



Distinct Subfamilies of Odorant Binding Proteins in Locust (Orthoptera, Acrididae): Molecular Evolution, Structural Variation, and Sensilla-Specific Expression

Xingcong Jiang¹, Jürgen Krieger², Heinz Breer¹ and Pablo Pregitzer^{1*}

¹ Institute of Physiology, University of Hohenheim, Stuttgart, Germany, ² Department of Animal Physiology, Institute of Biology/Zoology, Martin Luther University Halle-Wittenberg, Halle, Germany

OPEN ACCESS

Edited by:

Shuang-Lin Dong,
Nanjing Agricultural University, China

Reviewed by:

Paolo Pelosi,
University of Pisa, Italy
Dan-Dan Zhang,
Lund University, Sweden
Long Zhang,
China Agricultural University, China

*Correspondence:

Pablo Pregitzer
p_pregitzer@uni-hohenheim.de

Specialty section:

This article was submitted to
Invertebrate Physiology,
a section of the journal
Frontiers in Physiology

Received: 20 July 2017

Accepted: 11 September 2017

Published: 26 September 2017

Citation:

Jiang X, Krieger J, Breer H and
Pregitzer P (2017) Distinct Subfamilies
of Odorant Binding Proteins in Locust
(Orthoptera, Acrididae): Molecular
Evolution, Structural Variation, and
Sensilla-Specific Expression.
Front. Physiol. 8:734.
doi: 10.3389/fphys.2017.00734

Odorant binding proteins (OBPs) play an important role in insect olfaction, facilitating transportation of odorant molecules in the sensillum lymph. While most of the researches are concentrated on Lepidopteran and Dipteran species, our knowledge about Orthopteran species is still very limited. In this study, we have investigated OBPs of the desert locust *Schistocerca gregaria*, a representative Orthopteran species. We have identified 14 transcripts from a *S. gregaria* antennal transcriptome encoding SgreOBPs, and recapitulated the phylogenetic relationship of SgreOBPs together with OBPs from three other locust species. Two conserved subfamilies of classic OBPs have been identified, named I-A and II-A, exhibiting both common and subfamily-specific amino acid motifs. Distinct evolutionary features were observed for subfamily I-A and II-A OBPs. Surface topology and interior cavity were elucidated for OBP members from the two subfamilies. Antennal topographic expression revealed distinct sensilla- and cellular- specific expression patterns for SgreOBPs from subfamily I-A and II-A. These findings give first insight into the repertoire of locust OBPs with respect to their molecular and evolutionary features as well as their expression in the antenna, which may serve as an initial step to unravel specific roles of distinct OBP subfamilies in locust olfaction.

Keywords: locust, *Schistocerca gregaria*, odorant binding protein, evolution, structure, sensilla

INTRODUCTION

In insects, the process of olfactory signal processing begins in hair-like cuticle appendages, called sensilla, located mainly on the antennae and palps (Steinbrecht, 1996; Hansson and Stensmyr, 2011; Suh et al., 2014). Olfactory sensory neurons (OSNs) project their dendrites into the lumen of the sensillar hairs, which is filled with sensillum lymph (Hansson and Stensmyr, 2011; Suh et al., 2014). The hydrophobic odorant molecules enter the sensillum via the porous cuticle and have to pass the aqueous lymph till reaching the chemosensory membrane of the sensory neurons (Vogt et al., 1999; Leal, 2013; Suh et al., 2014). This passage is thought to be mediated by small soluble proteins enriched in the sensilla lymph, the so called odorant binding proteins (OBPs), which are produced and secreted by accessory cells (Pelosi et al., 2006, 2017). OBPs are polypeptides comprised of

~110–200 amino acids; usually they exhibit a considerable degree of sequence divergence. Based on the number of conserved cysteine (C)-residues, several subtypes are discriminated. Whereas, the pattern of six conserved C-residues represents a hallmark of classic OBPs (Pelosi et al., 2006), OBPs with more or with less C-residues are designated as plus-C and minus-C OBPs (Zhou et al., 2004; Foret and Maleszka, 2006). In addition, atypical OBPs have been classified which may originate from a fusion of two classic OBPs (Xu et al., 2003; Vieira and Rozas, 2011). Typically, the tertiary structure of insect OBPs consists of six α -helices forming an interior binding cavity. This structure is maintained and stabilized by disulfide bridges formed by conserved C-residues (Leal et al., 1999; Scaloni et al., 1999; Sandler et al., 2000). However, OBP structures with more than six helices have been reported (Horst et al., 2001; Lagarde et al., 2011). It is also proposed that the C-terminal domain that is variable in length can spatially interfere with the interior binding cavity and thus may affect the ligand binding mechanism (Damberger et al., 2000; Horst et al., 2001; Tegoni et al., 2004; Pelosi et al., 2017).

Most of our current knowledge of insect OBPs is based on studies of species from the taxa Lepidoptera and Diptera (Hekmat-Scafe et al., 2002; Vogt et al., 2002; Leal, 2013; Pelosi et al., 2017). The desert locust, *Schistocerca gregaria* is a representative of the taxa Orthoptera, which is quite distant from the orders Lepidoptera and Diptera on the phylogenetic scale (Wheeler et al., 2001; Vogt et al., 2015) and as hemimetabolous insects their developmental process differ significantly from that of holometabolous insects. Very little is known about OBPs in Orthoptera; only a limited number of sequences have recently been reported for a few locust species: *Locusta migratoria* (Ban et al., 2003; Xu et al., 2009; Yu et al., 2009), *Oedaleus asiaticus* (Zhang et al., 2015), and *Ceracris kiangsu*. Information about the expression of locust OBPs in the olfactory sensilla is limited to LmigOBP1, which was found to be expressed in sensilla trichodea and sensilla basiconica (Jin et al., 2005). Concerning another olfactory sensillum type, the sensilla coeloconica, a possible expression of OBPs has rarely been documented even in holometabolous insect species (Larter et al., 2016). Incidentally, the crystal structure of locust OBPs has only been resolved for LmigOBP1, which establishes a unique seven- α -helices structure (Zheng et al., 2015). The possibility of structural differences between locust OBPs is still an open question.

In order to extend our knowledge about OBPs in Orthopteran locust species, in the current study we have performed a systematic characterization of locust OBPs with respect to the molecular evolution, structural variation and sensilla-specific expression. Based on the OBP sequences of *S. gregaria* newly identified from an antennal transcriptome and the documented OBP sequences from other locust species, we conducted a phylogenetic analysis of the current locust OBP repertoire. The emerging two subfamilies of classic OBPs were compared for sequence divergence, selection pressure and variation of the predicted tertiary structure in detail. Analysis of the topographic expression pattern revealed that the molecular and phylogenetic distinctness between the two subfamilies are accompanied by a sensilla-specific expression pattern.

MATERIALS AND METHODS

Identification of *S. gregaria* OBP Transcripts

A *S. gregaria* antennal transcriptome database was generated comprising a total of 55,060 contigs with an N50 of 2,223 bp. The strategy of homology-mining was adopted to identify the candidate OBP transcripts. We retrieved documented OBPs from different insect species including *Anopheles gambiae* (AgamOBPs, Diptera), *Apis mellifera* (AmelOBPs, Hemiptera), *Drosophila melanogaster* (DmelOBPs, Diptera), *Tribolium castaneum* (TcasOBPs, Coleoptera), *Acyrtosyphon pisum* (ApisOBPs, Hemiptera), *Bombyx mori* (BmorOBPs, Lepidoptera) (Vieira and Rozas, 2011), *Blattella germanica* (BgerOBPs, Blattaria) (Niu et al., 2016), and *Zootermopsis nevadensis* (ZnevOBPs, Isoptera) (Terrapon et al., 2014), as well as from three other locust species, including *L. migratoria* (LmigOBPs) (Ban et al., 2003; Yu et al., 2009), *O. asiaticus* (OasiOBPs) (Zhang et al., 2015), and *C. kiangsu* (CkiaOBPs). Using the collected sequences as queries, we conducted a local tBLASTx search on BioEdit 7.2.5 against the transcriptome database with an *E*-value $< 10^{-5}$. Annotation of the screened contigs was inspected by performing tBLASTx and BLASTp search against non-redundant (nr) protein database in NCBI (Bethesda, MD, USA). The extracted contigs which putatively encode OBPs were in turn used as new queries to identify additional candidates using tBLASTx and BLASTp methods. Open reading frames in the identified OBP transcripts were inspected by Genamics Expression (Hamilton, New Zealand). Accession numbers for the newly identified SgreOBPs and other locust OBPs are deposited in the Supplementary Material.

