Insect Molecular Biology

Insect Molecular Biology (2016) **25**(5), 580–594 doi: 10.1111/imb.12243

Identification and expression profile of odorant-binding proteins in *Halyomorpha halys* (Hemiptera: Pentatomidae)

D. P. Paula*, R. C. Togawa*, M. M. C. Costa*, P. Grynberg*, N. F. Martins* and D. A. Andow†

**Parque Estação Biológica, Embrapa Genetic Resources and Biotechnology, Brasília, Brazil; and †Department of Entomology, University of Minnesota, St. Paul, MN, USA*

Abstract

The brown marmorated stink bug, *Halyomorpha halys*, is a devastating invasive species in the USA. Similar to other insects, olfaction plays an important role in its survival and reproduction. As odorant-binding proteins (OBPs) are involved in the initial semiochemical recognition steps, we used RNA-Sequencing (RNA-Seq) to identify OBPs in its antennae, and studied their expression pattern in different body parts under semiochemical stimulation by either aggregation or alarm pheromone or food odorants. Thirty full-length putative HhalOBPs were identified, corresponding to 22 'classic' OBPs and eight 'Plus-C' OBPs. The similarity amongst them ranged from 4.95–70.92%, and with another 325 hemipteran OBPs similarity ranged from 1.94–91.51%, the highest levels being with other stink bug OBPs. Phylogenetic analysis confirmed the monophyly of seven groups of stink bug and other hemipteran OBPs. All 30 HhalOBPs were expressed and about 2/3 were expressed primarily in antennae. The expression of 21 HhalOBPs was higher in the antennae under alarm pheromone stimulus, indicating that multiple OBPs may be responding to this pheromone. Two were highest in antennae under aggregation pheromone stimulus. These findings should provide a basis for understanding the physiological functions of HhalOBPs and the chemosensory perception of this pest, which may help to uncover new control targets for behavioural interference.

Keywords: brown marmorated stink bug, chemoreception, RNA-Seq, semiochemicals, transcriptome.

First published online 12 May 2016.

Correspondence: Débora P. Paula, Parque Estação Biológica, W5 Norte, P.O. Box 02372, Embrapa Genetic Resources and Biotechnology, Brasília, Brazil. Tel.: + 55 (61) 34484929; fax: + 55 (61) 34484672; e-mail: debora.pires@embrapa.br

Introduction

Halyomorpha halys (Stål) (Hemiptera: Pentatomidae), also known as the brown marmorated stink bug (BMSB), is a polyphagous stink bug native to China, Japan, Korea and Taiwan (Hoebeke & Carter, 2003; Lee *et al*., 2013). It is an invasive species that has ravaged farms and distressed homeowners in the mid-Atlantic region of the USA and has spread to 41 different states and the District of Columbia (DC) (http://www.stopbmsb.org/where-is-bmsb/state-by-state/). In North America, *H. halys* has become a major agricultural pest across a wide range of commodities because it is a generalist herbivore, capable of consuming more than 100 different species of host plants (Bergmann *et al*., 2013), often resulting in substantial economic damage. It is a pest of tree fruits, grapes, other small fruits, row crops including vegetables, legumes and cotton, shade trees, ornamentals and nursery crops (Panizzi *et al*., 2000; Nielson *et al*., 2008; Zhu *et al*., 2012).

Olfaction is the primary sensory modality in insects, guiding them to locate food, oviposition sites (host cues) and conspecifics (mating), and to avoid natural enemies (Field *et al*., 2000; Asahina *et al*., 2008). Antennae are the principal biosensors for insect olfaction, where small and amphipathic proteins, called odorant-binding proteins (OBPs), mediate the first steps in odorant perception (eg Vogt & Riddiford, 1981; Hekmat-Scafe *et al*., 2002; Pelosi *et al*., 2006). OBPs comprise a multigene family with well-known physicochemical characteristics (eg Zhou *et al*., 2004, 2010; Xu *et al*., 2009).

The function of OBPs in hemipteran communication first began to be understood with the identification of the *Lygus lineolaris* antennal-specific protein (Dickens *et al*.,

1995). Now several OBPs have been isolated from 32 hemipteran species (eg Ji *et al*., 2013; He & He, 2014; Hull *et al*., 2014; Farias *et al*., 2015; Yuan *et al*., 2015). Nonetheless, although the molecular basis of insect olfaction has been extensively examined in some holometabolous insects (eg lepidopterans and dipterans), the molecular components and mechanisms comprising the heteropteran olfactory system have not been well studied. Identification and expression profiling of chemosensory genes are of primary importance for exploring their functions and the mechanisms of insect olfaction. Transcriptomes based on next-generation sequencing data have now been commonly used to identify chemosensory genes in species for which a complete genomic sequence has been unavailable. However, little is known about OBP expression profiles in response to environmental semiochemical stimuli, such as if they are constitutively expressed or quickly triggered by an environmental elicitor. It is also not clear if a semiochemical cue/signal triggers the expression of a single specific OBP or triggers the expression of multiple OBPs.

In the field, an aggregation pheromone has been used to attract the BMSB (Weber *et al*., 2014). The BMSB is also attracted by the volatiles of several host plants, especially fruits (Rice *et al*., 2014). In addition, it is widely known that stink bugs and other heteropterans repel conspecifics by releasing alarm pheromone (Ishiwatari, 1974). Therefore, with this diversity of environmental semiochemicals, it is expected that stink bugs have a diversity of OBPs to perceive these environmental cues or signals. So far, the maximum number of OBPs identified in a heteropteran has been 38 for *Apolygus lucorum* (Yuan *et al*., 2015).

To elucidate the molecular basis for olfactory reception of the BMSB and to facilitate the design and implementation of novel intervention strategies against this invasive pest, we studied the antennal transcriptome using RNA-Sequencing (RNA-Seq) to identify putative OBP transcripts, and subsequently examined OBP expression in different body tissues under different semiochemical stimulations by using quantitative real-time PCR (qPCR). The theoretical physicochemical parameters of the putative OBPs were characterized, and the similarity and phylogeny amongst these and other available hemipteran OBPs were analysed. This effort will increase understanding of heteropteran olfaction and specifically that of the BMSB.

Results and discussion

BMSB antennal transcriptome and HhalOBP identifications

Antennal transcriptome data of the BMSB had good coverage, with 247 865 908 raw reads obtained (after quality control 241 182 920 reads remained). A total of 37 229 contigs was assembled with a maximum length of 12 892 bp (average length 741.3 ± 823.8, N50 = 1254 bp). Regarding transcriptome completeness, we identified 227 constitutive genes out of the 248 indicated by CEGMA software (Core Eukaryotic Genes Mapping Approach), ie 91.53% completeness (Table S1). A total of 5802 gene ontology (GO) annotations was recovered, and categorized by biological process, cellular component and molecular function (Fig. 1). In total, 27 biological processes, 11 cellular components and 20 molecular functions were identified. The categories more directly related to insect olfaction in the list of biological processes were localization, response to stimulus, cell communication and sensory organ morphogenesis. In the list of cellular components, these were membrane, macromolecular complex and extracellular region. Finally, in the list of molecular functions, these were transporter activity, signal transducer activity and odorant binding. These most common GO annotations in biological processes, cellular component and molecular function were similar to those found for the antennal transcriptome of the mirid *Apo. lucorum* (Ji *et al*., 2013), although the antenna of the BMSB had higher richness of biological processes and molecular function. However, the annotations were not similar to the most common biological processes and molecular functions found in the antennal transcriptomes of the pentatomids *Chinavia ubica*, *Dichelops melacanthus* or *Euschistus heros* (Farias *et al*., 2015). The BMSB transcriptome also revealed the presence of enzymes belonging to energy, amino acid, nucleotide, cofactor, vitamin, carbohydrate, glycan and lipid metabolic pathways (Fig. S1). The BLASTx search matched homologous sequences in 29 other species (Fig. 2).

The high transcriptome coverage and number of contigs enabled the identification of 30 full-length putative OBPs, more than identified for most other Heteroptera (including the Pentatomidae; Zhou *et al*., 2010; Gu *et al*., 2011a, 2013; Ji *et al*., 2013; Farias *et al*., 2015). The only exceptions were the tarnished plant bug, from which 33 putative OBPs were found using a high-throughput sequencing platform screening of all developmental stages (Hull *et al*., 2014), and *Apo. lucorum*, from which 38 putative OBPs were found (Yuan *et al*., 2015). The BMSB had a similar number of putative OBPs as other stink bugs: 25 for *E. heros* and *C. ubica* (Farias *et al*., 2015). These polyphagous stink bugs have a broad range of host plants, with some overlap amongst them.

HhalOBPs in silico *characterization*

OBPs are small, globular, water-soluble, acidic proteins (about 120–150 amino acids) with a signal peptide in

