Insect Molecular Biology

Insect Molecular Biology (2014) **23**(5), 621–631 doi: 10.1111/imb.12110

Odorant-binding protein (OBP) genes affect host specificity in a fig–pollinator mutualistic system

N. Wang*, N. X. Wang*, L. M. Niu†, S. N. Bian*, J. H. Xiao‡ and D. W. Huang*‡

**College of Plant Protection, Shandong Agricultural University, Tai'an, China; †Environment and Plant Protection Institute, Chinese Academy of Tropical Agricultural Sciences, Danzhou, Hainan, China; and ‡Key Laboratory of Zoological Systematics and Evolution, Institute of Zoology, Chinese Academy of Sciences, Beijing, China*

Abstract

The interaction between figs and their pollinating wasps is regarded as a model system for studying specialized co-evolved mutualism. Chemoreception of fig wasps plays an important role in this interaction, and odorant-binding proteins (OBP) function in the first step of odorant detection. The OBP repertoire of the fig wasp *Ceratosolen solmsi* is reported to be one of the smallest among insects; however, it is unknown how these OBPs are related to the complicated mating process occurring within the fig cavity and the extreme host specificity of the species. In the present study, we combined a structural analysis of the conserved cysteine pattern and motif order, a phylogenetic analysis, and previous studies on ligand-binding assays to deduce the function of OBPs. We also quantified the expression of OBP genes in different life stages of female and male fig wasps by using real-time quantitative PCR, which can help to predict the function of these genes. The results indicated that CsolOBP1 and CsolOBP2 (or CsolOBP5) in males may bind to pheromones and play important roles in mate choice, whereas CsolOBP4 and CsolOBP5 may primarily function in host localization by females through binding of volatile compounds emitted by receptive figs.

Keywords: fig wasps, habitat, predicting function.

First published online 10 July 2014.

Correspondence: Da-Wei Huang, College of Plant Protection, Shandong Agricultural University, Tai'an 271018, China; Key Laboratory of Zoological Systematics and Evolution, Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, China. Tel.: + 86 10 64807235; fax: + 86-10-64807235; e-mail: huangdw@ioz.ac.cn

Introduction

Chemoreception is essential for the survival and reproduction of insects through processes such as the detection of food, predators, hosts, oviposition sites and mates. There are two major chemosensory mechanisms, olfaction and taste (Sánchez-Gracia *et al.*, 2009), in which chemical signals are detected by members of multigene families that encode odorant-binding proteins (OBPs), chemosensory proteins, olfactory receptors, gustatory receptors and ionotropic receptors. OBPs participate in the first step in odour detection (Pelosi & Maida, 1990). OBP family members are highly divergent in insects, and their overall sequence identity is as low as 16.71% (Vieira & Rozas, 2011); however, almost all identified OBPs have some conserved characteristic features: a signal peptide on the N-terminal, a secondary structure containing at least six α-helices, and a highly conserved cysteine pattern (Pelosi & Maida, 1995; Vieira & Rozas, 2011). In general, the insect OBP family has been classified into several subfamilies based on structural features, functional information, and phylogenetic relationships: Classic, Minus-C, Plus-C, Dimer, PBP/GOBP, ABPI and ABPII, CRLBP, D7, and Double-domain (Vieira & Rozas, 2011; Vieira *et al.*, 2012). Most OBPs in insects have six cysteines and belong to the Classic subfamily (Pelosi & Maida, 1995; Vogt, 2005). The six-cysteine pattern is a common characteristic of OBPs in the Classic subfamily. For example, most lepidopterous and *Drosophila melanogaster* OBPs contain C1-X_{30}-C2-X_3-C3-X_{42}-C4-$X_{8–10}$-C5-X_8-C6-$X_{24–26}$ [C1–6 represents the conserved six cysteines and X_n represents any n amino acids (Field *et al.*, 2000)]. Subsequent studies have shown that the cysteine patterns of OBPs are similar within the same insect order; for example, a conserved pattern of C1-$X_{23–35}$-C2-X_3-C3-$X_{27–45}$-C4-$X_{7–14}$-C5-X_8-C6 exists in hymenopteran OBPs. There are marked differences,

however, in the number of amino acids between neighbouring cysteines; for instance, in Hymenoptera, the distance between the first and second cysteines has the highest variability, whereas in other orders, the highest variability occurs between the fourth and fifth cysteines (Xu *et al*., 2009).

At the protein sequence level, conserved motifs in the OBPs are important elements of functional domains (Xu *et al*., 2009). For example, although general OBPs (GOBPs) and pheromone-binding proteins (PBPs) in Lepidoptera have the same eight motifs, the motif patterns are quite different: all PBPs have a motif pattern of 6-1-2-8-3-4-5-7 and most GOBPs are 7-6-1-2-8-3-4-5. In addition, the motif patterns are quite different among different insect orders. For example, of the eight motifs in the Lepidoptera GOBPs, only two are found in Diptera, and the motif patterns of PBPs are also different between Lepidoptera and Hymenoptera; thus, it seems that the highly conserved cysteine pattern is the key structure of OBPs, whereas the motif patterns may fine-tune the functions of OBPs, leading to minor differences in the binding of diverse odour molecules (Xu *et al*., 2009).

Given the high divergence in OBP sequences, it is difficult to study the phylogenetic relationships among all insect OBPs. Previous phylogenetic analyses of OBPs across Hymenoptera showed that the evolution of OBP families is highly dynamic, exhibiting both species-specific branches and some groups of orthologous sequences. For example, Gp-9 sequences from 18 species of *Solenopsis* are clustered together (Krieger & Ross, 2005; Pelosi *et al*., 2006), and the honey bee *Apis mellifera* has extensive lineage-specific gene duplications of OBPs (Sánchez-Gracia *et al*., 2009).

The system of figs (*Ficus*: Moraceae) and their pollinating fig wasps (Agaonidae: Hymenoptera) is widely regarded as a model for the study of co-evolved mutualism (Anstett *et al*., 1997). The inflorescences of the fig inflate and form round, oval or pear-shaped fruits. The florets that develop in the intine form a syconium, which comprises the fig. Pollinating fig wasps only develop in the closed and dark galls formed from the fig ovaries within the syconium, and each wasp lives in its own gall. As in other insects, chemoreception plays a key physiological role in wasps' life history, particularly in mating and host specificity. The fig wasp is an example of protandry. When wasps mature, the males always emerge from the galls earlier than females; they emerge into the syconium, find a gall with a female inside, and mate with her (Galil & Eisikowitch, 1968). Because there are many others wasps (such as non-pollinating wasps) in the same fig, how do the male wasps exclude interference from other wasps and effectively find their mates in the dark figs? Clearly, the process by which the male wasps perceive the pheromones released by the female wasps of the same species is critical. The mated female wasps then fly away from their natal figs to search for a suitable fig to complete their oviposition and pollination. The high degree of specificity is presented by a phenomenon that a female is only attracted to its specific fig tree species through the volatile chemicals released by the figs when they are ready to be pollinated (Hossaert-McKey *et al*., 1994; Ware & Compton, 1994). The within-tree fruiting synchrony of the figs and the short lifespan of the wasps means that the female wasps must usually leave their natal trees to find figs that are suitable for oviposition, the wasps enter into next generation, and the generations can be extended (Ware & Compton, 1994). The precise seeking and identification of figs is therefore crucial for female wasps. By contrast, male wasps usually do not leave the dark fig in which they develop (Ramírez, 1974).