Characterization of Consensus Amino Acid Motifs

Signatures of sequence divergence underlying locust subfamily I-A and II-A OBPs were addressed by identifying consensus amino acid motifs. Toward that goal, the online MEME SUITE v. 4.11.2 (<http://meme-suite.org/tools/meme>) was used (Bailey et al., 2009), with the default setting (motif width: 6–50 amino acids; motif distribution: zero or one occurrence per sequence). The output comprised six consensus motifs which was ascertained to be sufficient to recapitulate the sequence information of subfamily I-A and II-A. The identified six motifs were also utilized to target sequences of the locust OBP repertoire to obtain the motif match degree (match *E*-value) using MAST module (Motif Alignment and Search Tool) implemented in MEME SUITE. The motif match *E*-value assesses statistical significance of the consensus motif toward a targeted sequence based on its log likelihood level and the occurrence frequencies of background amino acids. The default statistical significant threshold setting was e^{-5} .

Phylogenetic Analysis

The OBP amino acid sequences from four hitherto documented locust species were utilized to recapitulate the phylogenetic relationship: 16 from *L. migratoria*, 15 from *O. asiaticus*, 7 from *C. kiangsu* and the currently identified 14 candidates

from *S. gregaria*. Amino acid sequences of OBPs from the four locust species are deposited in the Supplementary Material. The predicted signal peptide (SP) on the N-terminal domain was deleted before the sequences being further investigated due to two reasons: (1) SP is cut off in post-translational modification when the protein is secreted into the sensillum lymph; (2) SP exhibits a certain degree of sequence divergence but may contain limited bio-information (Vieira et al., 2007). Prediction of SP was based on SignalP 4.1 (<http://www.cbs.dtu.dk/services/SignalP/>) (Petersen et al., 2011). Multiple sequence alignments were conducted by MAFFT v. 7 (<http://mafft.cbrc.jp/alignment/server/>) using the algorithm E-INS-I, which is accuracy favored and is suitable for sequences with multiple conserved domains (Kato and Standley, 2013). After the alignment, Gblocks v. 0.91b (http://molevol.cmima.csic.es/castresana/Gblocks_server.html) was used to inspect the poorly aligned sites and divergent regions (Castresana, 2000). To search an optimal amino acid substitution model, we chose the Find Best Protein Model implemented in MEGA 6.0 which performs a comprehensive parametric assessment (e.g., BIC scores, AICc value, lnL value) (Tamura et al., 2013). The Whelan and Goldman model (WAG), discrete GAMMA distribution (G) and an assumed fraction of evolutionary invariable sites (I) was considered to describe the substitution best. RAxML v. 8.2.9 implemented in the CIPRES Science Gateway (<https://www.phylo.org/>) was used for the locust OBP phylogeny construction (Miller et al., 2012; Stamatakis, 2014). A search of best scoring maximum likelihood tree (-f a) was launched, supported by 1,000 rapid bootstrap iterations (autoMRE based bootstopping criterion). The generated maximum likelihood tree was graphically edited by FigTree v. 1.4.3 (<http://tree.bio.ed.ac.uk/software/figtree/>). A similar strategy was applied to analyze the phylogenetic relationship between locust OBPs and OBPs from eight other insect species. In brief, SignalP, MAFFT, and Gblocks were used to prepare the multiple sequence alignment; RAxML was responsible for building the maximum likelihood tree (-f a, 1,000 iteration) using the proposed best fitting substitution model (WAG+G+I) by MEGA.

Selection Constraint on Locust OBP Repertoire

The nucleotide coding sequences underlying the locust OBP repertoire (see Supplementary Material) were aligned in accordance with the multiple sequence alignment from the above mentioned phylogenetic analysis using TranslatorX (<http://translatorx.co.uk/>). The sequence order of alignment was guided by the constructed phylogenetic tree mentioned above. The signatures of selection regime acting on sequences of the locust OBP phylogeny were estimated by resolving three principle concepts: the non-synonymous substitution rate (dN), synonymous substitution rate (dS) and the ω rate (dN/dS). Toward that, HyPhy batch program was utilized which implements maximum likelihood estimate and post-likelihood ratio test (Kosakovsky Pond et al., 2005). A local fit model (MG94xREV_3x4 substitution model) was adopted (Kosakovsky Pond et al., 2009), and each single branch in the locust OBP

phylogeny was assigned with a unique set of dN and dS values, assuming the branch-to-branch variant ω rates. To support the local fit model, we additionally conducted a coarse estimate of the ω rate using the alternative global fit model, assuming invariable ω rate shared by different phylogenetic branches. A likelihood ratio test compared the results obtained from two distinct models, and strongly favored the local fit model ($P = 10^{-3}$). Normality distribution of dN, dS, and the ω rates was assessed by D'Agostino-Pearson test, and the statistical difference was evaluated by non-parametric Mann-Whitney *U*-test. GraphPad Prism 5.0 was used to analyze the data and generate the diagrams (San Diego, CA, USA).

Synthesis of Riboprobes For *in Situ* Hybridization

The coding sequences of six SgreOBPs from locust OBP subfamily I-A and II-A were amplified, sequenced and then cloned into the pGEM-T vectors (Invitrogen) for subsequent transcription. Linearized pGEM-T vectors carrying SgreOBPs coding sequences were utilized to synthesize digoxigenin (Dig) and biotin (Bio) labeled anti-sense and sense RNA probes using the T7/SP6 RNA transcription system (Roche, Germany). The sense (s) and antisense (as) primers used for amplification of the SgreOBP sequences were:

SgreOBP1 s, ctgggacgtcaacatgaaact;
 SgreOBP1 as, aatgcacgaactaccaggctg;
 SgreOBP5 s, ggccgcgcttctctcataagga;
 SgreOBP5 as, cggccctggcgcagcacctgcatt;
 SgreOBP6 s, acagcacaccaccgtcacac;
 SgreOBP6 as, ggtgcttcttgaaagggcac;
 SgreOBP10 s, gcgtatcaccggctgtgta;
 SgreOBP10 as, agtctcactctgccagcga;
 SgreOBP11 s, tggaccgcagcacaaca;
 SgreOBP11 as, cgatagcgtatgccctttcac;
 SgreOBP14 s, ctgttgggtcagctcctgt;
 SgreOBP14 as, gtcgtgacagctcctccactg

In Situ Hybridization

Antennae of adult *S. gregaria* were dissected and embedded in Tissue-Tek O.C.T. Compound (Sakura Finetek Europe, The Netherlands). Cryosections at 12 μ m were thaw mounted on SuperFrost Plus slides (Menzel-Gläser, Braunschweig, Germany) at -21°C (Jung CM300 cryostat). RNA *in situ* hybridization (ISH) was conducted as previously reported (Yang et al., 2012; Guo et al., 2013; Jiang et al., 2016). Section were fixed (4% paraformaldehyde in 0.1 M NaHCO_3 , pH 9.5) at 4°C for 22 min. The following consecutive steps were conducted at room temperature: a wash for 1 min in PBS (phosphate buffered saline = 0.85% NaCl, 1.4 mM KH_2PO_4 , 8 mM Na_2HPO_4 , pH 7.1), an incubation for 10 min in 0.2 M HCl, another wash for 1 min in PBS, an incubation for 10 min in acetylation solution (0.25% acetic anhydride freshly added in 0.1 M triethanolamine) and washes for three times in PBS (3 min each). Sections were prehybridized for 1 h at 60°C in hybridization buffer (50% formamide, $5\times$ SSC, 50 $\mu\text{g/ml}$ heparin, and 0.1% Tween-20). 100 μl hybridization solution containing the labeled RNA in hybridization buffer was placed onto the tissue section. A

coverslip was placed on top and slides were incubated in a moisture box at 60°C overnight (18–20 h). After hybridization, slides were washed twice for 30 min in 0.1× SSC at 60°C, then each slide was treated with 1 ml 1% blocking reagent (Roche) for 40 min at room temperature.