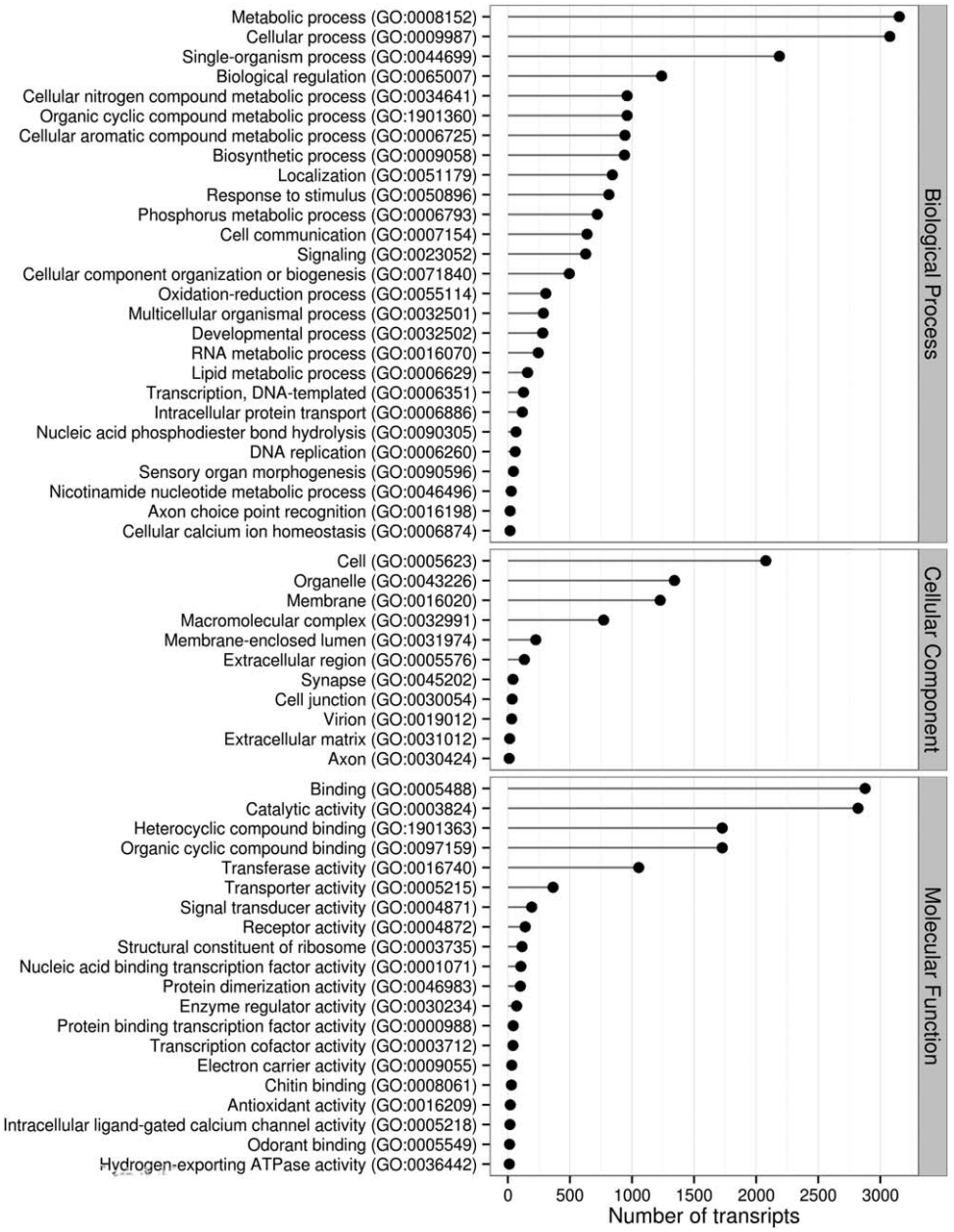

Figure 1. Summary of gene ontology (GO) annotation terms of the antennal brown marmorated stink bug transcripts. Transcript sequences generated by RNA-Sequencing (RNA-Seq) were annotated for GO terms using Blast2GO v. 2.8.0. The absolute number of annotated transcripts per GO term was plotted for each GO category (biological process, cellular component and molecular function). As expected, terms belonging to the first hierarchical levels had more transcripts.

the N-terminal region, and six cysteine residues (Cys) in conserved positions (Hekmat-Scafe *et al*., 2002; Zhou *et al*., 2004; Li *et al*., 2005; Forêt & Maleszka, 2006; Xu *et al*., 2009). This Cys motif forms a conserved tertiary protein structure of six alpha-helices coordinated by three disulphide bridges (Vogt & Riddiford, 1981; Pelosi *et al*., 2006; Zhou *et al*., 2010), and has been used as a signature for OBP identification. The BMSB putative OBPs (Table 1) had similar physicochemical properties as other OBPs, with some exceptions. Most of them ranged from 136 to 241 amino acid residues (14 582.91–26 738.95 Da), except HhalOBP5 with 333 residues (39 043.04 Da) (Fig. 3). Seven putative OBPs had the less common basic isoelectric point (pI), which was also observed in other species by Zhu *et al*. (2012), Farias *et al*. (2015) and Gong et al. (2015). As expected, all of the HhalOBPs had an amino terminal signal peptide, comprising the first 16 to 26 residues. Nearly all the putative HhalOBPs identified in the antenna had the predicted domain of pheromone/general odorant-binding protein (IPR006170), except in three cases, which were all Plus-C. The predicted secondary and tertiary structures had five to seven alpha-helixes (Fig. 4, Table S2), held together by three disulphide bridges between Cys1–Cys3, Cys2–Cys5 and Cys4–Cys6, as expected for an OBP structure (Fan *et al*., 2011).

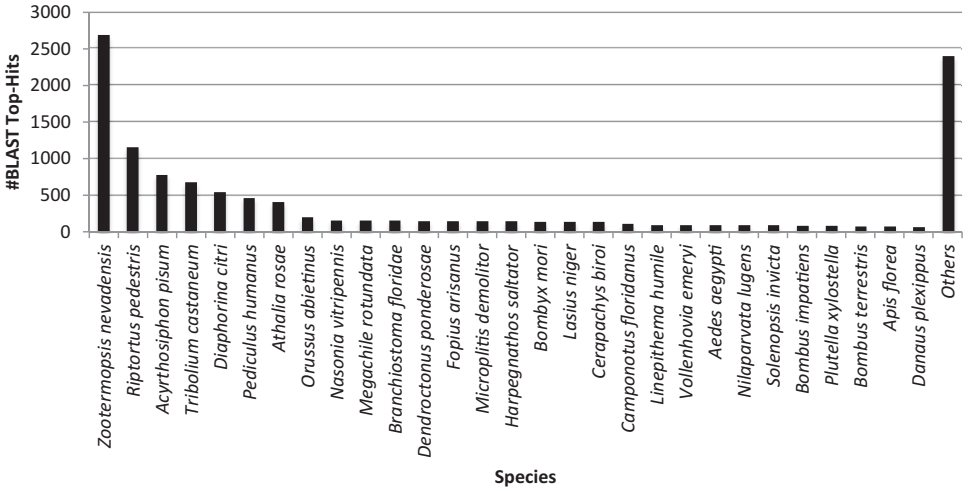

Figure 2. Top-hit species distribution with similarity to contigs from antennae of the brown marmorated stink bug (n = 11 894 top BLASTx hits, 10.4% total BLASTx hits).

Regarding the C-pattern, OBPs have been classified into subfamilies based on the number of conserved Cys residues: classical OBPs (six), Plus-C [more than six and a conserved proline, and according to Ji *et al.* (2013) with a C-terminal extension], minus-C (fewer than six) and OBP dimers (two complete OBP domains each with six conserved Cys residues) (eg Hekmat-Scafe *et al.*, 2002; Xu *et al.*, 2003; Zhou *et al.*, 2004, 2010). Most of the putative HhalOBPs had six Cys in the conserved positions, as in classical OBPs (Fig. 3, Table 1). Eight putative HhalOBPs had additional conserved Cys (C1a, C1b, C1c, C6a, C6b and C6c), a proline (Pro or P) after the C6a and an extended carboxy-terminal end, and were classified as Plus-C OBPs. Plus-C OBPs have also been found in other hemipterans (Zhou *et al.*, 2010; Gu *et al.*, 2011a; Ji *et al.*, 2013; Hull *et al.*, 2014;

Table 1. Physicochemical predictions for the *Halyomorpha halys* odorant-binding proteins (HhalOBPs)

z	Accession code	AA	pI	MW	Conserved Cys	SignalP	InterPro domain IPR006170 PBP/GOBP
HhalOBP1	KT875737	170	5.90	18899.59	6	1-21	73-166, 32-161
HhalOBP2	KT875738	175	8.81	19163.82	6	1-20	56-173, 51-164
HhalOBP3	KT875739	197	4.57	22546.47	12 (Plus-C)	1-24	57-162
HhalOBP4	KT875740	212	5.43	23699.71	12 (Plus-C)	1-20	48-135, 47-137
HhalOBP5	KT875741	333	5.77	39043.04	6	1-19	150-233, 143-223
HhalOBP6	KT875742	140	5.01	15653.63	6	1-20	21-128, 23-129
HhalOBP7	KT875743	141	8.66	14883.62	6	1-20	10-138, 21-130
HhalOBP8	KT875744	143	5.55	15872.7	6	1-23	26-132, 23-132
HhalOBP9	KT875745	204	5.35	23184.44	12 (Plus-C)	1-26	78-155, 61-155
HhalOBP10	KT875746	198	8.47	21994.06	12 (Plus-C)	1-25	Not predicted
HhalOBP11	KT875747	139	5.06	15507.58	6	1-20	23-136, 20-129
HhalOBP12	KT875748	149	9.01	16351.63	6	1-20	22-133, 22-133
HhalOBP13	KT875749	190	4.77	21633.36	12 (Plus-C)	1-24	Not predicted
HhalOBP14	KT875750	148	6.00	16671.42	6	1-20	6-139, 20-133
HhalOBP15	KT875751	149	5.39	16752.13	6	1-20	26-136, 27-135
HhalOBP16	KT875752	177	5.56	19863.64	6	1-23	63-174, 62-172
HhalOBP17	KT875753	150	5.34	16667.51	6	1-20	10-139, 21-133
HhalOBP18	KT875754	241	6.12	26738.95	12 (Plus-C)	1-21	121-224
HhalOBP19	KT875755	140	8.44	15017.61	6	1-20	22-137, 21-129
HhalOBP20	KT875756	138	4.84	15322.74	6	1-21	16-129, 18-130
HhalOBP21	KT875757	150	7.50	16687.47	6	1-16	6-139, 20-134
HhalOBP22	KT875758	146	8.84	16854.22	6	1-21	34-139, 32-137
HhalOBP23	KT875759	144	8.77	16012.09	6	1-21	25-137, 24-131
HhalOBP24	KT875760	145	5.50	16563.25	6	1-24	30-143, 25-138
HhalOBP25	KT875761	148	5.04	16394.06	6	1-20	26-141, 20-133
HhalOBP26	KT875762	149	5.52	16867.48	6	1-24	27-140, 32-136
HhalOBP27	KT875763	148	5.28	16611.42	6	1-20	7-140, 22-133
HhalOBP28	KT875764	196	4.86	21687.34	12 (Plus-C)	1-20	Not predicted
HhalOBP29	KT875765	237	6.33	27286.40	12 (Plus-C)	1-16	91-216
HhalOBP30	KU315186	136	4.47	14727.9	6	1-19	20-134, 22-127

AA, amino acid residue number; pI, isoelectric point; MW, molecular weight (Da); Cys, cysteine; SignalP, signal peptide amino acid location.
InterPro domain IPR006170 (pheromone/general odorant-binding protein, PBP/GOBP) classified in the SSF47565 and PF01395 families, respectively.