The diecious fig syconia of *Ficus hispida* provide shelter to only four fig wasp species: one pollinating wasp *Ceratosolen solmsi* and the three non-pollinators *Philotrypesis pilosa*, *Philotrypesis* sp., and *Apocrypta bakeri* (Abdurahiman & Joseph, 1978). The volatile compounds released throughout the fig development cycle vary dynamically, and the quantity and composition of volatile compounds emitted vary between female and male figs, but at the receptive phase, the volatile compounds of female and male figs are similar (Proffit *et al*., 2008). This is the phase in which the female wasps are attracted to the figs through the release of volatile chemicals when the figs are ready to be pollinated.

In the published genome data of *C. solmsi* (Xiao *et al*., 2013), the gene repertoires in the chemoreception are annotated: seven OBPs, seven chemosensory proteins, 46 olfactory receptors, 5 gustatory receptors and 11 ionotropic receptors. In particular, gustatory receptors, olfactory receptors and OBPs have dramatic restrictions, possibly reflecting host specificity (Xiao *et al*., 2013).

To determine the mechanisms of olfactory reception in *C. solmsi* related to mating and host specificity, we extensively analysed the structural features of the seven OBPs. We also used a real-time quantitative PCR (qPCR) method to examine the sex- and development-specific expression patterns of these *OBP* genes in two important life stages: pupae and adults.

Results

Gene identification

In total, seven putative *OBP* genes were annotated from the published *C. solmsi* genome (Table 1) using manual annotation (Xiao *et al*., 2013). We named these *OBP* genes using the '*CsolOBP*' prefix followed by a number incremented from the 5′ end of the scaffold to the 3′ end, in increasing scaffold order. Of the seven *OBP* genes, six were predicted to encode complete coding sequences,

Table 1. List of *OBP* genes in *Ceratosolen solmsi*

Gene name	Gene ID	ORF length (nt)	Protein length (aa)	Most similar homology search with known proteins				
				score	*E*-value	Identity (%)	Species	Protein ID
CsolOBP1	CSO_004664	456	151	248	1e-81	94	*Nasonia vitripenis*	XP_001606881.1
CsolOBP2	CSO_007091	435	144	158	1e-46	59	*N. vitripenis*	XP_001600769.1
CsolOBP3	CSO_009606	393	130	77.8	1e-15	31	*N. vitripenis*	XP_003428209.1
CsolOBP4	CSO_009614	399	132	127	3e-34	45	*N. vitripenis*	XP_001600962.2
CsolOBP5	CSO_009615	405	134	195	4e-61	67	*N. vitripenis*	XP_001601068.1
CsolOBP6	CSO_009620	327	108	166	2e-50	71	*N. vitripenis*	XP_001600607.1
CsolOBP7	CSO_010926	369	122	82.0	2e-17	33	*N. vitripenis*	XP_001606346.1

whereas *CsolOBP6* lacked the signal peptide at the N-terminal. Because of the highly diversified sequences of *OBP* genes and the resulting difficulty in designing degenerate primers, we only succeeded in cloning the partial sequences of *CsolOBP1* and *CsolOBP2* by using conventional PCR methods to amplify the *OBP* genes from total genomic DNA.

The length of the seven translated OBP protein sequences varied between 108 and 151 amino acids, and the sequences shared only 8.7–25.0% amino acid similarity with each other (Table S1). The CsolOBPs were homologous with known OBP sequences and were most similar to those in *Nasonia vitripennis*, with a similarity of 31–94% (Table 1).

Structural features

Except for CsolOBP6, all of the OBPs in *C. solmsi* had a signal peptide at the N-terminal with a length of 16–25 amino acids (Table 2). Regarding secondary structures, six of the proteins had at least six α-helices (Table 2). Globally, each of the amino acid sequences had six conserved cysteines (Fig. 1 and Table 2), and the cysteine pattern in the CsolOBPs was conserved: C1-$X_{25–29}$-C2-X_3-C3-$X_{27–42}$-C4-$X_{8–12}$-C5-$X_{4–8}$-C6. Eight residues were present between the fifth and sixth cysteines except in CsolOBP3 and CsolOBP6, in which only four residues were present between these cysteines. This pattern was slightly different from those reported in Hymenoptera previously (Xu *et al*., 2009).

To assess the degree of variation between neighbouring cysteines, we computed the coefficients of variation (CVs) as 5.29, 0, 13.32, 17.64 and 28.46. The highest variability, with a CV of 28.46, was observed for the distance between the fifth and sixth cysteines, followed by a CV of 5.29 for the distance between the first and second cysteines. These findings were also different from those of a previous analysis in which the highest variability was observed between the first and second cysteines in Hymenoptera (Xu *et al*., 2009).

In functional proteins, conserved motifs are important to maintain function. To explore the function of CsolOBPs, we introduced the OBP sequences of some other Hymenoptera as references to study the conserved motifs: 90 NvitOBPs in *N. vitripennis* (Vieira *et al*., 2012), nine MmedOBPs in *Microplitis mediator* (Zhang *et al*., 2009), 21 AmelOBPs in *A. mellifera* (Forêt & Maleszka, 2006), and 18 SinvOBPs in *Solenopsis invicta* (Gotzek *et al*., 2011), making a total of 145 OBPs including the seven in *C. solmsi*.

To compare the OBPs with different functions across the five species, we analysed the motifs of the 145 OBPs using the MEME online server (Fig. 2). Only four motifs were present in >50% of the input OBPs, and the motif orders in different OBPs were not identical. The order of 3-2-1-4 for the four motifs was detected in three proteins in *C. solmsi* (CsolOBP4, 5, and 7), nineteen proteins in *N. vitripennis*, three proteins in *M. mediator* (MmedOBP1, 3, and 6), six proteins in *A. mellifera* (AmelOBP4, 10, 11, 13, 14 and 21), and 10 proteins in *S. invicta*. The motif with

Table 2. Construction features of *Ceratosolen solmsi* odorant-binding proteins

Classic OBPs	Signal peptide length (aa number)	α-helices number	Cysteine pattern
CsolOBP1	24	7	C1-X_{26}-C2-X_3-C3-X_{39}-C4-X_{12}-C5-X_8-C6
CsolOBP2	25	6	C1-X_{26}-C2-X_3-C3-X_{37}-C4-X_8-C5-X_8-C6
CsolOBP3	16	6	C1-X_{25}-C2-X_3-C3-X_{42}-C4-X_8-C5-X_4-C6
CsolOBP4	21	6	C1-X_{25}-C2-X_3-C3-X_{41}-C4-X_8-C5-X_8-C6
CsolOBP5	17	7	C1-X_{27}-C2-X_3-C3-X_{38}-C4-X_8-C5-X_8-C6
CsolOBP6	N	6	C1-X_{29}-C2-X_3-C3-X_{36}-C4-X_8-C5-X_4-C6
CsolOBP7	18	5	C1-X_{27}-C2-X_3-C3-X_{27}-C4-X_8-C5-X_8-C6
Summary			C1-$X_{25–29}$-C2-X_3-C3-$X_{27–42}$-C4-$X_{8–12}$-C5-$X_{4–8}$-C6