Visualization of Dig-labeled probe hybridizations was achieved by using an anti-Dig alkaline phosphatase (AP) conjugated antibody (1:500, Roche) and NBT/BCIP substrate. Antennal sections were analyzed on a Zeiss Axioskope2 microscope (Zeiss, Oberkochen, Germany) equipped with Axiovision software. For two-color FISH visualization of hybridized probes was performed by using an anti-Dig AP-conjugated antibody in combination with HNPP/Fast Red (Roche) for Dig-labeled probes and an anti-biotin streptavidin horse radish peroxidase-conjugate together with fluorescein-tyrarnides as substrate (TSA kit, Perkin Elmer, MA, USA) for Bio-labeled probes. Sections from FISH experiments were analyzed with a Zeiss LSM510 Meta laser scanning microscope (Zeiss, Oberkochen, Germany). Confocal images stacks were processed by ZEN 2009 software. The pictures shown represent projections of optical planes selected from confocal image stacks. For clear data presentation, images were only adjusted in brightness and contrast. Antennal sections of both male and female antennae were analyzed using each generated probe. No obvious difference between sexes regarding the labeling intensity and labeling pattern was observed. Thus, only the images of male antenna were adopted in this study.

Structure Modeling and Electrostatic Potential

In silico simulation of OBP tertiary structure was performed by I-TASSER server (<http://zhanglab.ccmb.med.umich.edu/I-TASSER/>) (Roy et al., 2010), which implements the iterative template threading refinement making full use of established homologous protein structures. PyMol was used to visualize the simulated protein tertiary structures (DeLano, 2002). The molecular surface was solvent excluded and the solvent radius was set 1.4 as default. APBS plug (Unni et al., 2011) implemented in PyMol was employed to calculate the surface electrostatic potentials in the range of -6 to 6 kT/e, and was presented as blue-red hue gradient.

RESULTS

Identification, C-Skeleton Pattern and Phylogenetic Relationship of Locust OBPs

Toward an identification of OBPs from *S. gregaria* and a comprehensive characterization of OBPs in locust species, we have performed a homology-based data mining of an antennal transcriptome which resulted in 14 transcripts putatively encoding SgreOBPs. Subsequently, a multiple sequence alignment was conducted addressing the amino acid sequences of the newly identified SgreOBPs together with hitherto documented OBPs from three other locust species: 16 from *L. migratoria*, 15 from *O. asiaticus* and 7 from *C. kiangsu*. Several OBP subtypes could be categorized based on the number

of conserved C-residues (**Figure 1A**). First, 33 OBPs were classified as classic OBPs comprising six conserved C-residues, the hallmark of classic OBPs. Second, 15 OBPs were categorized in two types of plus-C OBPs harboring more than six conserved C-residues. Finally, only one minus-C OBP with less than six conserved C-residue and three atypical OBPs with extraordinary long stretches between conserved C1 and C2 were identified.

As a next step, we analyzed the phylogenetic relationship of the locust OBP repertoire by constructing a phylogenetic tree utilizing the maximum likelihood algorithm and bootstrap iterations. The emerging picture indicated that the repertoire of locust OBPs can be divided into four major families (I–IV), which apparently split at the internal nodes (**Figure 1B**). We further classified three additional subfamilies (I-A, II-A, and III-A), based on the presence of higher bootstrap support (above 80%) on the divergent nodes. It is noteworthy that subfamily I-A and II-A both represent classic OBPs and each subfamily apparently comprise three distinct groups with 3–4 orthologous OBPs from different locust species (**Figure 1B**). Within each subfamily, the sequence identity between OBPs from different groups ranged from 28 to 35%; OBP members within each ortholog group exhibit generally above 80% sequence identity. Incidentally, plus-C OBPs type-A converged onto a subfamily III-A and segregated from their counterparts plus-C OBPs type-B and classic OBPs. Together, the data indicate a considerable degree of orthology in the OBP repertoires across the four analyzed locust species and no marked species-specific expansion within the OBP phylogeny.

Elucidation of Subfamily-Specific Consensus Amino Acid Motifs

To better elucidate the clustering regime of individual subfamilies, we analyzed the consensus amino acid motifs characteristics underlying subfamily I-A and II-A OBPs. The local consensus motifs were calculated by recapitulating repeatedly occurring sequence patterns along OBP sequences. Six consensus motifs with various widths were identified and localized at distinct positions (**Figure 2**). The motif 1 and motif 2 appeared as common motifs in all OBPs of both subfamilies, whereas the other four motifs specifically fit either the repertoire of subfamily I-A OBPs (motif combination 4 and 6) or the repertoire of subfamily II-A OBPs (motif combination 3 and 5). Therefore, two less divergent sequence domains were unraveled by the presence of motif combination 1 and 2, spanning the domains of C2–C3 and C4–C6. In contrast, the sequence domains close to the N-terminus (42 amino acids, motif 3 and motif 4) and ahead of C4 (11–15 amino acids, motif 5 and motif 6) appeared to be more divergent.

Utilizing the six identified consensus motifs in **Figure 2** we have quantified the sequence divergence for the locust OBP repertoire at a local motif scale (**Figure S1**). Apart from subfamilies I-A and II-A, the common motif 1 and motif 2, especially the latter, recapitulate sequence information present in many of the other locust OBPs analyzed (*E*-value below 10^{-5}) indicating particular phylogenetic conservation of these regions. Not surprisingly, the subfamily-specific motifs 3–6 failed to match OBP members (*E*-value above 10^{-5}) other than subfamily