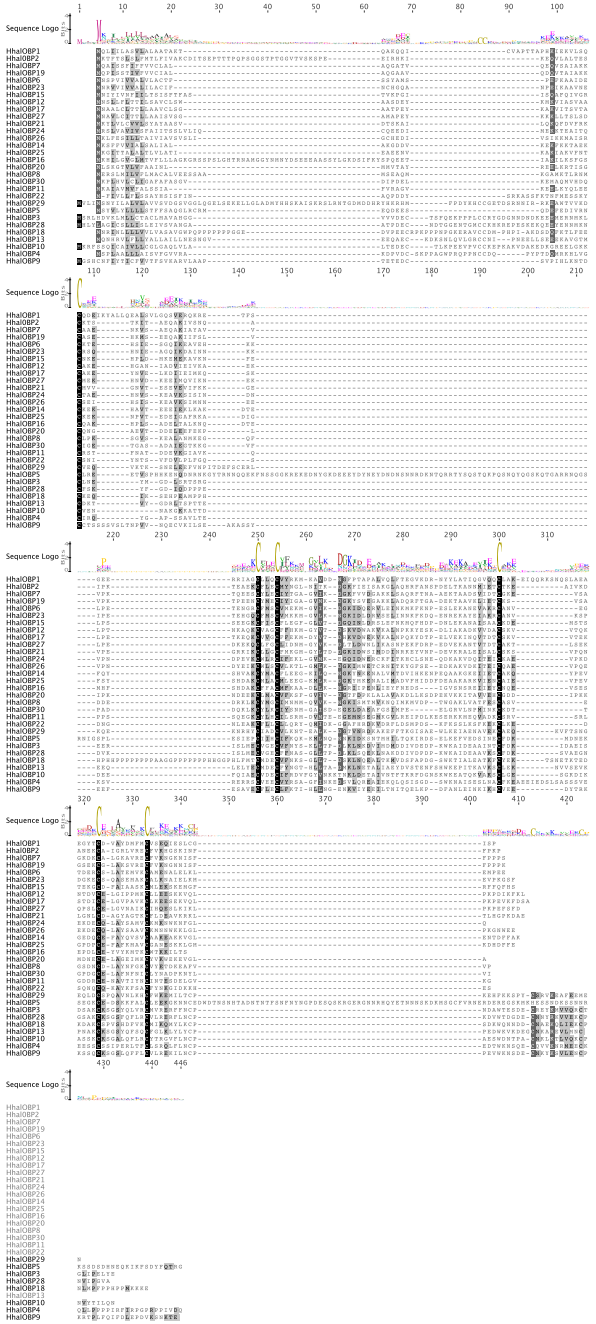

Figure 3. Alignment of the deduced amino acid sequences of the 30 *Halyomorpha halys* odorant-binding proteins (HhalOBPs). Similarity is indicated by the matrix Blosum62 where the black colour indicates 100% identity, dark grey 99% > identity ≥ 80%, light grey 80% > identity ≥ 60% and white colour identity < 60%. The sequence logo is at the top of the alignment. The amino acid percentage identity matrix is in Table S3. The conserved Cys are indicated in the sequence logo.

Farias *et al*., 2015). In this work, all putative OBPs had the strictly conserved three residues between C2–C3, which is the expected OBP motif C-pattern in insects (Xu *et al*., 2009). Only 10 out of 30 did not have the expected Cys motif-pattern for Hemiptera (C1-$X_{22\text{-}32}$-C2-X_3-C3-$X_{36\text{-}46}$-C4-$X_{8\text{-}14}$-C5-X_8-C6; Xu *et al*., 2009). These were: HhalOBP1 with C1-X_{37}-C2 and C4-X_{21}-C5; HhalOBP3, 4, 10, 13 and 28 with C5-X_9-C6; HhalOBP4 with C4-X_{22}-C5; HhalOBP5 with C1-X_{108}-C2; HhalOBP9 with C1-X_{38}-C2 and C5-X_9-C6; HhalOBP18 with C1-X_{48}-C2 and C4-X_{16}-C5-X_9-C6; and HhalOBP29 with C1-X_{35}-C2 and C5-X_9-C6; which included all eight Plus-C OBPs. Exceptions to the expected Cys motif-pattern have been identified in other stink bugs and Heteroptera, including *E. heros* and *C. ubica*. These were two EherOBPs with C1-X_{41}-C2, two EherOBPs and one CubiOBP with C4-X_{22}-C5, and one EherOBP and one CubiOBP with C5-X_9-C6 (Farias *et al*., 2015), *Adelphocoris lineolatus* with 14 AlinOBPs with C5-X_9-C6 and *Apo. lucorum* with five AlucOBPs with C5-X_9-C6 (Gu *et al*., 2011a; Ji *et al*., 2013).

HhalOBP similarity and phylogeny

OBPs generally have low similarity within and amongst species (Zhou *et al*., 2010). Exceptions have been observed in some intraspecific cases, such as with splice variants (Hull *et al*., 2014), and in some interspecific cases, such as when species share similar ligands (Vogt *et al*., 1991; Pelosi & Maida, 1995; Farias *et al*., 2015). The similarity of the mined putative OBPs from the BMSB amongst themselves and with other orthologous OBPs from other hemipteran species is represented respectively in Fig. 3 (Table S3) and Fig. 5. The similarity amongst HhalOBPs ranged from 4.95% (between HhalOBP5 and HhalOBP11) to 70.92% (between HhalOBP7 and HhalOBP19), but usually amongst HhalOBPs, it was less than 20.0%. The similarity amongst HhalOBPs and other hemipteran OBPs deposited in GenBank (325 sequences from 33 different species) ranged from 1.94% (between HhalOBP2 and RproOBP, GenBank JAA75164.1) to 91.51% (between HhalOBP4 and CubiOBP1, GenBank AIU64824.1).

The phylogenetic analysis for the HhalOBPs with the other 325 hemipteran sequences (Fig. 5A) confirmed again the intraspecific divergence of the putative OBPs (Farias *et al*., 2015). The HhalOBPs did not form a single clade, although three pairs formed sister pairs (HhalOBP3/HhalOBP28, HhalOBP7/HhalOBP19 and HhalOBP24/HhalOBP26) with bootstrap support ranging from 84% to 95%. At the same time, the phylogeny showed the convergence or shared derived character of several groups of stink bug OBPs: HhalOBP4 was related to the previously identified pair of EherOBP6/CubiOBP1, and HhalOBP11 was related to the EherOBP3/CubiOBP3 pair (Farias *et al*., 2015). Several new monophyletic groups involving only stink bug OBPs were identified (bootstrap support in parentheses), HhalOBP12/HhalOBP17/HhalOBP27/EherOBP4 (94%), HhalOBP6/

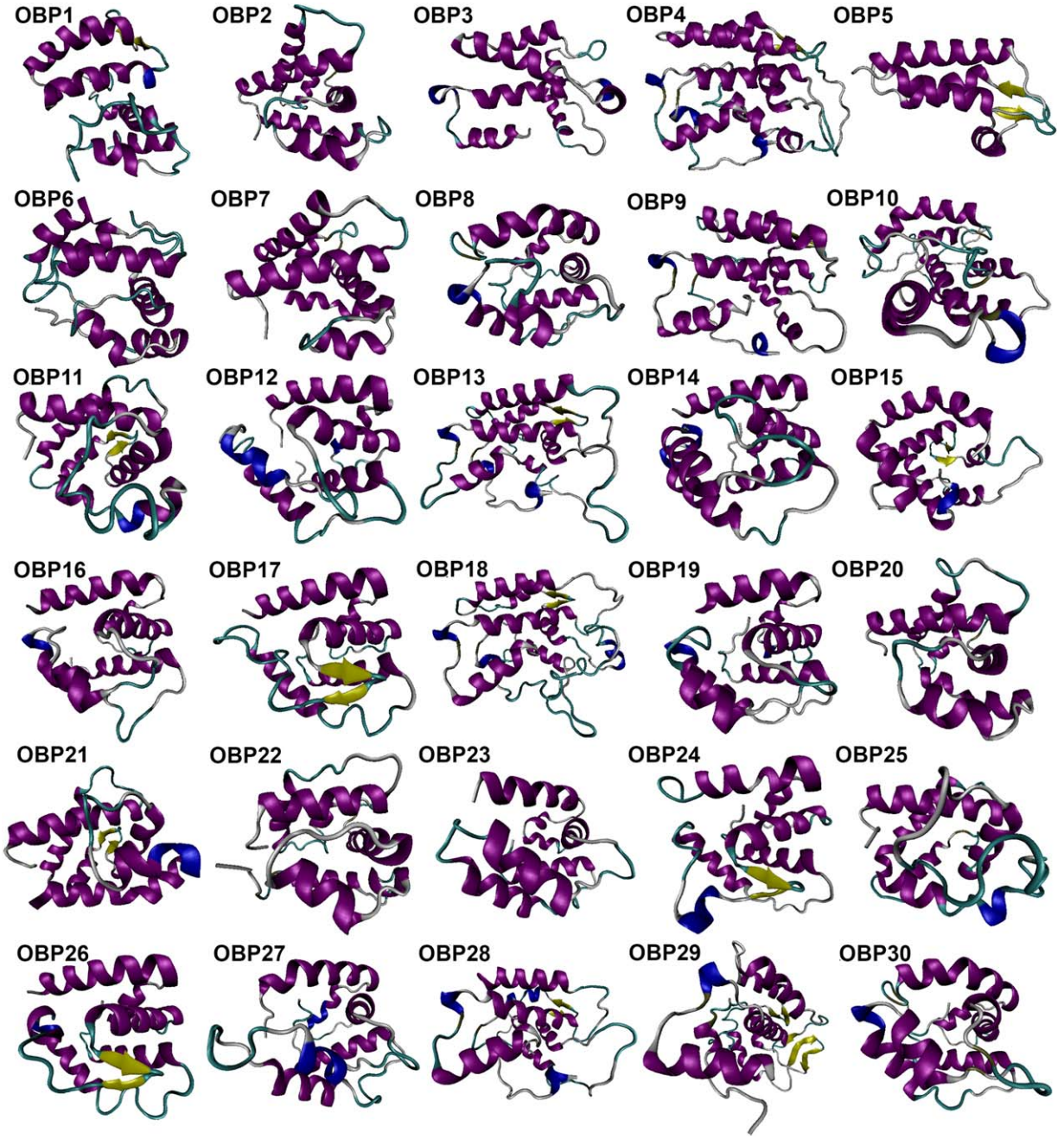

Figure 4. Three-dimensional models of the *Halyomorpha halys* odorant-binding proteins (HhalOBPs). The structures were orientated with the N-terminus to the right side. The structures in purple are α-helixes, in yellow are β-strands, in dark blue are turns, and in light blue and white are coils.

