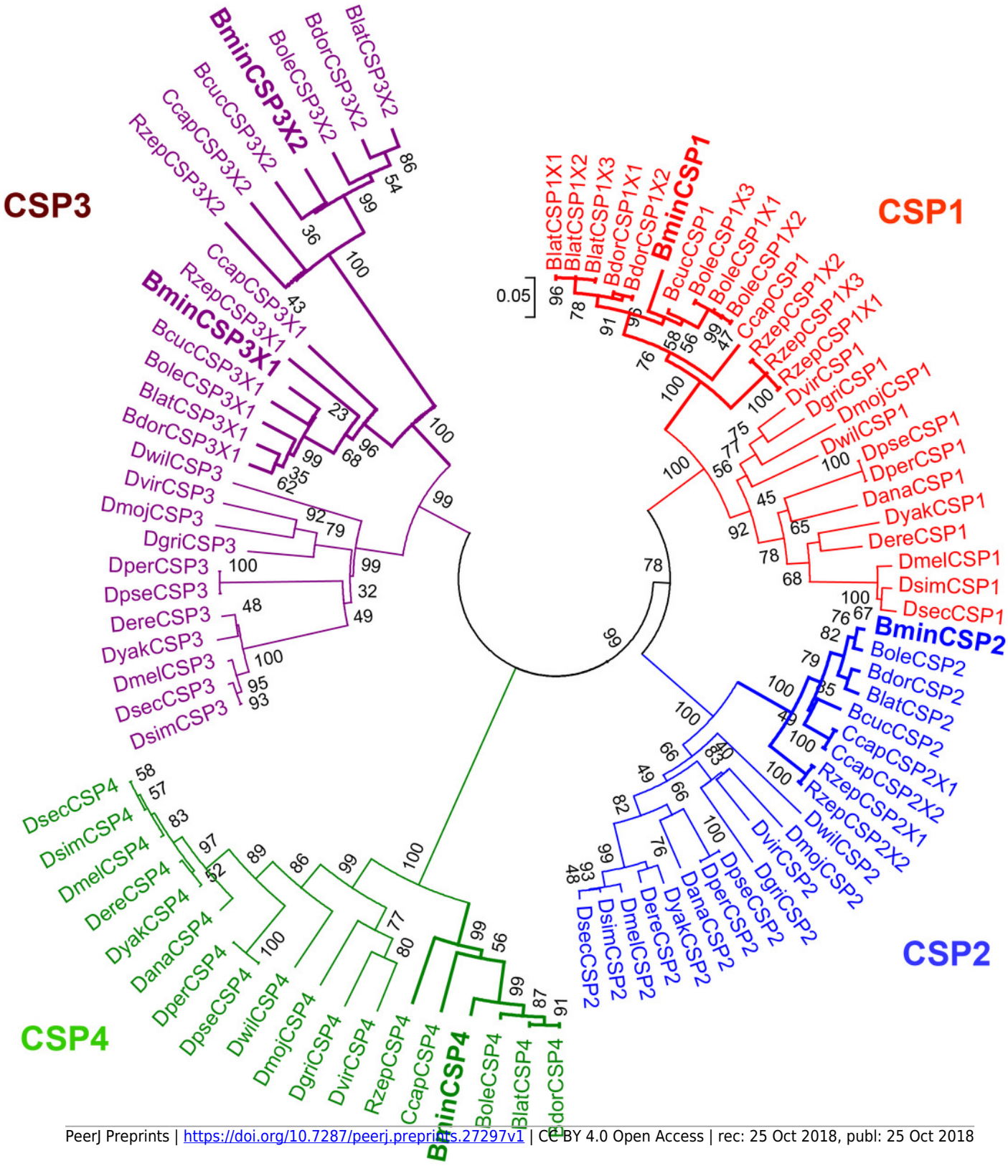

Table 1(on next page)

Identification of OBP genes in Tephritidae

1 **Table 1 Identification of OBP genes in Tephritidae**

OBPs	Tephritidae									
	Dmel	Bmin	Bdor	Bcuc	Blat	Bole	Rzep	Ccap	Tephritidae	Total
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

NOTE: The four-letter codes used for the species are *Bactrocera minax* (Bmin), *Bactrocera dorsalis* (Bdor), *Bactrocera cucurbitae* (Bcuc), *Bactrocera latifrons* (Blat), *Bactrocera oleae* (Bole), *Ceratitis capitata* (Ccap), *Rhagoletis zephyria* (Rzep), and *Drosophila melanogaster* (Dmel). A total of 25, 37, 33, 35, 30, 29, 34 OBP genes were identified in Bmin, Bdor, Bcuc, Blat, Bole, Rzep and Ccap, respectively.

Table 2(on next page)

Sequence homology of CSPs in Tephritidae

1 **Table 2 Sequence homology of CSPs in Tephritidae**

2

	CSP1	CSP2	CSP3X1	CSP3X2	CSP4
Intra-Tephritid	91%	91%	82%	80%	89%
Tephritid and fruitfly	64%	73%	50%	28%	66%

3

Table 3(on next page)

Identification of OR genes in Tephritidae

Table 3 Identification of OR genes in Tephritidae

	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	0	14	25	17	18	13	16	20	87
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

NOTE: The four-letter codes used for the species are *Bactrocera minax* (Bmin), *Bactrocera dorsalis* (Bdor), *Bactrocera cucurbitae* (Bcuc), *Bactrocera latifrons* (Blat), *Bactrocera oleae* (Bole), *Ceratitis capitata* (Ccap), *Rhagoletis zephyria* (Rzep), and *Drosophila melanogaster* (Dmel). A total of 53, 70, 61, 58, 59, 64, and 61 OR genes were identified in Bmin, Bdor, Bcuc, Blat, Bole, Rzep and Ccap, respectively.