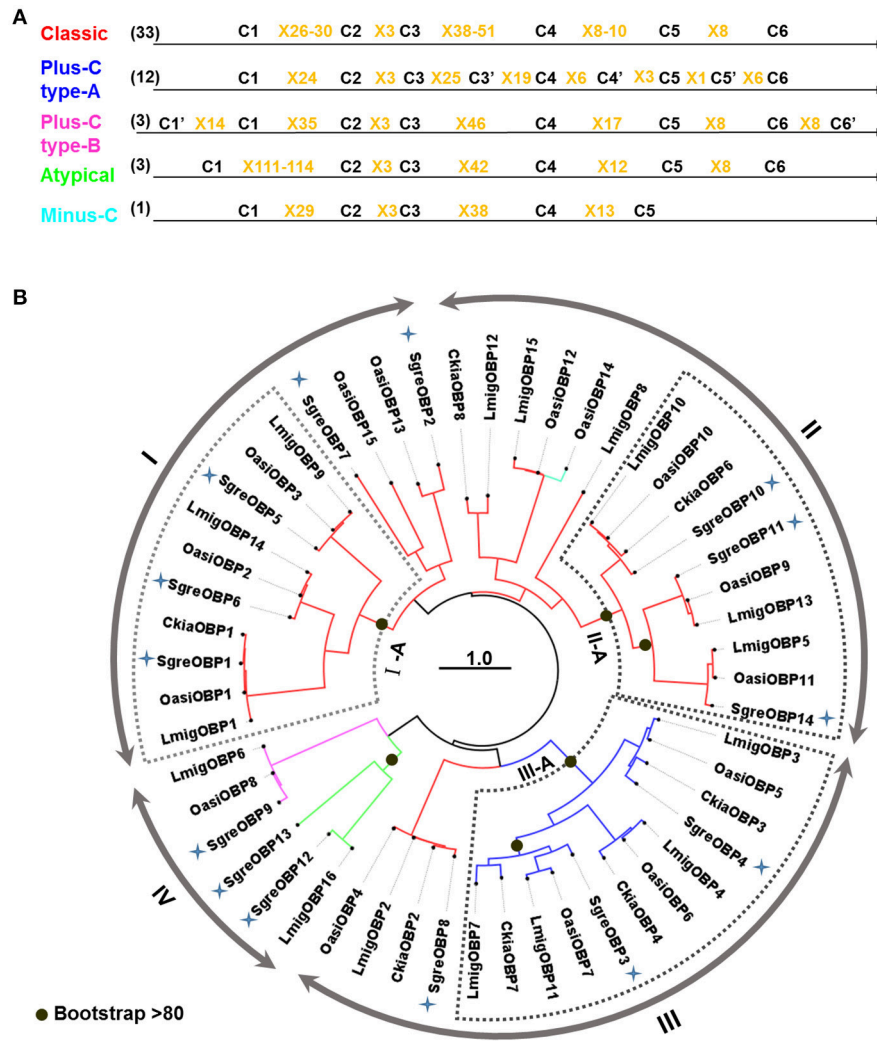


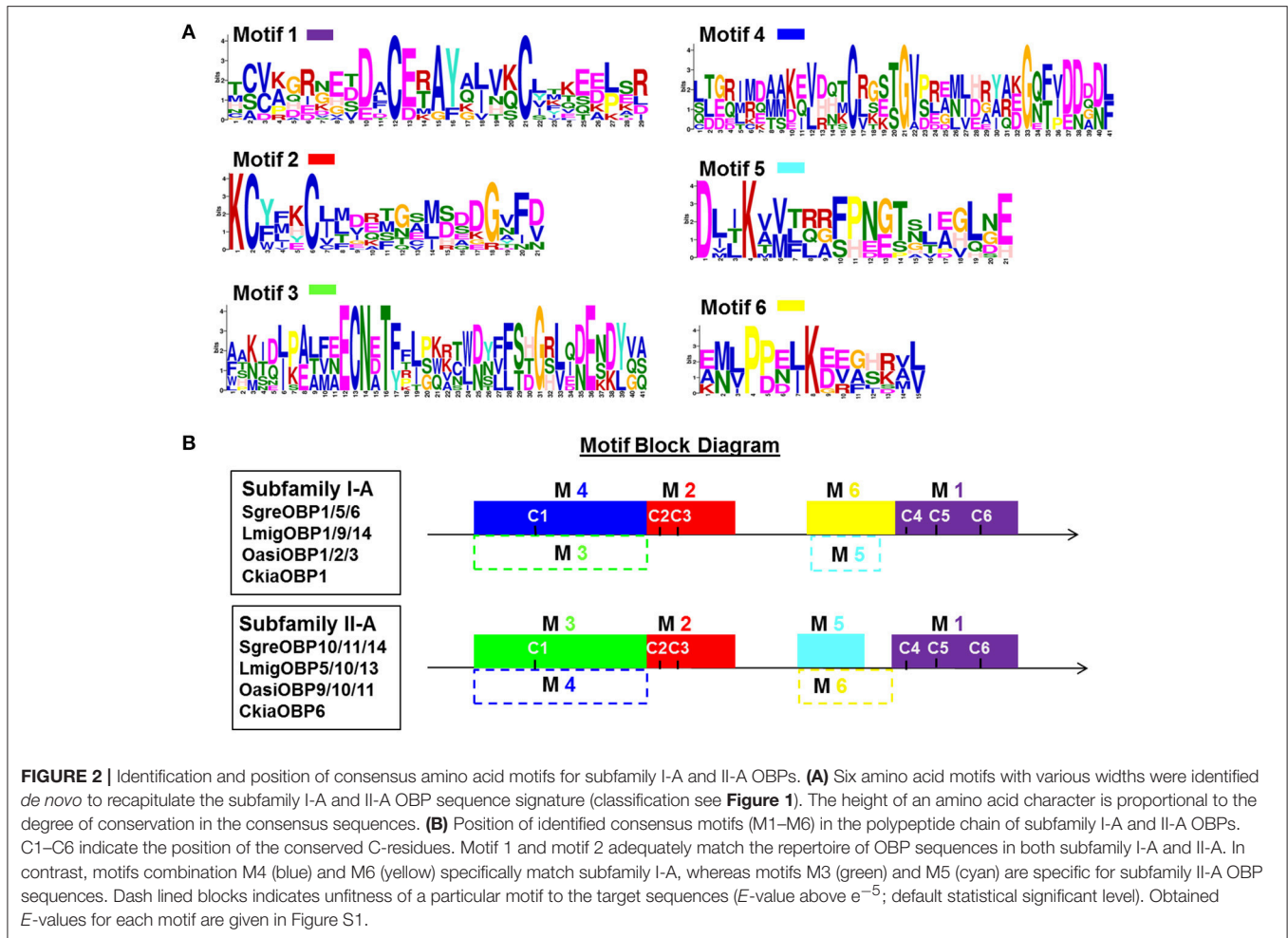
FIGURE 1 | C-residue skeletons and phylogeny of OBPs from four locust species. **(A)** OBPs subtypes were categorized based on the number of conserved C-residues. C-skeleton patterns are based on the multiple sequence alignment of 52 OBP sequences from four locust species. C-residues conserved in all OBPs are shown as C1-C6 in black characters; additional C-residues conserved in the two plus-C OBP types are shown as C'; amino acid between two C-residues are shown as X plus the number of amino acid. The number of each OBP subtype is given in the parenthesis. **(B)** The phylogenetic tree was constructed using the maximum likelihood algorithm supported by 1,000 bootstrap replicates. OBP sequences utilized to generate the tree were derived from four locust species: 14 from *Schistocerca gregaria* (SgreOBPs), 16 from *Locusta migratoria* (LmigOBPs), 15 from *Oedaleus asiaticus* (OasiOBPs) and 7 from *Ceracris kiangsu* (CkiaOBPs). Four primary families (I-IV) are denoted by arrow lines. Further classification of three subfamilies (I-A, II-A, and III-A) was based on the over 80% bootstrap support at the internal node (indicated by black dots). Inner branches in different colors represent OBP subtypes in **(A)**: red, classic OBPs; blue, plus-C OBPs type-A; magenta, plus-C OBPs type-B; green, atypical OBPs; cyan, minus-C OBP. Newly identified SgreOBPs are denoted by blue crosses. The tree is midpoint rooted. Scale bar represents one amino acid substitution per site.

I-A and II-A OBPs, despite a small number of OBPs in family I and family II (Figure S1). Taken together, the motif analysis unraveled the presence of both stabilized and diversified domains residing on the global sequences.

Selection Pressure and Orthology Evolution of Locust Subfamily I-A and II-A

The appearance of two distinct conserved subfamilies in the locust OBP phylogeny, coupled with the clustering pattern of different ortholog groups is presumably a consequence of particular selection regimes. To prove this notion, we have