HhalOBP23/CubiOBP2 (87%), HhalOBP25/EherOBP1 (98%), HhalOBP11/EherOBP3/CubiOBP3 (97%), HhalOBP21/CubiOBP4 (97%) and possibly HhalOBP30/EherOBP2 (75%). Several of these groups had mirid OBPs as the closest outgroups (eg Alin, Aluc, Asut, Llin), but the bootstrap support between the stink bug and mirid OBPs was low ($<$ 50%).

The phylogenetic relations of the stink bug OBPs and other nonpentatomid hemipteran species supported two previously identified patterns. The first was monophyly with early divergence from other OBPs. The four monophyletic groups that we had identified previously (Farias *et al.*, 2015) continued to be monophyletic and three more could be identified. The EherOBP1 group expanded (bootstrap support 83%) and included HhalOBP14 and 25, two reduviid OBPs and eight mirid OBPs. These additions now suggest that the reduviid OBPs would be a separate subgroup (bootstrap support 45%; Fig. 5B). The EherOBP5

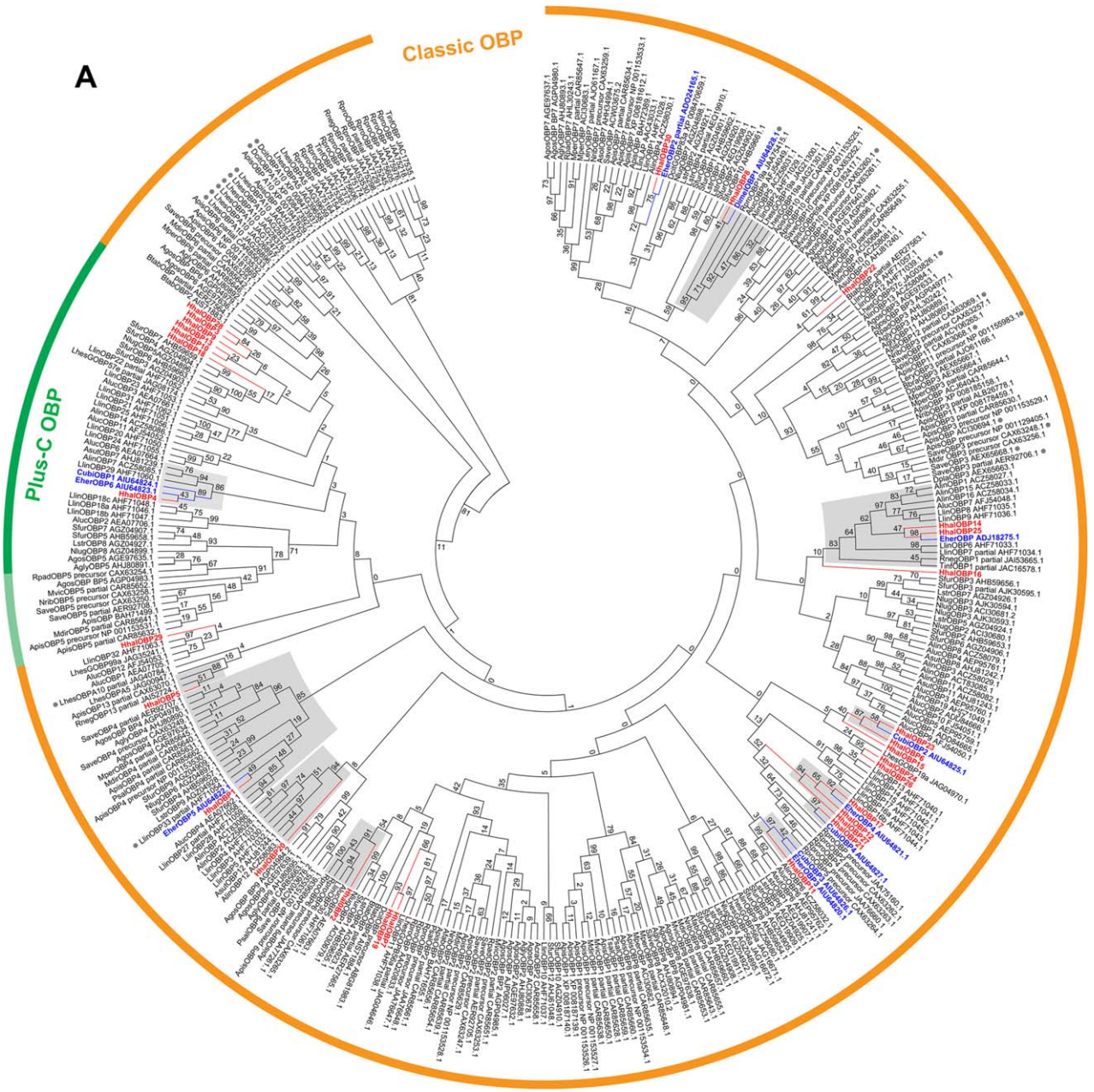

Figure 5. Phylogenetic relationships of: (A) *Halyomorpha halys* odorant-binding proteins (HhalOBPs; in red), other pentatomid (in blue) and other putative hemipteran OBPs; (B) Monophyletic relationships involving the HhalOBPs and other nonpentatomid hemipterans. The trees were constructed with MEGA 6.06 using a WAG + G + I model. Values indicated at the nodes are bootstrap values based on 1000 replicates. The species names are abbreviated with four letters together with the accession numbers of the deduced OBP amino acid sequences. Light green arc indicates possible Plus-C, and grey spots indicate known and possible Minus-C. Grey-shaded parts of the phylogeny in (A) are all monophyletic lineages involving more than one pentatomid species. Acra, *Aphis craccivora*; Afab, *Aphis fabae*; Agly, *Aphis glycines*; Agos, *Aphis gossypii*; Alin, *Adelphocoris lineolatus*; Aluc, *Apolygus lucorum*; Apis, *Acyrthosiphon pisum*; Asol, *Aulacorthum solani*; Asut, *Adelphocoris suturalis*; Bbra, *Brevicoryne brassicae*; Btab, *Bemisia tabaci*; Cubi, *Chinavia ubica*; Dcit, *Diaphorina citri*; Dimel, *Dichelops melacanthus*; Dpla, *Drepanosiphum platanoidis*; Eher, *Euschistus heros*; Lery, *Lipaphis erysimi*; Lhes, *Lygus hesperus*; Llin, *Lygus lineolaris*; Lstr, *Laodelphax striatella*; Mdir, *Metopolophium dirhodum*; Mper, *Myzus persicae*; Mvic, *Megoura viciae*; Nlug, *Nilaparvata lugens*; Nrib, *Nasonovia ribisnigri*; Psal, *Pterocomma salicis*; Rneg, *Rhodnius neglectus*; Rpad, *Rhopalosiphum padi*; Rpro, *Rhodnius prolixus*; Save, *Sitobion avenae*; Sfur, *Sogatella furcifera*; Tinf, *Triatoma infestans*; Tsal, *Tuberolachnus salignus*.

group also expanded (bootstrap support 85%) with HhalOBP1, four mirid OBPs, four delphacid OBPs and 10 aphid OBPs; the aphids were clearly a separate subgroup (bootstrap support 96%; Fig. 5B). The DimelOBP1 group expanded slightly with HhalOBP8, one reduviid OBP and five mirid OBPs (bootstrap support 95%), and the EherOBP6/CubiOBP1 group, previously with three mirid OBPs (bootstrap support 86%), now had the addition of