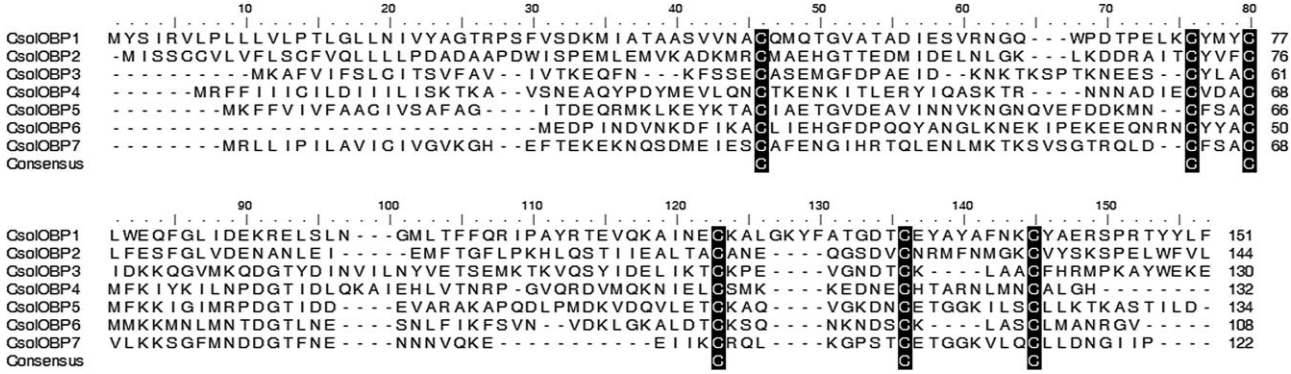
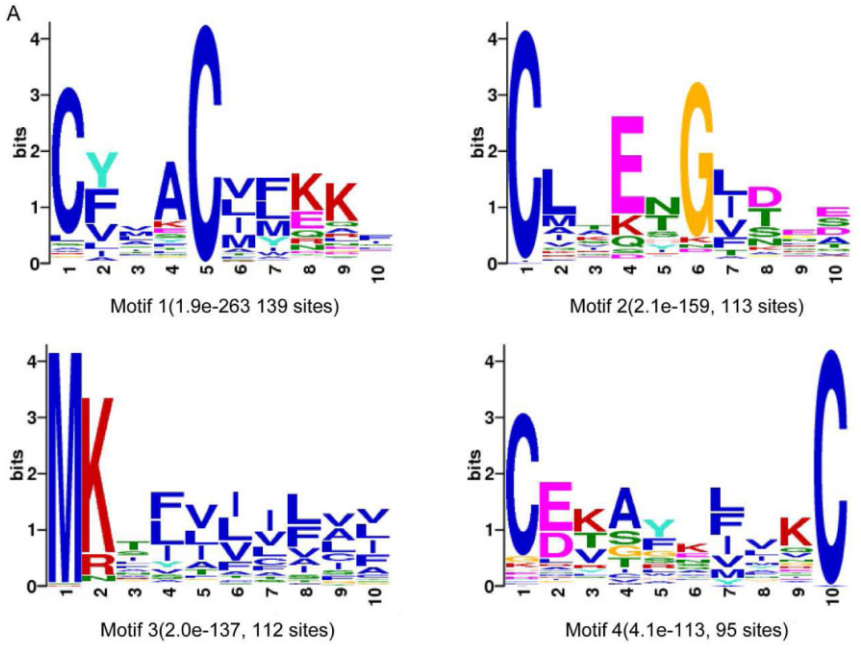


Figure 1. Alignment of *Ceratosolen solmsi* OBPs. Black boxes showed conserved cysteines residues.



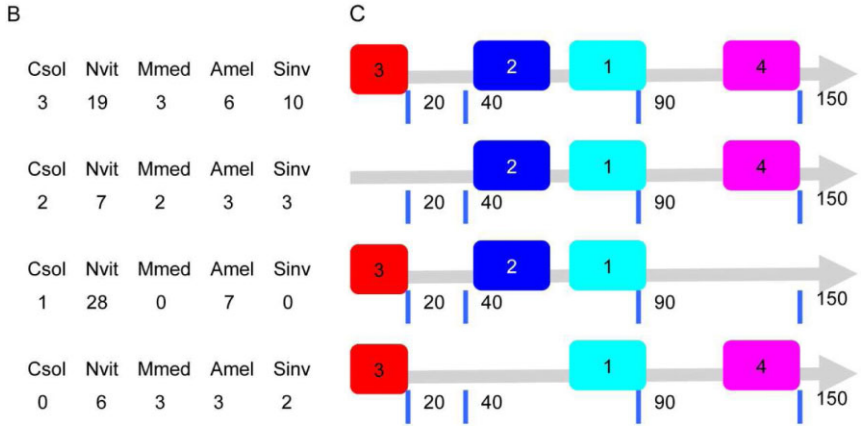


Figure 2. Motif analyses of the odorant-binding proteins (OBPs) of *Ceratosolen solmsi*, *Nasonia vitripennis*, *Microplitis mediator*, *Apis mellifera*, and *Solenopsis invicta*. (A) Motifs in the OBP proteins. (B) The numbers in the five columns are the numbers of OBPs corresponding to the MEME motif patterns shown in (C). (C) The numbers in the boxes correspond to the numbered motifs shown in (A); low numbers indicate high conservation and the numbers on the bottom indicate approximate locations of each motif on the protein sequence, starting from the N-terminal.

an order of 2-1-4, lacking motif 3, was found in two proteins in *C. solmsi* (CsolOBP1 and 2), seven proteins in *N. vitripennis* (NvitOBP2, 10, 11, 13, 70, 76, and 86), two proteins in *M. mediator* (MmedPBP1 and MmedOBP10), three proteins in *A. mellifera* (AmelOBP1, 5, and 9), and three proteins in *S. invicta* (SinvOBP1, 3, and 11). There was a specific motif order of 3-2-1, with motif 4 lost, in *C. solmsi* (CsolOBP3), *N. vitripennis* and *A. mellifera* (AmelOBP2, 15, 16, 17, 18, 19, and 20). In addition, there was a motif order of 3-1-4, with motif 2 lost, in all of the analysed species except for *C. solmsi*.

Phylogenetic analysis

The phylogenetic analysis was based on all of the OBPs of *C. solmsi, M. mediator, N. vitripennis*, *A. mellifera* and *S. invicta* (Fig. 3). Seven CsolOBPs were scattered in different branches and each CsolOBP was in the same branch as the homologous protein in *N. vitripennis*. Except for *C. solmsi*, all of the other species showed species-specific branches and/or lineage-specific gene duplications. There were no species-specific branches in *M. mediator*, but our results identified a lineage-specific gene duplication event in *M. mediator* (Fig. 3, MmedOBP1 and MmedOBP6).

Notably, two groups of orthologous sequences were shared across Hymenoptera: CsolOBP1-NvitOBP02-AmelOBP10-SinvOBP10, and MmedPBP1-CsolOBP2-NvitOBP79-AmelOBP1-SinvOBP1.