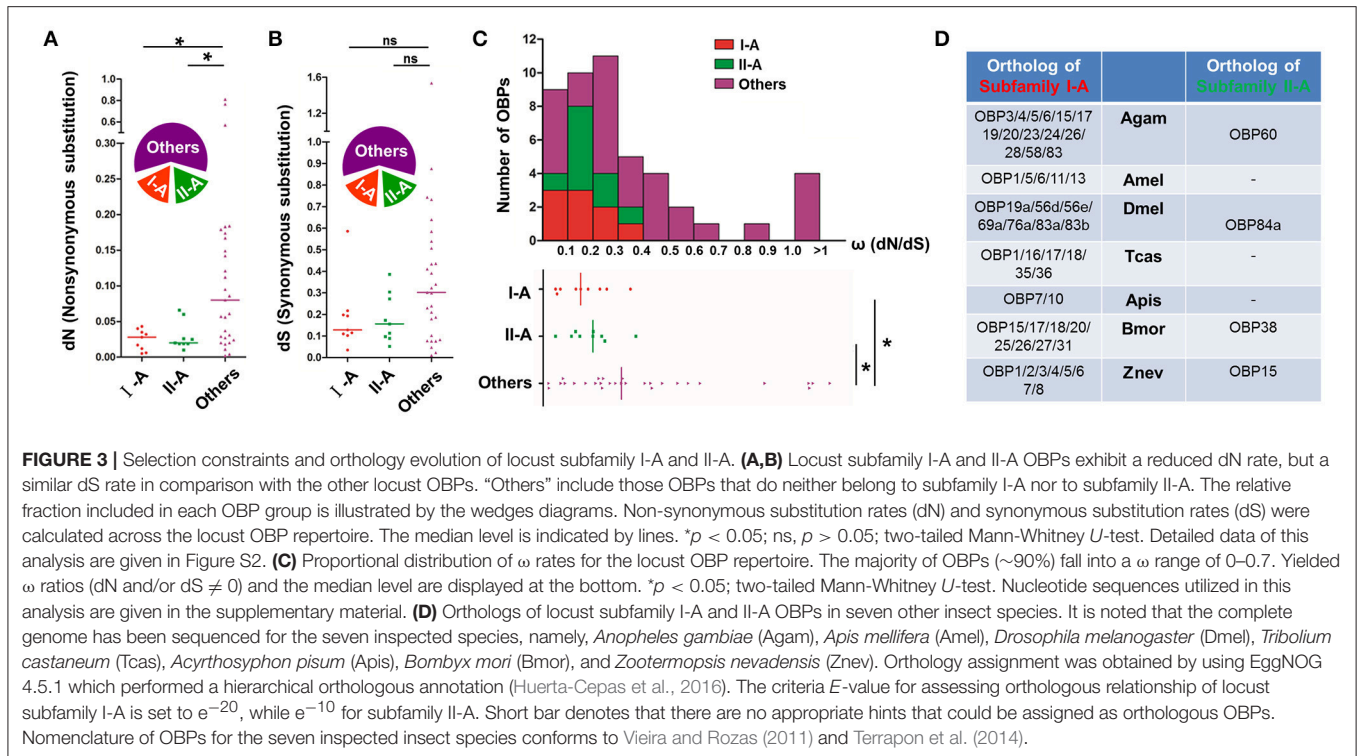
tried to quantify the strength of selection pressure acting on genes encoding the locust OBP repertoire. We analyzed three principal concepts which reflect the selection pressure, namely, the non-synonymous substitution rates (dN), the synonymous substitution rates (dS) and the ω rates (dN /dS) (Figure 3). We found a significantly reduced median dN level for both subfamily I-A (dN = 0.030, $U = 60$, $p = 0.016$, Mann-Whitney U -test) and subfamily II-A (dN = 0.028, $U = 60$, $p = 0.016$, Mann-Whitney U -test), in comparison with that of other OBP members (dN = 0.085, Figure 3A). However, the median dS level appeared to be quite similar among subfamily I-A (dS = 0.12, $p = 0.154$,



$U = 88.5$, Mann-Whitney U -test), subfamily II-A ($dS = 0.16$, $U = 86$, $p = 0.131$, Mann-Whitney U -test) and the other OBP members ($dS = 0.31$, **Figure 3B**). For the ω rates, the values ranged from 0 to 0.7 for nearly 90% of locust OBPs (**Figure 3C**), which is indicative of purifying selection acting on locust OBP repertoire in general. For a few exceptions, ω rates larger than one were found which may indicate a positive selection. Notably, median ω rates for OBPs of subfamily I-A ($\omega = 0.18$, $U = 63$, $p = 0.021$, Mann-Whitney U -test) and subfamily II-A ($\omega = 0.22$, $U = 69$, $p = 0.036$, Mann-Whitney U -test) were significantly reduced in comparison with other OBP members in the phylogeny ($\omega = 0.35$, **Figure 3C**).

Exposed to a similar selection regime, we wondered if orthologous OBPs in other species would undergo similar divergent events in relation to the two locust OBP subfamilies. To address the issue, we made a phylogenetic analysis of the two locust OBP subfamilies and the reference OBPs derived from 8 other insect species which gradually emerged in the course of insect evolution. The analysis revealed that locust subfamily II-A OBPs remained on an intact clade without intermingling with reference OBP genes on the newly constructed phylogenetic tree (**Figure S3**). A different result was obtained for the subfamily

I-A: the original clustering relationship of ortholog groups in locust phylogeny was disrupted and altered with a complex re-clustering pattern integrating reference OBPs. The orthologous relationship (Theißen, 2002) of OBPs between the two locust subfamilies and other species was also inferred. It is found that the number of locust subfamily I-A orthologous OBPs in the inspected insect species expanded considerably, and exhibited a many-to-many orthologous relationship with locust subfamily I-A (**Figure 3D**), with *A. pisum* as apparent exception likely due to a smaller OBP gene repertoire (Zhou et al., 2010). In contrast, the number of locust subfamily II-A orthologous OBPs in other species apparently decreased, and displayed a 1-to-many or 0-to-many orthologous relationship with locust subfamily II-A (**Figure 3D**). Moreover, it was found that locust subfamily II-A OBPs and their orthologous OBPs may share a common ancestor verified by the convergence of a mono phylogenetic clade with the bootstrap support above 70% at the basal divergent node (**Figure S3**). However, the common ancestral status for locust subfamily I-A OBPs and their orthologous OBPs appeared ambiguous because of the absence of evident bootstrap support (**Figure S3**). In sum, our results provide evidence that locust subfamily I-A and II-A OBPs are subject to mutually similar strengthened



purifying selection, whereas distinct divergent events occur during evolution of their orthologous OBPs in other species.

Prediction of Tertiary Structures for OBPs in Subfamily I-A and II-A

The intriguing sequence and evolutionary characteristics underlying locust subfamily I-A and II-A OBPs inspired us to explore the possible concurrent variation of their tertiary structures. Therefore, we have simulated the tertiary structures for OBP members from both two subfamilies covering different ortholog groups and locust species. Parametric estimates toward the accuracy and reliability of the structure prediction was scrutinized, which permitted to investigate structural variation as an exploratory trial. To unravel structural variation between the two subfamilies, we superimposed the backbone structures of those simulated OBPs to LmigOBP1, the hitherto only established crystal structure for the locust OBP repertoire (Zheng et al., 2015). The averaged RMSD score obtained by imposing subfamily II-A OBPs to LmigOBP1 (2.8) doubled that of imposing subfamily I-A OBPs to LmigOBP1 (1.39 in average, Figure S4), indicating an enhanced structural similarity within one subfamily.

Multiple sequence alignment of subfamily I-A OBPs revealed a striking variation on the C-terminal domain (Figure S4). It is known that LmigOBP1 has a prolonged C-terminus with ~17 amino acids to form a seventh α -helix (Zheng et al., 2015). In contrast, the C-terminus in OasiOBP3 and SgreOBP6 is shortened to a 7 amino acids motif and most likely constitute a coiled-coil strand instead of a seventh α -helix (Figure 4); a groove emerged on the collapsed surface due to the shortened

C-terminus. The electrostatic potential pattern varies greatly at a global surface scale as well as on the local C-terminal surface scale (cyan dash line, Figures 4A,C,E). Another striking structural difference is the enlarged cavity of LmigOBP1 bordered by the prolonged C-terminus, whereas the cavity for the other two counterparts, representative of different ortholog groups shrinks to some extent (white dash line, Figures 4B,D,F). Unlike subfamily I-A, the multiple sequence alignment of subfamily II-A OBPs exhibited an aligned C-terminus but an unaligned N-terminus, namely, an extra extension of a 9–10 amino acids motif in the LmigOBP10 ortholog group (Figure S4). Correspondingly, this alteration was predicted to result in a coiled-coil structure on the N-terminal domain for LmigOBP10; at the same surface position, an opening structure was observed on its two counterparts, the OasiOBP11 and SgreOBP11 (Figures 5A,C,E). Apart from that, the surface electrostatic potential profile seems to vary slightly, both at the global surface scale and at the local N-terminal surface scale (cyan dash line, Figures 5C,E), regardless of the extra N-terminal coil present on LmigOBP10. However, the interior cavity could be enriched with negative potentials (LmigOBP10 and SgreOBP11, Figures 5B,F), or with positive potentials (OasiOBP11, Figure 5C).