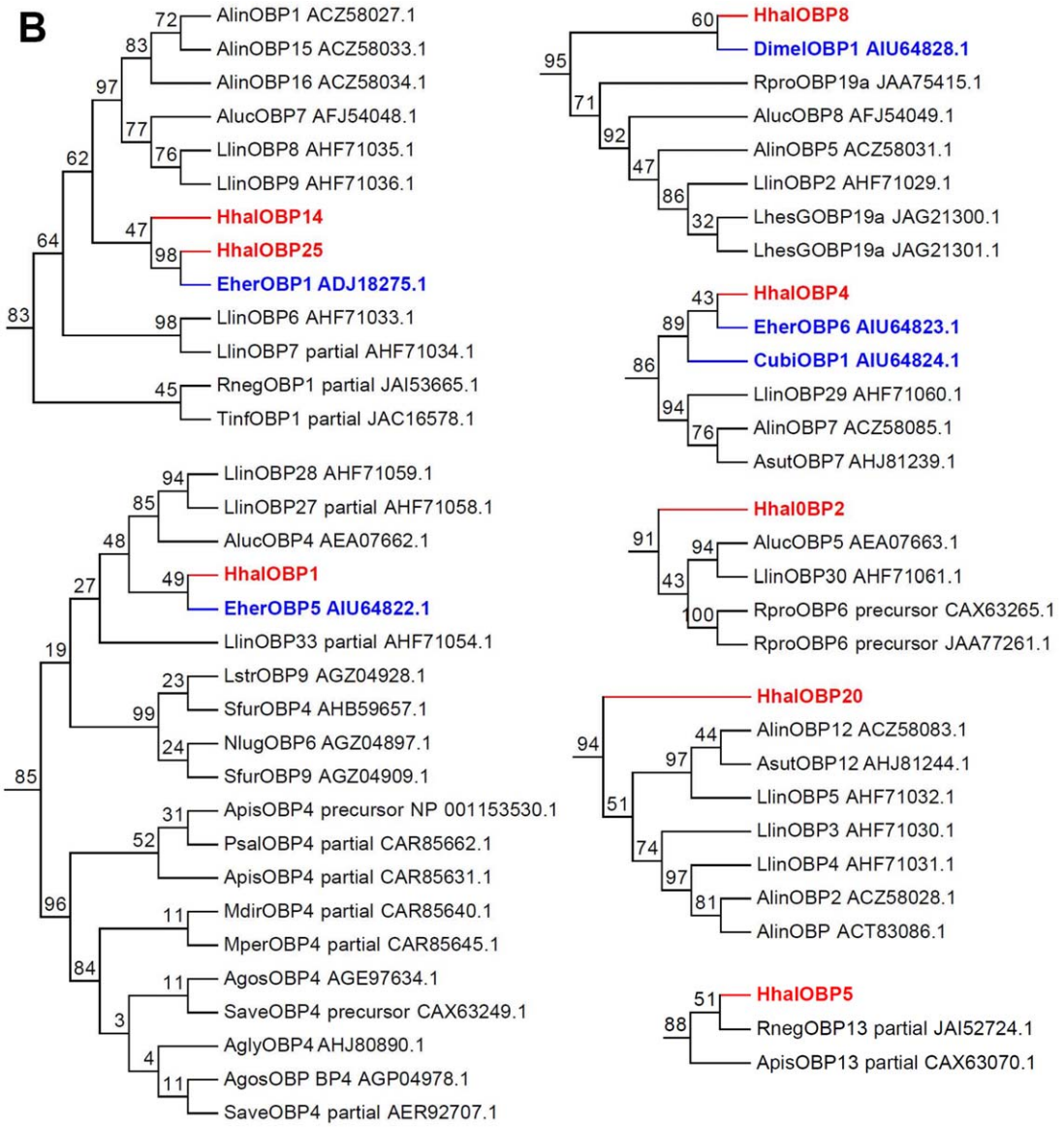

Figure 5. Continued.

HhalOBP4 and separated into stink bug (bootstrap support 89%) and mirid (bootstrap support 94%) subgroups. The three new monophyletic groups were the ones associated with HhalOBP5, 2 and 20. HhalOBP5 was monophyletic with an aphid (ApisOBP13) and reduviid (RnegOBP13) OBP with bootstrap support 88%; HhalOBP2 was monophyletic with AlucOBP5, LlinOBP30 and RproOBP6 with bootstrap support 91%; and HhalOBP20 was monophyletic with seven mirid OBPs (AlinOBP, AlinOBP2 and 12, LlinOBP3, 4 and 5, and AsutOBP12) with bootstrap support 94%. As we predicted previously (Farias *et al.*, 2015), the enrichment of the heteropteran database has begun to clarify the phylogenetic pattern within the monophyletic clades.

The second pattern was exhibited by HhalOBP9, 10, 13, 14, 15, 16, 18, 22 and 29, which had uncertain phylogenetic positions in the Hemiptera, because there were no similar OBPs in the hemipteran database. These had bootstrap supports between 4% and 61% with other species. It is likely that database enrichment will clarify the phylogenetic relationships of these OBPs. For example, EherOBP2, EherOBP4, CubiOBP2, CubiOBP4 and the EherOBP3/CubiOBP3 pair, which previously had uncertain phylogenetic positions (Farias *et al.*, 2015), are all now monophyletic with one to three HhalOBPs.

All of the hemipteran Plus-C OBPs grouped together in the phylogeny, albeit with weak bootstrap support (= 8%), and they were paraphyletic with two clades of

classic OBPs (Fig. 5A). Some of the OBPs included in the Plus-C clades were incomplete (partial or precursors), but as they were monophyletic with known Plus-C OBPs, we designated them as possible Plus-C OBPs (light green arc, Fig. 5A). Future investigations may confirm that these incomplete OBPs are indeed Plus-C OBPs.

These results support the hypothesis that within the Heteroptera there has been extensive duplication and differentiation of OBPs (Vieira *et al*., 2012), especially compared with the conservative evolution of these proteins in the aphids (Gu *et al*., 2013) and Lepidoptera (Vogt *et al*., 1991).

HhalOBP expression

All 30 HhalOBPs were expressed in adult BMSBs (Fig. 6). Overall, HhalOBPs had higher expression levels in antennae than in the other parts of the body evaluated (heads without antennae, legs, and wings). Twenty-one HhalOBPs were expressed more in the antennae for all bioassays (HhalOBP1 to 4, 7 to 9, 11 to 16, 19, 20, 22, 25, 27 to 30). The other nine OBPs had qualitative differences in expression amongst the bioassay treatments. HhalOBP6, 17, 23 and 24 had lower expression in antennae in the aggregation and food bioassay treatments, and higher expression in the alarm treatment. HhalOBP18 and 21 had higher expression in the antennae in the aggregation and alarm treatments and either lower or no difference in the food treatment. HhalOBP26 had higher expression in the antennae in the alarm treatment, lower expression in the food treatment and no difference in the aggregation treatment. HhalOBP10 was unusual because it was the only OBP that did not show higher expression in the antennae in any of the treatments. HhalOBP5 had missing values in the alarm treatment and so it could not be similarly evaluated.

The aggregation bioassay treatment had 24 OBPs with higher expression in the antennae, and these increases ranged from 1.61-fold for HhalOBP21 to 6580-fold for HhalOBP25. The alarm treatment had 28 OBPs with higher expression in the antennae, ranging from 3.91-fold for HhalOBP21 to 94 600-fold for HhalOBP25. The food treatment had 23 OBPs with higher expression in the antennae, ranging from 1.43-fold in HhalOBP2 to 18 200-fold in HhalOBP25.

The expression of the HhalOBPs was higher in the antennae in the alarm treatment than any of the other treatments for 21 of the OBPs. Of these 21, seven had higher expression in the aggregation treatment than the food treatment (HhalOBP2, 6, 7, 9, 18, 21 and 27). Two HhalOBPs had highest expression in the aggregation treatment, HhalOBP4 and HhalOBP8. These results indicate that the alarm and aggregation pheromones each triggered the expression of multiple OBPs, and these were not the same set of OBPs. HhalOBP3, 5 and 22 were expressed similarly in the antennae of the three bioassay treatments. HhalOBP1, 10, 16 and 29 had similar antennal expression in the aggregation and alarm treatments and lower expression in the food treatment.

These experiments showed that: (1) most of the OBPs are being expressed at least at a basal level, no matter the semiochemical stimulus; (2) the expression of 29 of the 30 OBPs were triggered or accentuated in the antennae by one or more of the semiochemical stimuli; (3) one semiochemical stimulus triggered or accentuated the antennal expression of several OBPs. Each of the semiochemicals stimulated the expression of 23–28 HhalOBPs in the antennae, possibly to enable the species to recognize better other environmental cues or signals related to the species' survival and reproduction.

In Hemiptera, such as *Apo. lucorum*, *E. heros*, *L. lineolaris*, *Ad. lineolatus*, *Aphis gossypii*, *Nilaparvata lugens* and *Sogatella furcifera*, a large proportion of identified OBP genes are highly expressed in the antennae (Gu *et al*., 2011a, 2013; Ji *et al*., 2013; He & He, 2014; Hull *et al*., 2014; Zhou *et al*., 2014; Farias *et al*., 2014). The high antennal expression of OBPs is clearly correlated with a role in olfaction. In Hemiptera, OBPs are expressed in different types of sensillae, with some OBPs located in the sensilla trichodea and sensilla basiconica (Gu *et al*., 2011b; Ji *et al*. 2013), and some only located in the short sensilla basiconica (Sun *et al*., 2014). Remarkably, subsequent studies have identified OBPs that are highly expressed not only in the antennae but also in other tissues, such as the legs and heads. Many scholars have suggested that the OBPs that are highly expressed in these tissues might be associated with taste perception and might participate in other physiological functions (Shanbhag *et al*., 2001a; Jeong *et al*., 2013). The diverse expression profiles of insect OBPs make it possible for these proteins to participate in different physiological functions.

In hemipteran OBPs, a high degree of expression was observed in the antennae. A total of 21 *L. lineolaris* OBPs, 16 *Apo. lucorum* OBPs, 12 *Ad. lineolatus* OBPs, five *Aph. gossypii* OBPs, more than six *N. lugens* OBPs and three *S. furcifera* OBPs were more highly expressed in the antennae than in other tissues (Gu *et al*., 2011a, 2013; Hua *et al*., 2012; Ji *et al*., 2013; He & He, 2014; Hull *et al*., 2014; Zhou *et al*., 2014). In *L. lineolaris*, OBPs not only were shown to have antennae-specific expression, but were also expressed in other tissues, such as heads, stylets, legs and wings (Hull *et al*., 2014). *Drosophila melanogaster* OBPs were also expressed in gustatory organs on the wings and legs (Galindo & Smith, 2001; Shanbhag *et al*., 2001b). In our research, all HhalOBPs were expressed in non-antennal body parts, suggesting that these genes might also participate in taste functions.