Gene expression patterns

To further determine the putative function of the OBPs in the fig wasp *C. solmsi*, we used real-time qPCR to compare the relative transcript levels of each gene in different sexes and developmental stages including pupae and adults (virgin adults: mature but still remaining inside galls; mated adults: emerged from galls inside the fig syconium) (Fig. 4).

First, we compared the *OBP* gene expression patterns between female and male wasps in the same developmental stage. No significant differences were observed at the pupal stage, but in the adult stages, significant differences were observed between females and males (in most cases, $P < 0.01$; $P < 0.05$ for *CsolOBP3*). It is interesting that in the mated adult stages, two of the genes (*CsolOBP4* and *CsolOBP5*) presented higher expression in females than in males, whereas the other five genes showed the opposite pattern.

We then compared the expression patterns among different developmental stages in females and males (Fig. 4 and Table 3). In females, all seven genes showed significantly different expression between the pupal and adult stages, with six genes having higher expression in adults, whereas *CsolOBP6* had significantly higher expression during the pupal stage ($P < 0.001$). Four genes (*CsolOBP1*, *3*, *4*, and *7*) had significantly higher expression in the adult stages than in the pupal stage ($P < 0.001$), but the highest expression level was in the mated females when they emerged from the galls. In the four genes, *CsolOBP4* was an extreme case for which the transcription level in mated females was 147.0-fold higher than that in the pupal stage. *CsolOBP2* and *CsolOBP5* also had a significantly higher expression level in adults than in pupae; however, the difference was only significant between mated females and pupal females (~35.4-fold and ~3579.3-fold higher than in pupae for *CsolOBP2* and *CsolOBP5*, respectively, $P < 0.001$); no significant difference was observed between pupal and virgin females ($P = 0.665$ and $P = 1.000$).

For males, the expression patterns were more complicated. *CsolOBP3* and *CsolOBP4* had the highest expression levels in the virgin stage, i.e., approximately 4226.5- and 11697.8-fold higher than in the pupal stage, respectively, and ~134.8- and ~93.3-fold higher than in the mated stage, respectively; there was no significant difference between the pupal and mated stages ($P = 0.967$ and $P = 0.878$). *CsolOBP1*, *2*, and *5* had the highest expression levels in the mated stage ($P < 0.001$), i.e. ~599.4-, ~196.0-, and ~2145.0-fold higher than that in the pupal stage, respectively; no significant difference existed in the pupal and virgin stages ($P = 0.979$, $P = 0.290$, and $P = 0.880$). *CsolOBP6* and *CsolOBP7* had significantly higher expression in the adult stages than in the pupal stage ($P < 0.001$), and the highest expression was observed in mated wasps.

Discussion

The fig wasp *C. solmsi* specifically colonizes and pollinates *F. hispida*. This obligate host specificity may be reflected by a reduction of its gene repertoire in the chemosensory system. In previous studies we have found that the genome of *C. solmsi* contains only seven OBP members, representing a dramatic reduction in variety, thus resulting in one of the smallest OBP families among reported insects (Xiao *et al*., 2013). Another insect that specifically parasitizes human bodies is *Pediculus humanus* which expresses only five OBPs (Kirkness *et al*., 2010). *Copidosoma floridanum*, an endoparasitoid that oviposits into the eggs of the moth *Trichoplusia ni*, expresses only eight OBPs (Donnell & Strand, 2006). Another insect, *Agrilus planipennis*, which mainly parasitizes trees of the genus *Fraxinus* and whose larvae eat the phloem, cambium and xylem inside the trunk, expresses nine OBPs (Mamidala *et al*., 2013). The habitats of these four species are relatively concealed and homogenous. By contrast, many other species live in

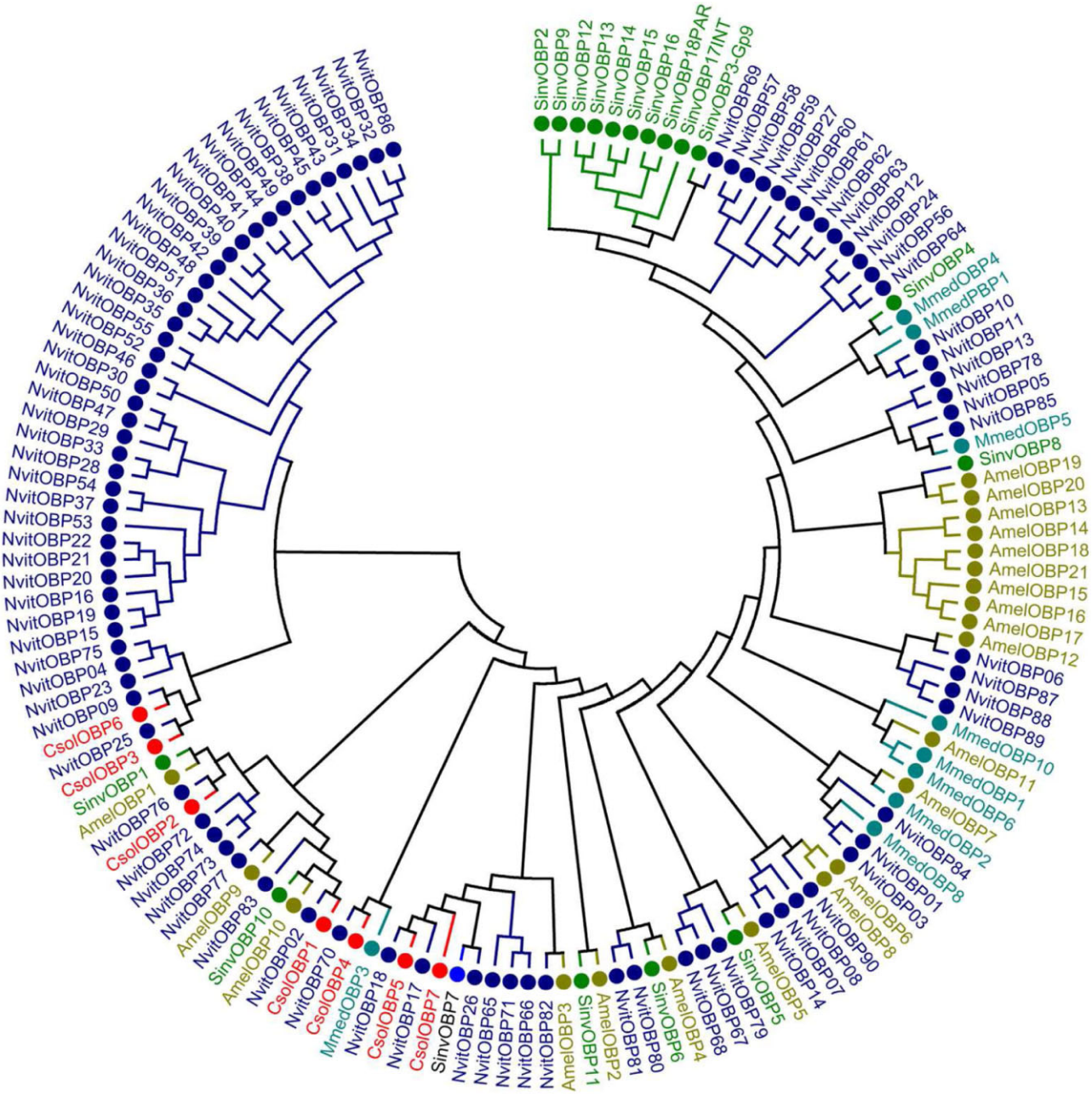


Figure 3. Phylogenetic tree of odorant-binding proteins (OBPs) in Hymenoptera. The tree was constructed using the maximum likelihood method based on the Poisson correction model. OBP names are presented as a four-letter code (first letter of the genus followed by the first three letters of the species name) + OBP + numbers. The OBPs for different insect species are indicated by coloured markers: *Ceratosolen solmsi* (Csol) in red, *Nasonia vitripennis* (Nvit) in navy blue, *Microplitis mediator* (Mmed) in cyan, *Apis mellifera* (Amel) in olive, and *Solenopsis invicta* (Sinv) in green.