Topographic Expression Patterns of SgreOBPs from Subfamily I-A and II-A

To approach this question, whether locust subfamily I-A and II-A OBPs may be expressed in different sensillum types and different cells, we set out to unravel the expression patterns of SgreOBPs from the two locust subfamilies in sensilla on the antenna, the major olfactory organ. By adopting RNA *in*

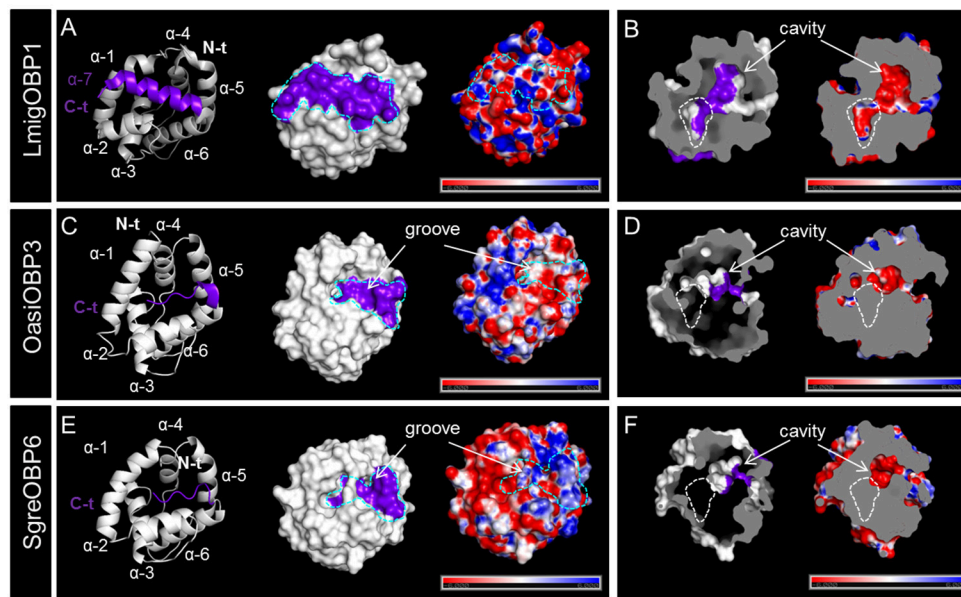


FIGURE 4 | Variations in the C-terminal domain and the interior cavity of subfamily I-A OBPs. **(A,C,E)** Comparison of the backbone structures, surface topologies and surface potentials of LmigOBP1, OasiOBP3 and SgreOBP6 which represent the three different ortholog groups in subfamily I-A. The C-terminal domains (see also Figure S4A) are highlighted in purple on both the backbone structures (left) and the molecular surfaces (middle). The dash line in cyan sketches the surface topology of the C-terminal domain (middle and right). Left and middle: an additional α -helix (α -7) is formed by the prolonged C-terminus in LmigOBP1 (Zheng et al., 2015) **(A)**. Instead of a seventh α -helix, the shortened C-terminus in OasiOBP3 **(C)** and SgreOBP6 **(E)** are likely to constitute a groove structure on the collapsed surface. Right: a map of electrostatic potential on the molecular surface. The electrostatic potential pattern of LmigOBP1 **(A)**, OasiOBP3 **(C)** and SgreOBP6 **(E)** varies greatly at a global surface scale as well as on the local C-terminal surface scale. **(B,D,F)** Depiction of the interior cavity (left) which is bordered by C-terminal domain (highlighted in purple) and the corresponding electrostatic potential map (right). The assumed enlarged interior cavity in LmigOBP1 **(B)** relative to OasiOBP3 **(D)** and SgreOBP6 **(F)** is outlined with a white dash line. Electrostatic potential was calculated in the range of -6 to 6 kT/e and was presented as blue-red hue gradient. Blue, negative potential; red, positive potential; k, Boltzmann's constant; T, temperature; e, charge of an electron.

situ hybridization (ISH) on antennal sections using specific OBP probes, we acquired a strikingly sensilla-specific expression pattern for SgreOBPs in the two subfamilies. For SgreOBP1, SgreOBP5 and SgreOBP6, the members of subfamily I-A, we found alike expression in the cells of both sensilla basiconica and sensilla trichodea (Figure 6). In contrast, none of the subfamily I-A SgreOBPs was expressed in sensilla coeloconica or sensilla chaetica. Conversely, for members of subfamily II-A SgreOBPs, namely, SgreOBP10, OBP11, and OBP14, the expression was found to be restricted to the cells of sensilla coeloconica; there was no evidence for an expression in cells of any other sensillum type (Figure 6). The notion that a similar expression pattern is conserved for orthologous OBPs from other locust species is supported by the finding that LmigOBP1 is specifically expressed in sensilla basiconica and sensilla trichodea of *L. migratoria* (Jin et al., 2005), alike its ortholog in *S. gregaria*, the SgreOBP1.

Thus, an apparent sensilla-specific expression pattern for each locust OBP subfamily emerged. To extend and specify this aspect, the expression of OBP subtypes was compared with the expression of sensilla-specific receptor types. The odorant receptor co-receptor Orco and the ionotropic receptor (IR) type IR8a are ubiquitous co-receptors expressed in insect OSNs, either together with ligand-specific ORs or with IRs, and are considered as general markers for sensilla basiconica/sensilla trichodea and sensilla coeloconica, respectively (Yang et al., 2012; Guo

et al., 2013). As a marker specific for distinct sensilla trichodea, the expression of the sensilla-specific receptor type OR3 in *S. gregaria* was monitored (Pregitzer et al., 2017). We designed riboprobes labeled by either Dig or Bio, which specifically targeted the distinct sensory neuron markers and SgreOBPs of the two subfamilies. Subsequently, two-color fluorescent *in situ* hybridization (FISH) experiments were performed to visualize the expressing cells (Figure S5). The results indicated that SgreOBPs of subfamily I-A are expressed in cells located in sensilla basiconica; these cells extended cytoplasmic processes and enclosed clusters of Orco expressing neurons. Similarly, SgreOBPs of subfamily I-A were found to be expressed in cells located in sensilla trichodea, as characterized by their close association with OR3 expressing OSNs. In the sensilla coeloconica, characterized by the IR8a-positive neurons, the neurons were found to be engulfed by cells which express OBPs of the subfamily II-A.

Although, our data demonstrated that SgreOBPs from different ortholog groups in each subfamily are expressed in the same sensillum type, it remained unclear to what extent they are expressed in the same set of sensilla and whether they are co-expressed in the same cells within a distinct sensillum. To resolve this question, we performed two-color FISH on sections through the antenna of *S. gregaria* using riboprobes targeting SgreOBPs from different ortholog groups. The results for SgreOBPs in