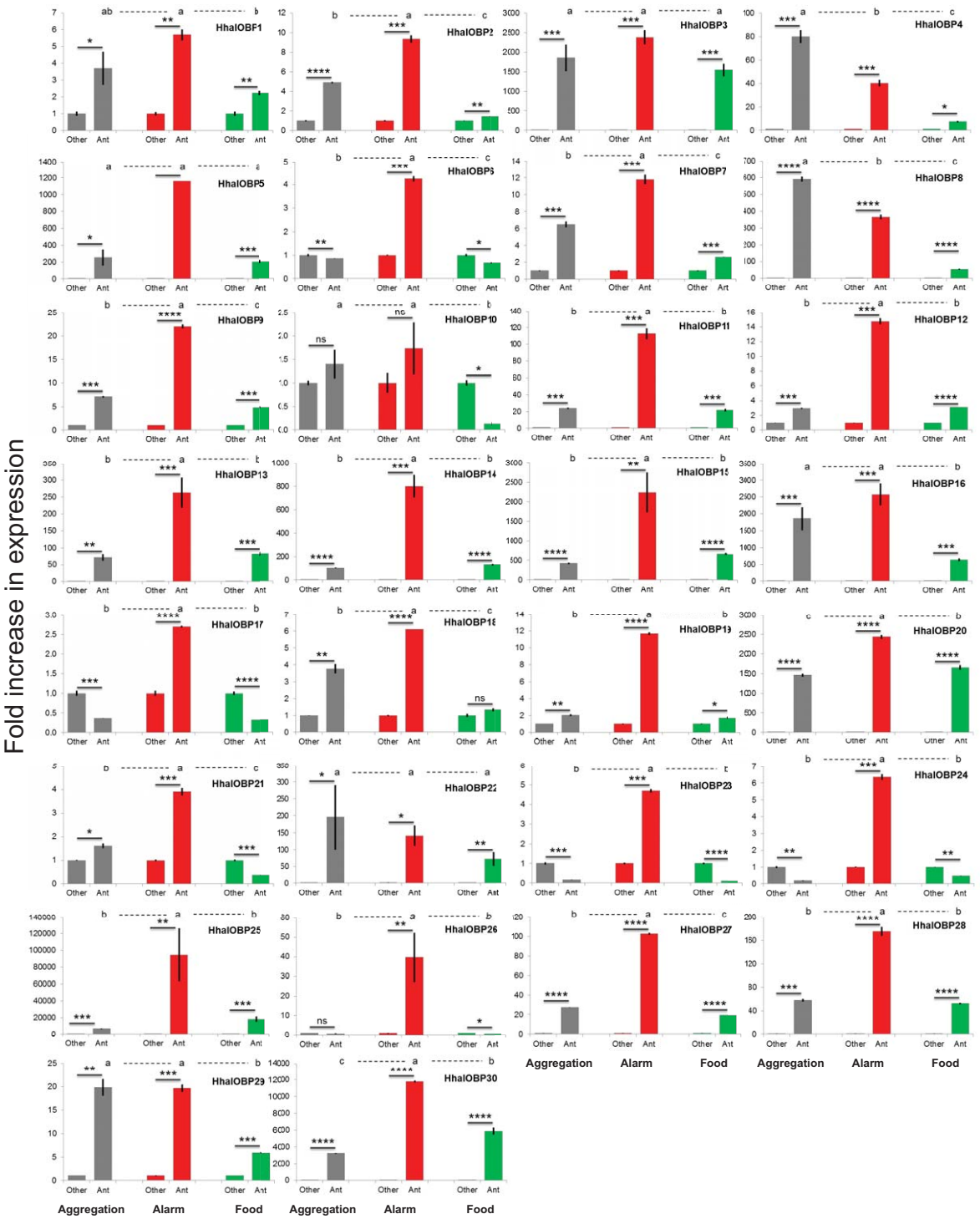

Figure 6. Relative expression levels of the *Halyomorpha halys* odorant-binding proteins (HhalOBPs) in antennae and other body parts (heads without antennae, legs, and wings) under different semiochemical stimuli (aggregation pheromone, alarm pheromone and food), using quantitative real-time PCR. The SD of the mean for three replicates is represented by the error bar, except for the HhalOBP5 alarm treatment, which is missing the SD. Difference in expression between antennae and other body parts is indicated above the bars for each semiochemical bioassay and each OBP separately using a solid line and stars (*$<$0.05; **$<$0.01; ***$<$0.001; ****$<$0.0001). Significant differences in expression in the antennae amongst the three bioassay treatments are indicated above the paired bars of the three bioassays for each OBP separately using a dashed line and letters; different letters within an OBP indicate a significant difference by Tukey's HSD (P $<$0.05). Ant, antennae; ns, not significant.

In conclusion, in this first report of OBPs in the BMSB, we identified 30 full-length HhalOBPs highly likely to be related to olfaction owing to their predominant expression in the antennas. More importantly, we showed that multiple OBPs were expressed from one semiochemical stimulus. Our findings are an initial contribution towards understanding OBP functions in the BMSB. Future studies of the binding characteristics of these OBPs are warranted to determine their roles in the odorant reception in this invasive species.

Experimental procedures

Insects

Adult BMSBs were obtained from laboratory colonies at Corvallis, OR, Lexington, KY, and St Paul, MN. They were held at 25 ± 2°C under a 16 h light : 8 h dark photoperiod for at least 72 h in environmental chambers in 10-cm Petri dishes containing fresh green beans and water to standardize pre-experimental conditions.

RNA-Seq

A total of 50 pairs of antennae from 3-day-old stink bug males and females (1:1) was flash frozen in liquid nitrogen prior to disruption using a mortar and pestle to grind them with TRIzol® reagent (Invitrogen, Life Technologies, Carlsbad, CA, USA). Total RNA was isolated using a Qiagen RNeasy Mini Kit, following the protocol for animal tissues (Qiagen, Germantown, MD, USA). RNA was eluted in RNase-free water, and the quality and quantity were assessed with an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). After extraction, 1 μg of the total RNA was placed in RNAstable tubes (Biomatrica, San Diego, CA, USA) and shipped to Beijing Genomics Institute (China) for sequencing. The RNA sample was used for TruSeq RNA library construction and sequencing using the Illumina HiSeq 2500 (Illumina, San Francisco, CA, USA) (one lane, 100 bp paired-end).

De novo *transcriptome assembly, annotation and OBP mining*

The transcriptome data set was received in Fastq format. Quality control was performed by Trimmomatic (http://etal.usadellab.org/cms/?page=trimmomatic) using leading 3, trailing 3, sliding window 4:15 and minimum length 36. The assembly was conducted with the Trinity package (Haas *et al.*, 2013) v. r2013-02-05 with the jellyfish kmer method (default parameters). cegma software was used (default parameters) to assess the completeness of the transcriptome assembly and identify the presence of a core consisting of 248 highly conserved proteins that are found in a wide range of eukaryotes, mostly housekeeping genes, that can be expected to be expressed (Parra *et al.*, 2009). The raw sequence transcriptome data were deposited in the Sequence Read Archive at the (DNA Data Bank of Japan DDBJ)/European Molecular Biology Laboratory (EMBL)/GenBank database and can be accessed under BioProject PRJNA263721. The assembly results were deposited in the Transcriptome Shotgun Assembly. The version described in this paper is the first version, GBTB01000000.

To search for OBP transcripts, the assembled transcripts (contigs) were annotated using Blast2GO v. 2.8.0 (Conesa *et al.*, 2005). The initial searches of the National Center for Biotechnology Information nonredundant (nr) protein database were conducted with the BLASTx algorithm with an E-value less than 1.0E-5, followed by collection of GO terms from the GO database and retrieval of Kyoto Encyclopedia of Genes and Genomes (KEGG) Pathway designations. The GO terms with absolute counts from all hierarchical levels were analysed with the Revigo suite to exclude redundant terms (Supek *et al.*, 2011) using the following parameters: allowed similarity of 0.05 and the numbers associated to GO categories are 'some other quantity where higher is better'. The advanced options were not changed. Graphics were produced using the R ggplot2 package (R Development Core Team 2008). The Enzyme Comission (EC) numbers were used to create customized metabolic and biosynthesis pathways of secondary metabolites with iPath2 suite (Yamada *et al.*, 2011).

The putative OBP contigs were screened for open reading frames (ORFs) using the software ORF Finder (http://www.ncbi.nlm.nih.gov/gorf/gorf.html) to identify the coding regions. In addition, the OBP domain IPR006170 (pheromone/general odorant-binding protein) was checked by InterProScan 5 (Jones *et al.*, 2014). We have named the mined putative full-length OBP ORFs by the first letter of the genus followed by the first three letters of the species name (ie HhalOBP) followed by a number in increasing scaffold order.

In silico *characterization of the HhalOBPs*

The full-length putative OBPs were submitted to a series of *in silico* analyses to predict: the deduced amino acid sequences (Expasy Translate Tool, http://web.expasy.org/translate/), the pI and monoisotopic molecular weight (MW) (Compute pI/Mw, http://web.expasy.org/compute_pi/) and the presence of the signal peptide (SignalP 4.1, http://etal.cbs.dtu.dk/services/SignalP/) (Petersen *et al.*, 2011). To predict the tertiary structure, a template search with BLAST and HHblits was performed against the SWISS-MODEL template library (SMTL, last update: 2015 - 10 - 28, last included Protein Data Bank release: 2015-10-23). The target sequence was searched with BLAST (Altschul *et al.*, 1997) against the primary amino acid sequence contained in the SMTL. A total of three templates was found. An initial HHblits profile was built using the procedure outlined in Remmert *et al.* (2012), followed by one iteration of HHblits against Non-redundant database 20. The obtained profile was searched against all profiles of the SMTL. For each identified template, the template quality was predicted from features of the target-template alignment, and the template with the highest quality was then selected for model building. Models were built based on the target-template alignment using Promod-II (Guex & Peitsch, 1997). Coordinates, which are conserved between the target and the template, were copied from the template to the model. Insertions and deletions were remodelled using a fragment library. Side chains were then rebuilt. Finally, the geometry of the resulting model was regularized by using a force field. In the event that loop modelling with ProMod-II did not produce satisfactory

results, an alternative model was built with MODELLER (Sali & Blundell, 1993). The global and per residue model quality was assessed using the QMEAN scoring function (Benkert *et al.*, 2011). For improved performance, weights of the individual QMEAN terms were trained specifically for SWISS - MODEL.

We also analysed the similarity amongst the deduced amino acid sequences of the putative 30 full-length identified in the antenna of the BMSB using MAFFT 7.017 (Katoh *et al.*, 2002) with the algorithm auto, scoring matrix BLOSUM62, gap opening penalty 1.53 and offset value of 0.123, in GENEIOUS 7.0.5 (Kearse *et al.*, 2012). This analysis enabled us to identify the conserved Cys motif-pattern. The similarity of the HhalOBPs was also compared with 325 hemipteran OBPs from 33 species available in DDBJ/EMBL/GenBank by constructing a phylogenetic tree using MEGA 6.06 (Tamura *et al.*, 2013) to find the best maximum likelihood (ML) model with the lowest small sample Akaike information criterion value. The unrooted phylogenetic tree was constructed using the statistical method of maximum likelihood, WAG + G + I model with 1000 bootstrap replications, partial deletion of gaps, coverage cut-off 95%, and nearest-neighbour interchange ML heuristic method, initial tree for ML default: neighbour-joining/BIONJ, with moderate branch swap filter. We identified monophyletic lineages using bootstrap values > 80 (Gascuel, 1997).