Figure 4. *CsolOBP* expression patterns (relative to stage 1) of female and male *Ceratosolen solmsi* in different development stages. The *x*-axis indicates the stages of development (pupae; virgin adults: developmentally mature, but collected when they were inside galls; mated adults: mature and emerged from the galls), the *y*-axis shows fold-increase in expression (mean ± SEM). Different letters (upper case: female; lower case: male) above each bar indicate a statistical differences ($P < 0.05$) and the same letters indicate no differences between mean expression levels. * indicates a significant difference between females and males at the 0.05 level ($P < 0.05$). ** indicates extremely significant difference between females and males at the 0.01 level ($P < 0.01$).

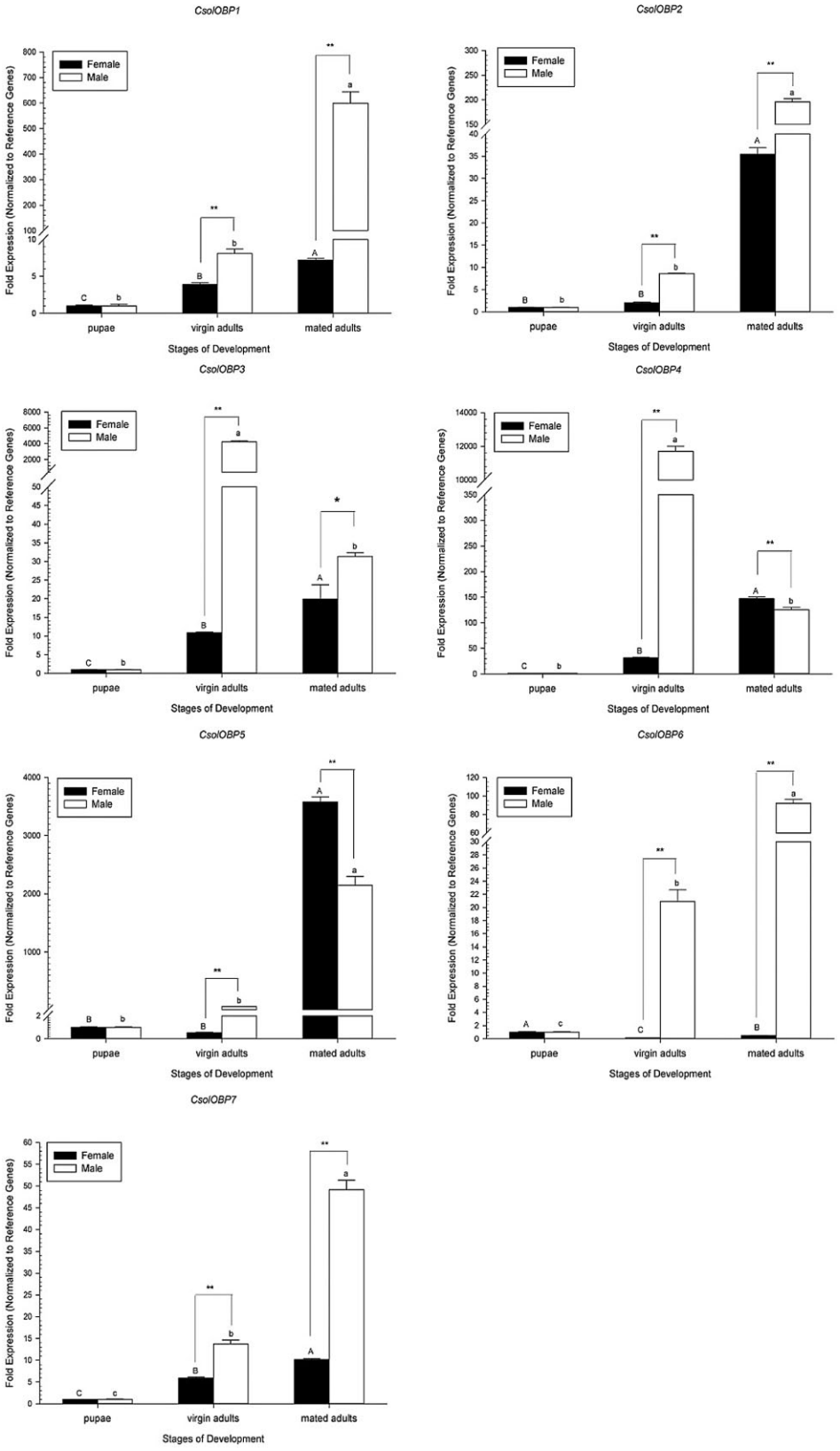

CsolOBP1
CsolOBP2
CsolOBP3
CsolOBP4
CsolOBP5
CsolOBP6
CsolOBP7
Female
Male
Fold Expression (Normalized to Reference Genes)
pupae
virgin adults
mated adults
Stages of Development

Table 3. Statistical analysis of *OBP* genes expression in *Ceratosolen solmsi*

Tukey HSD			Gene						
			CsolOBP1	*CsolOBP2*	*CsolOBP3*	*CsolOBP4*	*CsolOBP5*	*CsolOBP6*	*CsolOBP7*
Female	pupae	virgin adults	***P* < 0.001**	*P* = 0.665	***P* = 0.012**	***P* < 0.001**	*P* = 1.000	***P* < 0.001**	***P* < 0.001**
	pupae	mated adults	***P* < 0.001**	***P* < 0.001**	***P* < 0.001**	***P* < 0.001**	***P* < 0.001**	***P* < 0.001**	***P* < 0.001**
	virgin adults	mated adults	***P* < 0.001**	***P* < 0.001**	***P* = 0.005**	***P* < 0.001**	***P* < 0.001**	***P* = 0.009**	***P* < 0.001**
Male	pupae	virgin adults	*P* = 0.979	*P* = 0.290	***P* < 0.001**	***P* < 0.001**	*P* = 0.880	***P* < 0.001**	***P* < 0.001**
	pupae	mated adults	***P* < 0.001**	***P* < 0.001**	*P* = 0.967	*P* = 0.878	***P* < 0.001**	***P* < 0.001**	***P* < 0.001**
	virgin adults	mated adults	***P* < 0.001**	***P* < 0.001**	***P* < 0.001**	***P* < 0.001**	***P* < 0.001**	***P* < 0.001**	***P* < 0.001**

Terms in bold indicate $P < 0.05$.

open environments and their habitats are complicated, such as *S. invicta, A. mellifera, Tribolium castaneum, D. melanogaster, Plutella xylostella, Anopheles gambiae, N. vitripennis, Culex quinquefasciatus and Aedes aegypti.* The number of OBPs in these species are 18, 21, 50, 51, 64, 69, 90, 109 and 111, respectively (Hekmat-Scafe *et al*., 2002; Forêt & Maleszka, 2006; Tribolium Genome Sequencing *et al*., 2008; Gotzek *et al*., 2011; Vieira *et al*., 2012; Manoharan *et al*., 2013; You *et al*., 2013); therefore, we suggest that insects whose habitats are relatively concealed and homogenous express fewer OBPs than insects living in open habitats.