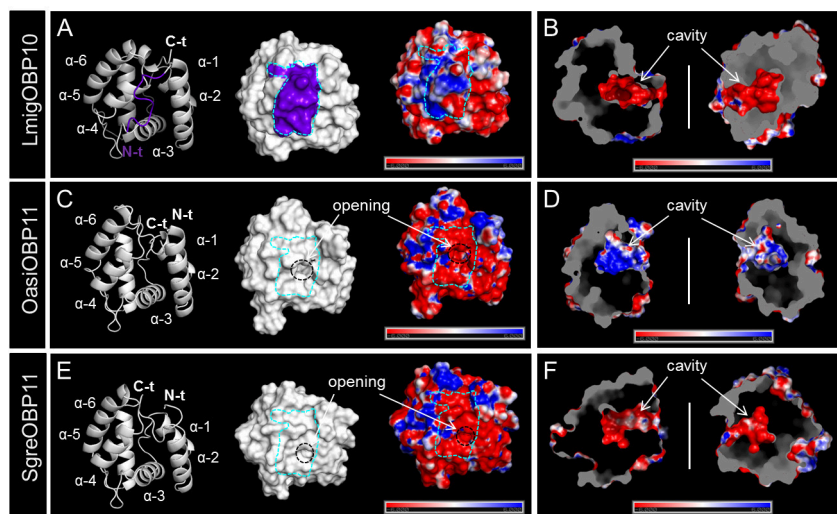


FIGURE 5 | The surface topologies and interior cavities of subfamily II-A OBPs. **(A,C,E)** Comparison of the backbone structures, surface topologies, and surface potentials of LmigOBP10, OasiOBP11, and SgreOBP11 which represent the three different ortholog groups in subfamily II-A. Left: the prolonged N-terminus in LmigOBP10 (see Figure S4B) was predicted to form a short coiled-coil shown on the backbone structure (highlighted in purple, **A**), but was absent from OasiOBP11 **(C)** and SgreOBP11 **(E)**. Middle: the N-terminal domain of LmigOBP10 was plotted on the surface and sketched by a cyan dash line **(A)**. The N-terminal domain of LmigOBP10 was labeled on the same surface position for OasiOBP11 **(C)** and SgreOBP11 **(E)**. The visible opening structure is denoted by a black circle for OasiOBP11 **(C)** and SgreOBP11 **(E)**. Right: a map of electrostatic potential on the molecular surface. Generally similar electrostatic potential pattern is observed among LmigOBP10 **(A)**, OasiOBP11 **(C)** and SgreOBP11 **(E)**. **(B,D,F)** A symmetric presentation of the interior cavity with the electrostatic potential. Electrostatic potential was calculated in the range of -6 to 6 kT/e and was presented as blue-red hue gradient. Blue, negative potential; red, positive potential; k, Boltzmann's constant; T, temperature; e, charge of an electron.

subfamily I-A indicate that SgreOBP1 was expressed in a cell population present in almost all basiconic and trichoid sensilla, whereas SgreOBP5 and SgreOBP6 were expressed only in a much smaller subset of cells than SgreOBP1 in the same sensillum (Figure 7). These differences became apparent in both horizontal sections giving a view onto superficial cellular layer (no cytoplasmic process expected, Figure 7A) as well as in longitudinal sections which allowed a view into deeper layers (cytoplasmic process expected, Figure 7B) of the antenna. Unlike SgreOBP1-positive cells which could be visualized both at the superficial and the deeper cellular layer, most of SgreOBP5- and SgreOBP6-positive cells appeared to be restricted to the superficial cellular layer close to the cuticle; slim cytoplasmic processes stretched to deeper cellular layers. Incidentally, there was evidence that SgreOBP5 and SgreOBP6 were expressed in the same set of cells of a sensillum (Figures 7E,F).

In contrast to the subfamily I-A, for subfamily II-A we did not find any evidence for an OBP subtype that was ubiquitously expressed in coeloconic sensilla (Figure S5). This result has led to the notion that particular OBP members of subfamily II-A may be specifically expressed in subsets of coeloconic sensilla. In fact, we frequently observed that expression of SgreOBP10 and SgreOBP14 were restricted to different cells in sensilla coeloconica (Figures 8A,B). For the subtypes SgreOBP11 and SgreOBP14 a co-expression in the same cells or expression in different cells were observed at a similar rate (Figures 8C,D). For the subtypes SgreOBP10 and SgreOBP11 it was frequently observed that they were co-expressed in the

same cells (Figure 8E), indeed, more often than an expression in different cells (Figure 8F). Moreover, we verified the spatially separated expression of SgreOBPs from subfamily I-A and II-A (Figure S6), consistent with the results in Figures 6, 7. Taken together, the results unravel a characteristic subfamily-dependent cellular expression pattern for different OBP subtypes.

DISCUSSION

The complex behavior of locust species, including the unique switch between a solitary phase and a gregarious phase, is strongly based on a sophisticated chemical communication system (Pener and Yerushalmi, 1998; Hassanali et al., 2005; Wang and Kang, 2014). Great efforts have been made to unravel the chemical cues and underlying chemosensory mechanisms in mediating locust enigmatic behavior (Heifetz et al., 1996; Anton et al., 2007). Out of these efforts, a variety of olfactory genes, including gene families encoding odorant receptors and candidate pheromone receptors have recently been identified (Guo et al., 2011; Wang et al., 2015; Pregitzer et al., 2017). Since much less was known about their counterparts which deliver the olfactory signal molecules to the receptors, the OBPs, this study was concentrating on a systematic analysis of locust OBPs with respect to their molecular evolution as well as on an evaluation of predicted protein structures for OBP subtypes and their expression pattern in distinct sensillum types.

The in-depth analysis of locust OBP sequences uncovered the presence of both common and specific amino acid

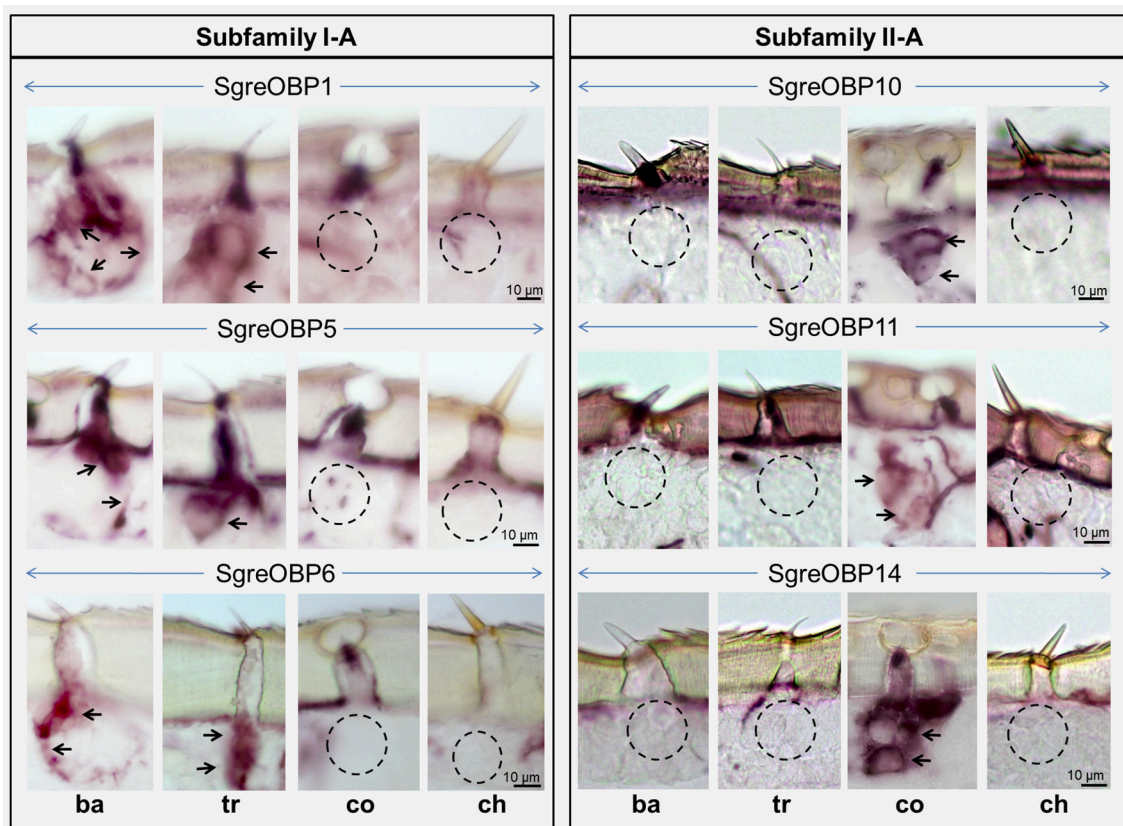


FIGURE 6 | Sensilla-specific expression of subfamily I-A and II-A OBPs in the antenna of *S. gregaria*. Antisense riboprobes which specifically target the SgreOBPs were used to visualize the appropriate structures by means of chromogenic *in situ* hybridization (ISH). SgreOBP1, SgreOBP5, and SgreOBP6 are representing three different ortholog groups of subfamily I-A, whereas SgreOBP10, SgreOBP11, and SgreOBP14 are representing three different ortholog groups of subfamily II-A. Labeling obtained with probes for subfamily I-A SgreOBPs was restricted to sensilla basiconica (ba) and sensilla trichodea (tr), but was absent in sensilla coeloconica (co) and sensilla chaetica (ch). Labeling obtained with probes for subfamily II-A SgreOBPs was detected only in sensilla coeloconica (co), but was absent in the other three sensillum types. Black arrows indicate the visible OBP labeling while black circles denote the absence of OBP labeling.