Semiochemical stimulation to study HhalOBP expression

Three bioassays were conducted to test the expression of the HhalOBPs in the presence of food, aggregation pheromone and alarm pheromone. The bioassays were run in the laboratory at 25 ± 2°C using an acrylic cage (80 × 55 × 60 cm). Adults obtained from the different sources were randomly assigned to a bioassay treatment to control for differences amongst the sources and differences in age. Females and males were stimulated separately in each bioassay. For the food scent attraction bioassay, three mature organic peaches were placed inside the cages, and starved (for 24 h) females ($n = 27$) or males ($n = 29$) were released in the corners of the cage so that there was a maximum of four individuals in the cage at any one time. For the aggregation pheromone attraction, one lure (AlphaScents (West Linn, OR, USA), BMSB lure, methyl E2,E4,Z6-decatrienoate) was placed in the centre of the cage, and either females ($n = 15$) or males ($n = 8$) were released as in the food scent bioassay. For the alarm pheromone stimulation, females ($n = 24$) or males ($n = 45$) were put in a small dissecting jar (25 cm diameter) to force them to disturb each other and release alarm pheromone. The bugs were considering stimulated when they presented the behaviour response (individuals took between 10 to 60 min) to walk onto the food or lure. Immediately after the desired response behaviour, the bugs were collected and flash-frozen in liquid nitrogen (5–10 s) and stored in RNAlater (Thermo Fisher Scientific, Waltham, MA, USA) at 4°C until dissection and molecular analysis.

HhalOBP expression analysis

For each semiochemical stimulation bioassay, antennae, heads, legs and wings of the males and females were dissected using RNase-free entomological scissors and tweezers (cleaned with RNaseZAP, Thermo Fisher Scientific). Each body part was immediately immersed in 1 ml TRIzol (Invitrogen), grouped according to the semiochemical stimulation and sex. They were macerated and homogenized using a RNase-free glass rod (different for each sample) until maceration of the tissues was complete. The lysates were incubated with TRIzol at room temperature for 5 min and the total RNA extraction was processed using a PureLink RNA Mini kit (Ambion by Thermo Fisher Scientific), according to the manufacturer's instructions. The DNase treatment and purification were processed on-column using a PureLink DNase provided by a complementary kit of the same brand. The RNA yield and quality were analysed by agarose gel electrophoresis and by fluorescence using a Qubit® RNA HS Assay Kit Thermo Fisher Scientific. For the first-strand cDNA synthesis, we pooled 1 μg total RNA of males and females in six sample groups: (1) antennae from the alarm pheromone bioassay, (2) antennae from the aggregation pheromone bioassay, (3) antennae from the food bioassay, (4) heads (without antennae), legs and wings from the aggregation pheromone bioassay, (5) heads (without antennae), legs and wings from the alarm pheromone bioassay, (6) heads (without antennae), legs and wings from the food bioassay. The cDNAs were generated using the SuperScript III First-Strand Synthesis System for RT-PCR (Invitrogen), according to the manufacturer's instructions. Primer pairs were designed (Table S4) based on the putative transcripts in order to determine transcript abundance of HhalOBPs in these six treatments using qPCR. The qPCR reactions (25 μl) were prepared using Thermo Scientific Maximo SYBR Green/ROX qPCR Master Mix (2x), with 500 ng cDNA and the HhalOBP primer pair at 0.3 μM. The amplifications were performed in a StepOnePlus Real-Time PCR System (Thermo Fisher Scientific) with two-step cycling parameters, as indicated by the manufacturer. Each qPCR reaction for each sample was performed in three technical replicates, designed so that one technical replicate of each of the six sample groups for a particular primer pair were together on a plate. This allowed us to conduct statistical analyses to estimate and eliminate any plate effects. No-template controls were included in each experiment. The relative fold change in different tissues was assessed by comparing the expression level of each OBP in antennal tissues to that in the other body parts (Gu *et al.*, 2011a). Melt curve analysis was performed to test locus-specificity of reaction products. The baseline fluorescence per reaction, the efficiency for each primer pair, the quantification cycle value and the starting concentration per sample (N_0) were estimated from the raw fluorescence data using the software LINREG v. 2015.3 (Ruijter *et al.*, 2009, 2014; Tuomi *et al.*, 2010). As the qPCR samples from the different bioassays had different numbers of individual adults in them, the N_0 values were normalized by the number of individuals to enable comparison across bioassays. Two statistical analyses were conducted using these normalized N_0 values. First, for each bioassay and OBP, the expression in the antennae was compared to that in the head/wings/legs by a paired *t*-test using the $\log_{10}$-transformed N_0 values paired by qPCR plate. This allowed determination of whether or not there was higher (or lower) expression in the antennae for each OBP and bioassay independently. Second, the relative increase (or decrease) of expression in the

antennae (compared to the other body parts) was compared for the three bioassays to determine if expression was higher in one or more of the bioassays. The difference in $\log_{10} N_0$ in the antennae and other body parts was calculated, which is the same as the $\log_{10}$ transformation of the fold-increase in expression in the antennae. These differences were analysed by two-way analysis of variance separately for each OBP, blocked for the qPCR plate, with bioassay as the treatment effect. Differences amongst the bioassay treatments were tested using Tukey's honest significant difference (HSD). The results were displayed as fold-increase, as typically done, rather than the $\log_{10}$-transformed scale. The significance of the differences amongst body parts was indicated with * above the two body parts for each bioassay and OBP. The significance amongst the bioassays was indicated with a dashed line and letters above the bars for each OBP. The standard errors in the figures were calculated using the formula for the variance of the product of correlated random variables.

Acknowledgements

We thank Jana Lee (United States Department of Agriculture-Agricultural Research Service), Ric Bessin (University of Kentucky) and Theresa Cira (University of Minnesota) for providing adults for this study. We also thank Elise Rosengren for assistance. This work was partially funded by USDA Regional Research Project NC205.

References

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W. *et al.* (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**: 3389–3402.

Asahina, K., Pavlenkovich, V. and Vosshall, L.B. (2008) The survival advantage of olfaction in a competitive environment. *Curr Biol* **18**: 1153–1155.

Benkert, P., Biasini, M. and Schwede, T. (2011) Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics* **27**: 343–350.

Bergmann, E., Bernhard, K.M., Bernon, G., Bickerton, M., Gill, S., Gonzales, C. *et al.* (2013) *Host plants of the brown marmorated stink bug in the U.S.* http://www.stopbmsb.org/where-is-bmsb/host-plants/#host_plants_table. Accessed on 15 September 2013.

Conesa, A., Götz, S., García-Gómez, J.M., Terol, J., Talón, M. and Robles, M. (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**: 3674–3676.

Dickens, J.C., Callahan, F.E., Wergin, W.P. and Erbe, E.F. (1995) Olfaction in a hemimetabolous insect antennal-specific protein in adult *Lygus lineolaris* (Heteroptera, Miridae). *J Insect Physiol* **41**: 857–867.

Fan, J., Francis, F., Liu, Y., Chen, J.L. and Cheng, D.F. (2011) An overview of odorant-binding protein functions in insect peripheral olfactory reception. *Genet Mol Res* **10**: 3056–3069.

Farias, L.R., Paula, D.P., Zhou, J.J., Liu, R., Pappas, G.J., Moraes, M.C.B. *et al.* (2014) Identification and expression profile of two putative odorant-binding proteins from the neotropical brown stink bug, *Euschistus heros* (Fabricius) (Hemiptera: Pentatomidae). *Neotropical Entomol* **43**:106–114.

Farias, L.R., Schimmelpfeng, P.H.C., Togawa, R.C., Costa, M.M.C., Grynberg, P., Martins, N.F. *et al.* (2015) Transcriptome-based identification of highly similar odorant-binding proteins among neotropical stink bugs and their egg parasitoid. *PLoS ONE* **10**: 0132286doi: 10.1371/journal.pone.0132286.

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.

Gascuel, O. (1997) BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data. *Mol Biol Evol* **14**:685–695.

Galindo, K. and Smith, D.P. (2001) A large family of divergent *Drosophila* odorant-binding proteins expressed in gustatory and olfactory sensilla. *Genetics* **159**: 1059–1072.

Gu, S.-H., Wang, S.-P., Zhang, X.-Y., Wu, K.-M., Guo, Y.-Y., Zhou, J.-J. *et al.* (2011a) Identification and tissue distribution of odorant binding protein genes in the lucerne plant bug *Adelphocoris lineolatus* (Goeze). *Insect Biochem Mol Biol* **41**: 254–263.

Gu, S.H., Wang, W.X., Wang, G.R., Zhang, X.Y., Guo, Y.Y., Zhang, Z.D. *et al.* (2011b) Functional characterization and immunolocalization of odorant binding protein 1 in the lucerne plant bug, *Adelphocoris lineolatus* (Goeze). *Arch Insect Biochem* **77**: 81–98.

Gu, S.-H., Wu, K.-M., Guo, Y.-Y., Field, L.M., Pickett, J.A., Zhang, Y.-J. *et al.* (2013) Identification and expression profiling of odorant binding proteins and chemosensory proteins between two wingless morphs and a winged morph of the cotton aphid *Aphis gossypii* Glover. *PLoS ONE* **8**: e73524.

Guex, N. and Peitsch, M.C. (1997) SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* **18**: 2714–2723.

Gong, Z.J., Liu, S., Jiang, Y.D., Zhou, W.-W., Liang, Q.-M., Cheng, J. *et al.* (2015) Construction and analysis of antennal cDNA library from rice striped stem borer, Chilo suppressalis (walker) (Lepidoptera: Pyralidae), and expression profiles of putative odorant-binding protein and chemosensory protein genes. *Arch Insect Biochem Physiol* **89**: 35–53.

Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J. *et al.* (2013) *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc* **8**: 1494–1512.