The presence of conserved cysteines is a typical feature of OBPs that allows them to maintain their secondary structure. The conserved cysteine pattern in *C. solmsi* OBPs is C1-$X_{25–29}$-C2-X_3-C3-$X_{27–42}$-C4-$X_{8–12}$-C5-$X_{4–8}$-C6, which is only slightly different from the previously reported conserved cysteine pattern of C1-$X_{23–35}$-C2-X_3-C3-$X_{27–45}$-C4-$X_{7–14}$-C5-X_8-C6 in Hymenoptera (Xu *et al*., 2009). This suggests that the cysteine pattern is the key structure of OBPs and should thus be highly conserved (Pelosi & Maida, 1995). In most previously studied OBPs in insects, eight residues were located between the fifth and sixth cysteines (Field *et al*., 2000; Xu *et al*., 2009); however in the present study, we detected two unusual proteins, CsolOBP3 and CsolOBP6, in *C. solmsi* in which the highest CVs was in the four amino acids between the fifth and sixth cysteines. These results indicate that the cysteine patterns of OBPs in *C. solmsi* may be still evolving.

At present, there is less available research on OBPs in Hymenoptera than in Lepidoptera and little is known regarding the function of OBPs in Hymenoptera; however, the composition and organization patterns of conserved motifs in OBPs may be associated with the fine-tuning of protein function (Xu *et al*., 2009). We can thus predict the functions of OBPs by studying the motif patterns. Previous ligand-binding assays revealed that MmedOBP1 and MmedOBP3 are GOBPs in *M. mediator* (Zhang *et al*., 2011), and the motif analyses showed that both OBPs contained the motif order of 3-2-1-4; AmelOBP2 (also called ASP2) was also found to be a GOBP (Lescop *et al*., 2009) and contained the motif order of 3-2-1. We therefore speculate that the OBPs of CsolOBP3, 4, 5 and 7 in *C. solmsi*, containing the motif orders of 3-2-1-4 and 3-2-1, can bind to general odorants. AmelOBP1 (also called ASP1) binds at least one of the major queen pheromone components (Danty *et al*., 1999), which contain the motif order of 2-1-4; thus, CsolOBP1 and CsolOBP2 in *C. solmsi*, which contain the motif order of 2-1-4, may bind to pheromones. Such rules may also apply to predict the OBP function of other insect species.

Phylogenetic analyses of the OBPs in hymenopteran insects showed that two groups of orthologous sequences were shared across Hymenoptera: CsolOBP1-NvitOBP02-AmelOBP10-SinvOBP10 and CsolOBP2-NvitOBP79-AmelOBP1-SinvOBP1. Among them, NvitOBP02 and NvitOBP79 are known as PBPs (Werren *et al*., 2010), and AmelOBP1 is also a PBP (Danty *et al*., 1999). The chemical structures of plant volatile compounds were complex and belonging to different chemical classes. In contrast to plant volatile compounds, insect sex pheromones were similar to a certain extent, especially the female sex pheromones. As a result, during the evolution of Hymenoptera, some homologous PBP genes may have been retained in each species within this lineage. Considering their phylogenetic homologous relationships with other PBPs, CsolOBP1 and CsolOBP2 are further speculated to be PBPs in *C. solmsi*.

Among fig pollinator species, the male wasps often emerge from their galls earlier than the females. The emerged males then chew small holes in the walls of galls containing females to enable mating. After mating, the males may help to enlarge the mating holes to enable the females to emerge into the central cavity of the fig (Galil & Eisikowitch, 1968) and so play a leading role in the mating process. In the expression profiling of the *OBP* genes of male fig wasps in different developmental stages, we detected that five *OBP* genes (*CsolOBP1, CsolOBP2, CsolOBP5, CsolOBP6* and *CsolOBP7*) were expressed at a dramatically higher level when the mature males were in the process of mating or searching for mates (emerging

from the galls into the fig syconia) than in the pupal stage, among which, three genes (*CsolOBP1, CsolOBP2* and *CsolOBP5*) showed extreme upregulation of >190-fold. These findings indicate that the products of these three genes highly expressed in mated males may bind to pheromones released from females. In summary, by combining the results of motif order and phylogenetic and gene expression analyses, we suggest that CsolOBP1 and CsolOBP2 (or CsolOBP5) may participate in the sensing of female pheromones in male fig wasps.

Host-specificity of fig wasps is demonstrated by the adult females. Mated females will fly away from their natal figs and search for new host figs mainly based on the volatile compounds emanating from the figs (Ramírez, 1970). Expression analysis in females indicated that six of the seven *OBP* genes (except for *CsloOBP6*) had the highest expression levels when the females emerged from the galls, especially *CsolOBP4* and *CsolOBP5*, both of which were also the only two genes that presented higher expression in females than in males at this stage. The results suggest that *CsolOBPs*, especially *CsolOBP4* and *CsolOBP5*, have vital functions in the process of host searching by detecting the volatile compounds emitted by the figs when they are receptive. Because mated pollinating wasps immediately need to search for suitable figs for spawning and pollination, it is vital for the wasps to determine the quantitative and qualitative differences in volatile compound levels among the fig's developmental phases (Proffit *et al*., 2008; Borges *et al*., 2013). Temporal patterns for the emission of volatile compounds show one major peak at receptivity (Proffit *et al*., 2008), and some compounds are most abundant and specifically present in the receptive phase, such as Linalool in *F. hispida* (Proffit *et al*., 2008) and α-muurolene in *F. racemosa* (Borges *et al*., 2013). These compounds may serve as specific attractors (Proffit *et al*., 2008; Borges *et al*., 2013) that are perceived by the olfactory systems of the mated female wasps, therefore, the OBPs appear to be the key proteins that bind to these compounds.

Above all, we combined structural analysis of the conserved cysteine pattern and motif order, phylogenetic analysis, previous studies on ligand-binding assays, and real-time qPCR to deduce the function of OBPs in the fig wasp. *OBP* genes play important roles in mating and the host specificity in the fig–pollinator mutualistic system. Such forecasting methods may also apply to predict the OBP function of other insect species.

Experimental procedures

Gene identification

In the *C. solmsi* genome, multiple approaches were used to predict genes including *de novo*, homology-based, expressed sequence tag and RNA-seq-based predictions. OBPs homologous to those in the following species were collected from the National Center for Biotechnology Information (NCBI), Hymenoptera Genome Database and FlyBase databases: *N. vitripennis*, *A. mellifera*, and *D. melanogaster*. The genes encoding OBPs in the assembled genome of *C. solmsi* were identified using the TBLASTN and BLASTP tools with an *E*-value cutoff of 0.005. A BLAST search of candidate OBP genes in the NCBI nonredundant protein database confirmed the orthology. The predicted genes were confirmed to match the corresponding transcriptomic data. Iterative searches were also conducted with each new protein as a query until no new genes were identified. The OBPs selected met both of the following criteria: sequences included the region of six conserved cysteines and the N-terminus contained the signal peptides. Because *C. solmsi* had only seven *OBP* genes, it was necessary to delve further into the raw genomic reads to confirm that the absence of genes did not result from poor or incomplete assembly or annotation. We then compared the unmapped reads to the protein sequences of all OBPs of *C. solmsi*, *N. vitripennis*, and *A. mellifera* using the BLASTX tool with an *E*-value of 1e-5. The detailed methods have been described previously (Xiao *et al*., 2013).

Structural features

The putative N-terminal signal peptides of the seven *C. solmsi* OBP amino acid sequences were predicted using the SignalP 4.1 Server http://www.cbs.dtu.dk/services/SignalP/ (Petersen *et al*., 2011). The identity values of the amino acid sequences of the seven OBPs were calculated using BioEdit (Hall, 1999). The secondary structure including α-helices was predicted using the PSIPRED online server http://bioinf.cs.ucl.ac.uk/psipred/ (McGuffin *et al*., 2000).

The protein sequences were aligned using CLUSTALX 2.1 (Thompson *et al*., 1997; Larkin *et al*., 2007) with default gap penalty parameters of a gap opening of 10 and extension of 0.2 to locate six conserved cysteines. The number of amino acids between the cysteines was counted separately. To evaluate the variability in the distance between each pair of neighbouring cysteines, we calculated the CV [= standard deviation of the number of amino acids between cysteines/the mean of the number of amino acids (Xu *et al*., 2009)].

Motif analysis was conducted using the MEME version 4.9.0 online server http://meme.nbcr.net/meme/ (Bailey & Elkan, 1994). Parameter settings were: distribution of motif occurrences = 0 or 1 per sequence, minimum width = 6, maximum width = 10, maximum number of motifs to find = 6. The motifs identified in the input sequences with $P < 0.0001$ were counted.

Phylogenetic analysis

The phylogenetic analysis was established based on the OBPs of *C. solmsi* and the total OBPs of *M. mediator, N. vitripennis*, *A. mellifera* and *S. invicta*. In total, 145 protein sequences were aligned by CLUSTALX 2.1 and the phylogenetic trees were inferred using the maximum likelihood (ML) method based on the Poisson correction model in MEGA5.2 (Tamura *et al*., 2011). All positions containing gaps and missing data were eliminated. An unrooted tree was generated with 1000 bootstrap replications.

Sample preparation for real-time quantitative PCR

We collected fig fruits of *F. hispida* from Danzhou (19°52′ N, 19°52′ E), Hainan province, China in July and August 2012. The

pupae (females and males) and virgin adults (females and males) were collected from galls that did not have a hole. The mated adults (females and males) that climbed out of the galls were collected. The samples were flash-frozen in liquid nitrogen and then stored in RNAhold™ (TransGen Biotech, Beijing, China) at −20 °C. Six sampling groups were collected for each biological experiment, and, there were three biological replicates for each experiment. For each group, total RNA was extracted and cDNA was synthesized.

Primers design and verification for real-time quantitative PCR

Gene-specific primers (Table S2) were designed using Primer Premier 5.0 according to several criteria: amplicon size between 90 and 220 base pairs, and an amplification product spanning a predicted intron to avoid contamination by genomic DNA. *RPL13a* (the gene encoding ribosomal protein L13a) and *UBC* (the gene encoding ubiquitin-conjugating enzyme) were used to normalize the target genes expression and to correct for sample-to-sample variation. The primers were previously validated experimentally (Wang *et al*., 2013).

Real-time quantitative PCR and statistical analysis

Real-time qPCR was performed using the iCycler iQ RealTime PCR Detection System iQ™5 (Bio-Rad, Hercules, CA, USA). A 20-μl PCR mixture contained 10 μl of 2 × TransStart™ Green qPCR SuperMix UDG (TransGen Biotech), 7.8 μl of ddH_2O, 0.4 μl Passive Reference Dye III (50×) (TransGen Biotech), 0.8 μl of primer mix (0.2 μM), and 1 μl of cDNA template. A no-sample reaction performed in triplicate served as the negative control. Real-time qPCR reactions for all genes for each sample were performed with three technical replicates to account for variation between runs. Cycling parameters included 50 °C for 2 min (UDG incubation), 94 °C for 10 min (UDG inactivation), 40 cycles at 94 °C for 5 s, and 60 °C for 30 s.

The amplification efficiencies of the target and reference genes were found to be similar, and the absolute values of the slopes of the target and reference genes were close to zero (<0.1), therefore, relative quantification was performed using the comparative $2^{-\Delta\Delta CT}$ method (Livak & Schmittgen, 2001). All data were normalized to expression levels of reference genes from samples from the same stages, and the relative fold-change in different stages was calculated using the transcript level in the pupae as a calibrator. Thus, the relative fold-change in different stages was assessed by comparing the expression level of each OBP in other stages to that in the pupae.

The data were statistically analysed used SPSS statistics 17.0 to obtain the means and standard errors. The comparative analyses were conducted with *t*-tests for the expression of each transcript in either sex in the same stage and with one-way nested ANOVA for developmental stages in the same sex, followed by a Tukey's honestly significant difference test ($P < 0.05$).

We followed the Minimum Information for Publication of Quantitative Real-Time PCR Experiments (MIQE) guidelines (Bustin *et al*., 2009) to increase the integrity of the results. An MIQE checklist is provided in Table S3.

Acknowledgements

We sincerely thank TransGen Biotech for providing most of the experimental reagents for free. This project was supported by the National Natural Science Foundation of China (NSFC grant no. 31090253, 31101634, 31210103912), partially by a grant (No. O529YX5105) from the Key Laboratory of the Zoological Systematics and Evolution of the Chinese Academy of Sciences, and by National Science Fund for Fostering Talents in Basic Research (Special subjects in animal taxonomy, NSFC-J1210002).

References

Abdurahiman, U.C. and Joseph, K.J. (1978) Cleptoparasitism of the fig wasps (Torymidae: Chalcidoidea) in *Ficus hispida* L. *Entomon* **3**: 181–186.

Anstett, M.C., Hossaert-McKey, M. and Kjellberg, F. (1997) Figs and fig pollinators: evolutionary conflicts in a coevolved mutualism. *TREE* **12**: 94–99.

Bailey, T.L. and Elkan, C. (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol* **2**: 28–36.

Borges, R.M., Bessiere, J.M. and Ranganathan, Y. (2013) Diel variation in fig volatiles across syconium development: making sense of scents. *J Chem Ecol* **39**: 630–642.

Bustin, S.A., Benes, V., Garson, J.A., Hellemans, J., Huggett, J., Kubista, M. *et al*. (2009) The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin Chem* **55**: 611–622.

Danty, E., Briand, L., Michard-Vanhée, C., Perez, V., Arnold, G., Gaudemer, O. *et al*. (1999) Cloning and expression of a queen pheromone-binding protein in the honeybee: an olfactory-specific, developmentally regulated protein. *J Neurosci* **19**: 7468–7475.

Donnell, D.M. and Strand, M.R. (2006) Caste-based differences in gene expression in the polyembryonic wasp *Copidosoma floridanum*. *Insect Biochem Mol Biol* **36**: 141–153.