He, M. and He, P. (2014) Molecular characterization, expression profiling, and binding properties of odorant binding protein genes in the whitebacked planthopper, *Sogatella furcifera*. *Comp Biochem Physiol Part B* **174**: 1–8.

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.

Hoebeke, E.R. and Carter, M.E. (2003) *Halyomorpha halys* (Stål) (Heteroptera: Pentatomidae): a polyphagous plant pest from Asia newly detected in North America. *Proc Entomol Soc Wash* **105**: 225–237.

Hua, J.F., Zhang, S., Cui, J.J., Wang, D.J., Wang, C.Y., Luo, J.Y. *et al*. (2012) Identification and binding characterization of three odorant binding proteins and one chemosensory protein from *Apolygus lucorum* (Meyer-Dur). *J Chem Ecol* **38**: 1163–1170.

Hull, J.J., Perera, O.P. and Snodgrass, G.L. (2014) Cloning and expression profiling of odorant- binding proteins in the tarnished plant bug, *Lygus lineolaris*. *Insect Mol Biol* **23**: 78–97.

Ishiwatari, T. (1974) Studies on the scent of stink bugs (Hemiptera: Pentatomidae), alarm pheromone activity. *Appl Ent Zool* **9**: 153–158.

Jeong, Y.T., Shim, J., Oh, S.R., Yoon, H.I., Kim, C.H., Moon, S.J. *et al*. (2013) An odorant-binding protein required for suppression of sweet taste by bitter chemicals. *Neuron* **79**: 725–737.

Ji, P., Gu, S.-H., Liu, J.-T., Zhu, X.-Q., Guo, Y.-Y., Zhou, J.-J. *et al*. (2013) Identification and expression profile analysis of odorant-binding protein genes in *Apolygus lucorum* (Hemiptera: Miridae). *Appl Entomol Zool* **48**: 301–311.

Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C. *et al*. (2014) InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**: 1236–1240.

Katoh, K., Misawa, K., Kuma, K. and Miyata, T. (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* **30**: 3059–3066.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S. *et al*. (2012) Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**: 1647–1649.

Lee, D.H., Short, B.D., Joseph, S.V., Bergh, J.C. and Leskey, T.C. (2013) Review of the biology, ecology, and management of *Halyomorpha halys* (Hemiptera: Pentatomidae) in China, Japan, and the Republic of Korea. *Environ. Entomol* **42**: 627–641.

Li, Z.-X., Pickett, J.A., Field, L.M. and Zhou, J.-J. (2005) Identification and expression of odorant-binding proteins of the malaria-carrying mosquitoes *Anopheles gambiae* and *Anopheles arabiensis*. *Arch Insect Biochem Physiol* **58**: 175–189.

Nielson, A.L., Shearer, P.W. and Hamilton, G.C. (2008) Toxicity of insecticides to *Halyomorpha halys* (Hemiptera: Pentatomidae) using glass-vial bioassays. *J Econ Entomol* **101**: 1439–1442.

Panizzi, A.R., McPherson, J.E., James, D.G., Javahery, M. and McPherson, B.A. (2000) Stink bugs (Pentatomidae). In *Heteroptera of Economic Importance* (Panizi, C.W. and Sa, A.R.), p. 421–474. CRC, New York, NY.

Parra, G., Bradnam, K., Ning, Z., Keane, T. and Korf, I. (2009) Assessing the gene space in draft genomes. *Nucleic Acids Res* **37**: 289–297.

Pelosi, P. and Maida, R. (1995) Odorant-binding proteins in insects. *Comp Biochem Physiol B* **111**: 503e–5514.

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

R Development Core Team (2008) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Remmert, M., Biegert, A., Hauser, A. and Soding, J. (2012) HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nat Methods* **9**: 173–175.

Rice, K.B., Bergh, C.J., Bergmann, E.J., Biddinger, D.J., Dieckhoff, C., Dively, G. *et al*. (2014) Biology, ecology, and management of brown marmorated stink bug (Hemiptera: Pentatomidae). *J Integr Pest Mgmt* **5**: 2014. DOI: http://dx.doi.org/10.1603/IPM14002.

Ruijter, J.M., Ramakers, C., Hoogaars, W.M., Karlen, Y., Bakker, O., van den Hoff, M.J. *et al*. (2009) Amplification efficiency: linking baseline and bias in the analysis of quantitative PCR data. *Nucleic Acids Res* **37**: e45.

Ruijter, J.M., Lorenz, P., Tuomi, J.M., Hecker, M. and van den Hoff, M.J.B. (2014) Fluorescent-increase kinetics of different fluorescent reporters used for qPCR depend on monitoring chemistry, targeted sequence, type of DNA input and PCR efficiency. *Microchim Acta* **181**: 1689–1696.

Sali, A. and Blundell, T.L. (1993) Comparative protein modeling by satisfaction of spatial restraints. *J Mol Biol* **234**: 779–815.

Shanbhag, S.R., Hekmat-Scafe, D., Kim, M.-S., Park, S.-K., Carlson, J.R., Pikielny, C. *et al*. (2001a) Expression mosaic of odorant-binding proteins in *Drosophila* olfactory organs. *Microsc Res Tech* **55**: 297–306.

Shanbhag, S.R., Park, S.K., Pikielny, C.W. and Steinbrecht, R.A. (2001b) Gustatory organs of *Drosophila melanogaster*: fine structure and expression of the putative odorant-binding protein PBPRP2. *Cell Tissue Res* **304**: 423–437.

Sun, L., Xiao, H.J., Gu, S.H., Zhou, J.J., Guo, Y.Y., Liu, Z.W. *et al*. (2014) The antennas-specific odorant-binding protein AlinOBP13 of the alfalfa plant bug *Adelphocoris lineolatus* is expressed specifically in basiconic sensilla and has high binding affinity to terpenoids. *Insect Mol Biol* **23**: 417–434.

Supek, F., Bošnjak, M., Škunca, N. and Šmuc, T. (2011) REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS ONE* **6**: e21800. doi:10.1371/journal.pone.0021800

Tamura, K., Stecher, G., Peterson, D., Filipski, A. and Kumar, S. (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol* **30**: 2725–2729.

Tuomi, J.M., Voorbraak, F., Jones, D.L. and Ruijter, J.M. (2010) Bias in the Cq value observed with hydrolysis probe based quantitative PCR can be corrected with the estimated PCR efficiency value. *Methods* **50**: 313–322.

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. and Riddiford, L.M. (1981) Pheromone binding and inactivation by moth antennae. *Nature* **293**: 161–163.

Vogt, R.G., Rybczynski, R. and Lerner, M.R. (1991) Molecular cloning and sequencing of general odorant-binding proteins GOBPI and GOBP2 from the tobacco hawk moth *Manduca sexta*: comparisons with other insect OBPs and their signal peptides. *J Neurosci* **17**: 2972–2984.

Weber, D.C., Leskey, T.C., Walsh, G.C. and Khrimian, A. (2014) Synergy of aggregation pheromone with methyl (*E,E,Z*)−2,4,6-decatrienoate in attraction of *Halyomorpha halys* (Hemiptera: Pentatomidae). *J Econ Entomol* **107**: 1061–1068.

Xu, P.X., Zwiebel, L.J. and Smith, D.P. (2003) Identification of a distinct family of genes encoding atypical odorant-binding proteins in the malaria vector mosquito, *Anopheles gambiae*. *Insect Mol Biol* **12**: 549–560.

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**: 632–645.

Yamada, T., Letunic, I., Okuda, S., Kanehisa, M. and Bork, P. (2011) iPath2.0: interactive pathway explorer. *Nucleic Acids Res* **39**: W412–W415

Yuan, H.-B., Ding, Y.-X., Gu, S.-H., Sun, L., Zhu, X.-Q., Liu, H.-W. *et al*. (2015) Molecular characterization and expression profiling of odorant-binding proteins in *Apolygus lucorum*. *PLoS ONE* **10**: e0140562.

Zhou, J.J., Huang, W.S., Zhang, G.A., Pickett, J.A. and Field, L.M. (2004) "Plus-C" odorant- binding protein genes in two *Drosophila* species and the malaria mosquito *Anopheles gambiae*. *Gene* **327**: 117–129.

Zhou, J.J., Vieira, F.G., He, X.L., Smadja, C., Liu, R.H., Rozas, J. *et al*. (2010) Genome annotation and comparative analyses of the odorant-binding proteins and chemosensory proteins in the pea aphid *Acyrthosiphon pisum*. *Insect Mol Biol* **19**: 113–122.

Zhou, S.S., Sun, Z., Ma, W., Chen, W. and Wang, M.Q. (2014) *De novo* analysis of the *Nilaparvata lugens* (Stal) antennas transcriptome and expression patterns of olfactory genes. *Comp Biochem Phys D* **9**: 31–39.

Zhu, G., Bu, W., Gao, Y. and Liu, G. (2012) Potential geographic distribution of brown marmorated stink bug invasion (*Halyomorpha halys*). *PLoS ONE* **7**: e31246.

Supporting information

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

Table S1. Number (and % completeness) of core eukaryotic genes (CEGs) mapped to the contigs of the brown marmorated stink bug. Group 1 is represented by the least conserved CEGs whereas group 4 has the most conserved ones. #Prots, number of 248 ultra-conserved CEGs present in the transcriptome; % compl, percentage of 248 ultra-conserved CEGs present; Total, total number of CEGs present including putative orthologues; Average, average number of orthologues per CEG; % Ortho, percentage of detected CEGs that have more than one orthologue.

Table S2. Protein Data Bank (PDB) templates used in homology modelling of odorant-binding protein (OBP) candidate proteins, showing per cent sequence similarity and coverage.

Table S3. Identity matrix (%) used in the alignments amongst the brown marmorated stink bug putative odorant-binding proteins (OBPs).

Table S4. Nucleotide sequences of the primer pairs used for the amplification of the putative *Halyomorpha halys* odorant-binding proteins (HhalOBPs) identified in the brown marmorated stink bug during the gene expression analysis by qPCR.

Figure S1. Metabolic pathway mapping of the antennal brown marmorated stink bug transcripts. Metabolic pathways were generated using iPath v. 2. The connections in red are the ones present in the antennae.