Field, L.M., Pickett, J.A. and Wadhams, L.J. (2000) Molecular studies in insect olfaction. *Insect Mol Biol* **9**: 545–551.

Forêt, S. and Maleszka, R. (2006) Function and evolution of a gene family encoding odorant binding-like proteins in a social insect, the honey bee (*Apis mellifera*). *Genome Res* **16**: 1404–1413.

Galil, J. and Eisikowitch, D. (1968) On the pollination ecology of *Ficus sycomorus* in East Africa. *Ecology* **49**: 259–269.

Gotzek, D., Robertson, H.M., Wurm, Y. and Shoemaker, D. (2011) Odorant binding proteins of the red imported fire ant, *Solenopsis invicta*: an example of the problems facing the analysis of widely divergent proteins. *PLoS ONE* **6**: e16289.

Hall, T.A. (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl Acids Symp Ser* **41**: 95–98.

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

Hossaert-McKey, M., Gibernau, M. and Frey, J. (1994) Chemosensory attraction of fig wasps to substances produced by receptive figs. *Entomol Exp Appl* **70**: 185–191.

Kirkness, E.F., Haas, B.J., Sun, W., Braig, H.R., Perotti, M.A., Clark, J.M. *et al*. (2010) Genome sequences of the human

body louse and its primary endosymbiont provide insights into the permanent parasitic lifestyle. *Proc Natl Acad Sci U S A* **107**: 12168–12173.

Krieger, M.J. and Ross, K.G. (2005) Molecular evolutionary analyses of the odorant-binding protein gene *Gp-9* in fire ants and other *Solenopsis* species. *Mol Biol Evol* **22**: 2090–2103.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H. *et al.* (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* **23**: 2947–2948.

Lescop, E., Briand, L., Pernollet, J.C. and Guittet, E. (2009) Structural basis of the broad specificity of a general odorant-binding protein from honeybee. *Biochemistry* **48**: 2431–2441.

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

Mamidala, P., Wijeratne, A.J., Wijeratne, S., Poland, T., Qazi, S.S., Doucet, D. *et al.* (2013) Identification of odor-processing genes in the emerald ash borer, *Agrilus planipennis*. *PLoS ONE* **8**: e56555.

Manoharan, M., Ng Fuk Chong, M., Vaïtinadapoulé, A., Frumence, E., Sowdhamini, R. and Offmann, B. (2013) Comparative genomics of odorant binding proteins in *Anopheles gambiae*, *Aedes aegypti*, and *Culex quinquefasciatus*. *Genome Biol Evol* **5**: 163–180.

McGuffin, L.J., Bryson, K. and Jones, D.T. (2000) The PSIPRED protein structure prediction server. *Bioinformatics* **16**: 404–405.

Pelosi, P. and Maida, R. (1990) Odorant-binding proteins in vertebrates and insects: similarities and possible common function. *Chem Senses* **15**: 205–215.

Pelosi, P. and Maida, R. (1995) Odorant-binding proteins in insects. *Comp Biochem Physiol B* **111**: 503–514.

Pelosi, P., Zhou, J.J., Ban, L.P. and Calvello, M. (2006) Soluble proteins in insect chemical communication. *Cell Mol Life Sci* **63**: 1658–1676.

Petersen, T.N., Brunak, S., von Heijne, G. and Nielsen, H. (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods* **8**: 785–786.

Proffit, M., Schatz, B., Bessière, J.-M., Chen, C., Soler, C. and Hossaert-McKey, M. (2008) Signalling receptivity: Comparison of the emission of volatile compounds by figs of *Ficus hispida* before, during and after the phase of receptivity to pollinators. *Symbiosis* **45**: 15–24.

Ramírez, B.W. (1970) Host specificity of fig wasps (Agaonidae). *Evolution* **24**: 680–691.

Ramírez, B.W. (1974) Coevolution of *Ficus* and Agaonidae. *Ann Mo Bot Gard* **61**: 770–780.

Tribolium Genome Sequencing, Richards, C., Gibbs, S., Weinstock, R.A., Brown, G.M., Denell, S.J. *et al.* (2008) The genome of the model beetle and pest *Tribolium castaneum*. *Nature* **452**: 949–955.

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

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. and Kumar, S. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* **28**: 2731–2739.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F. and Higgins, D.G. (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* **25**: 4876–4882.

Vieira, F.G. and 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 Biol Evol* **3**: 476–490.

Vieira, F.G., Forêt, S., He, X., Rozas, J., Field, L.M. and Zhou, J.J. (2012) Unique features of odorant-binding proteins of the parasitoid wasp *Nasonia vitripennis* revealed by genome annotation and comparative analyses. *PLoS ONE* **7**: e43034.

Vogt, R.G. (2005) Molecular basis of pheromone detection in insects. In *Comprehensive Molecular Insect Science* (Gilbert, L.I., ed.), pp. 753–803. Elsevier, Amsterdam.

Wang, B., Xiao, J.H., Bian, S.N., Niu, L.M., Murphy, R.W. and Huang, D.W. (2013) Evolution and expression plasticity of opsin genes in a fig pollinator, *Ceratosolen solmsi*. *PLoS ONE* **8**: e53907.

Ware, A.B. and Compton, S.G. (1994) Responses of fig wasps to host plant volatile cues. *J Chem Ecol* **20**: 785–802.

Werren, J.H., Richards, S., Desjardins, C.A., Niehuis, O., Gadau, J., Colbourne, J.K. *et al.* (2010) Functional and evolutionary insights from the genomes of three parasitoid *Nasonia* species. *Science* **327**: 343–348.

Xiao, J.H., Yue, Z., Jia, L.Y., Yang, X.H., Niu, L.H., Wang, Z. *et al.* (2013) Obligate mutualism within a host drives the extreme specialization of a fig wasp genome. *Genome Biol* **14**: R141.

Xu, Y.L., He, P., Zhang, L., Fang, S.Q., Dong, S.L., Zhang, Y.J. *et al.* (2009) Large-scale identification of odorant-binding proteins and chemosensory proteins from expressed sequence tags in insects. *BMC Genomics* **10**: 632.

You, M., Yue, Z., He, W., Yang, X., Yang, G., Xie, M. *et al.* (2013) A heterozygous moth genome provides insights into herbivory and detoxification. *Nat Genet* **45**: 220–225.

Zhang, S., Zhang, Y.J., Su, H.H., Gao, X.W. and Guo, Y.Y. (2009) Identification and expression pattern of putative odorant-binding proteins and chemosensory proteins in antennae of the *Microplitis mediator* (Hymenoptera: Braconidae). *Chem Senses* **34**: 503–512.

Zhang, S., Chen, L.Z., Gu, S.H., Cui, J.J., Gao, X.W., Zhang, Y.J. *et al.* (2011) Binding characterization of recombinant odorant-binding proteins from the parasitic wasp, *Microplitis mediator* (Hymenoptera: Braconidae). *J Chem Ecol* **37**: 189–194.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Table S1. Identity matrix of *Ceratosolen solmsi* OBPs.

Table S2. Primers and standard curve parameters used for expression analysis of the *OBPs* in *Ceratosolen solmsi*.

Table S3. MIQE checklist.