motifs (Figure 2). The common motifs adequately recapitulate sequence information in most of the locust OBPs, while specific motifs selectively represent locust OBP subfamilies which may contribute to the clustering of sequences on the phylogenetic tree (Figure 1). The mixed common and specific motif profile is reminiscent of the findings that selection regimes may vary among different sequence domains (Policy and Conway, 2001; Sawyer et al., 2005). The subfamily specific motifs define sequence domains that apparently withstand diversifying selection constraints, presumably shaped by the sensilla environment, including their likely interplay-partner, the endogenous receptor types (Figure S5). In contrast, the common motifs define sequence domains that appear to share similar stabilizing selection constraints, presumably required for the maintenance of the common globular structures of the proteins (Pelosi et al., 2017), or for retaining the conserved ligand binding sites (Yu et al., 2009).

The four locust species tackled in this study differ significantly in their geographic distribution. While *S. gregaria* (the desert locust) occurs in Africa, the Middle East and Asia and *L. migratoria* (the migratory locust) in Africa and Asia, but

also in Australia and New Zealand, the locusts *O. asiaticus* and *C. kiangsu* (the yellow-spined bamboo locust) appear to live locally in North China and South China. Nevertheless, a molecular and evolutionary stabilized status can be assigned to locust OBP subfamily I-A and II-A that appear to be subject to purifying selection pressure (Figure 3C), indicative for conserved chemosensory roles. In addition, the chemosensory adaptation to different habitats supposedly implies positive selection constraints (Cicconardi et al., 2017), and several of the locust OBPs appear to reflect such a selection regime (Figure 3C).

For the locust OBP subfamily I-A, the selective expression in two distinct sensillum types, sensilla basiconica, and sensilla trichodea, appears to be a characteristic hallmark (Figure 6 and Figure S5). This feature is also found for OBPs from other species, which are orthologous of locust OBPs subfamily I-A (Figure 3D). For example, in *Drosophila melanogaster*, most of the subfamily I-A orthologous OBPs are associated with sensilla basiconica and sensilla trichodea, similar to their locust counterparts. It was found that DmelOBP83a and DmelOBP83b were associated with sensilla basiconica and sensilla trichodea, while DmelOBP69a and DmelOBP76a seemed to be restricted to sensilla trichodea

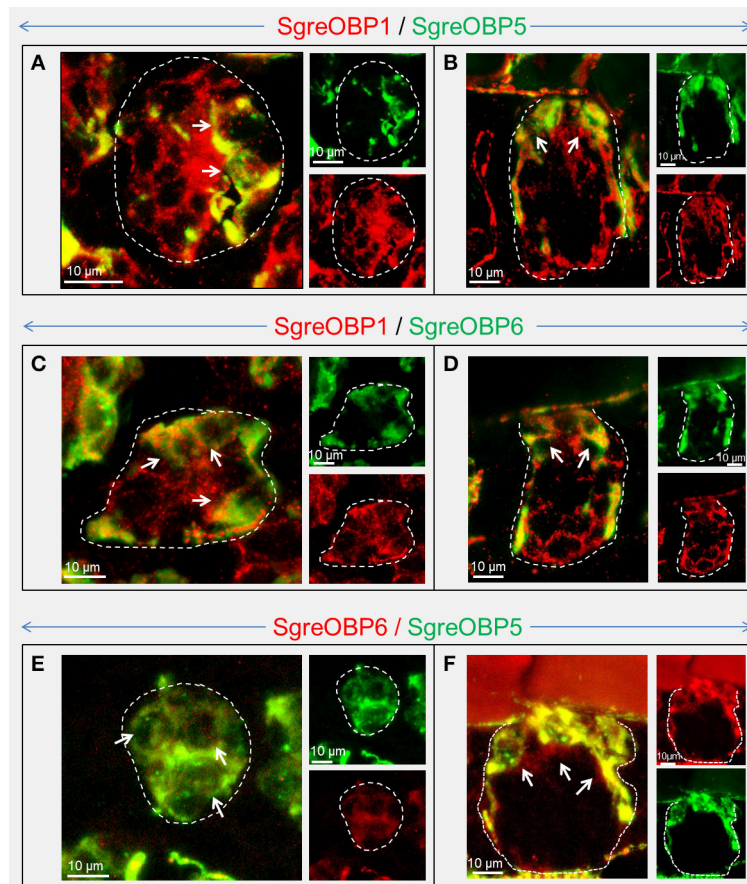


FIGURE 7 | Visualization of cells expressing distinct subtypes of subfamily I-A SgreOBPs. The confocal images show the co-localization of three SgreOBPs from subfamily I-A in the cellular compartment of a sensillum basiconicum. Cells expressing distinct subtypes of subfamily I-A SgreOBPs were visualized by two-color FISH employing subtype specific antisense riboprobes. Confocal images of the overlaid green and red fluorescence channel are shown at higher magnification on the left, the red and green fluorescent channels are shown separately at lower magnification on the right. Cells that are apparently assigned to the cell-cluster belonging to one sensillum basiconicum are outlined in a white dash line. Cells that co-express two distinct OBP subtypes are indicated by white arrows. **(A,C,E)** A horizontal perspective of the superficial cellular layer close to the cuticle is shown where the cytoplasmic processes exhibited by subfamily I-A SgreOBP-positive cells are less likely to be visualized. **(B,D,F)** A longitudinal perspective of a deep layer beneath the cuticle is shown where the cytoplasmic processes are likely to be visualized. **(A–D)** It is noted that a smaller number of cells are labeled in green compared to the number of cells labeled in red.

(Larter et al., 2016). However, for a few orthologous OBPs such as DmelOBP56d an extra sensillar expression has been reported (Larter et al., 2016). The concept of a sensilla-specific expression pattern for orthologous OBPs of locust subfamily I-A is also supported by the finding in the moth *Manduca sexta*, where two orthologous OBPs of locust subfamily I-A, named MsexABP2 and MsexABPx, are specifically expressed in sensilla basiconica (Nardi et al., 2003). Since the Orthopteran locust species emerged at a much earlier stage than the moth and fly species during the insect species divergence (Vieira and Rozas, 2011; Vogt et al., 2015), it is conceivable that a dual expression of subfamily I-A OBPs in both sensilla basiconica and sensilla trichodea may represent an ancestral status. In insect species like moths and flies, which emerged later in evolution, some OBP subtypes may have evolved towards a more specific function and expression in either sensilla basiconica or sensilla trichodea (Maida et al., 2005; Larter et al., 2016).

Our analysis suggests that the locust OBPs of subfamily II-A and their orthologous OBPs in other species have originated from a common ancestor (Figure S3), and may share a sensilla coeloconica specific expression pattern (Figure 6, Figure S5). In *Drosophila melanogaster*, DmelOBP84a, the only orthologous OBP of locust subfamily II-A is actually among the few OBPs that have been reported to be specifically expressed in sensilla coeloconica (Larter et al., 2016). Interestingly, the gene encoding OBP84a is retained in most, if not all, *Drosophila* species genomes (Cicconardi et al., 2017). Moreover, the OBP84a ortholog group in *Drosophila* species withstands apparent purifying selection pressure (Vieira et al., 2007) and converges onto a segregated phylogenetic clade (Cicconardi et al., 2017), which is very similar to the locust OBP subfamily II-A. These molecular and phylogenetic commonalities may point to some similarities with regard to their functional roles. In this regard, it is interesting to note that single sensillum recordings from sensilla coeloconica

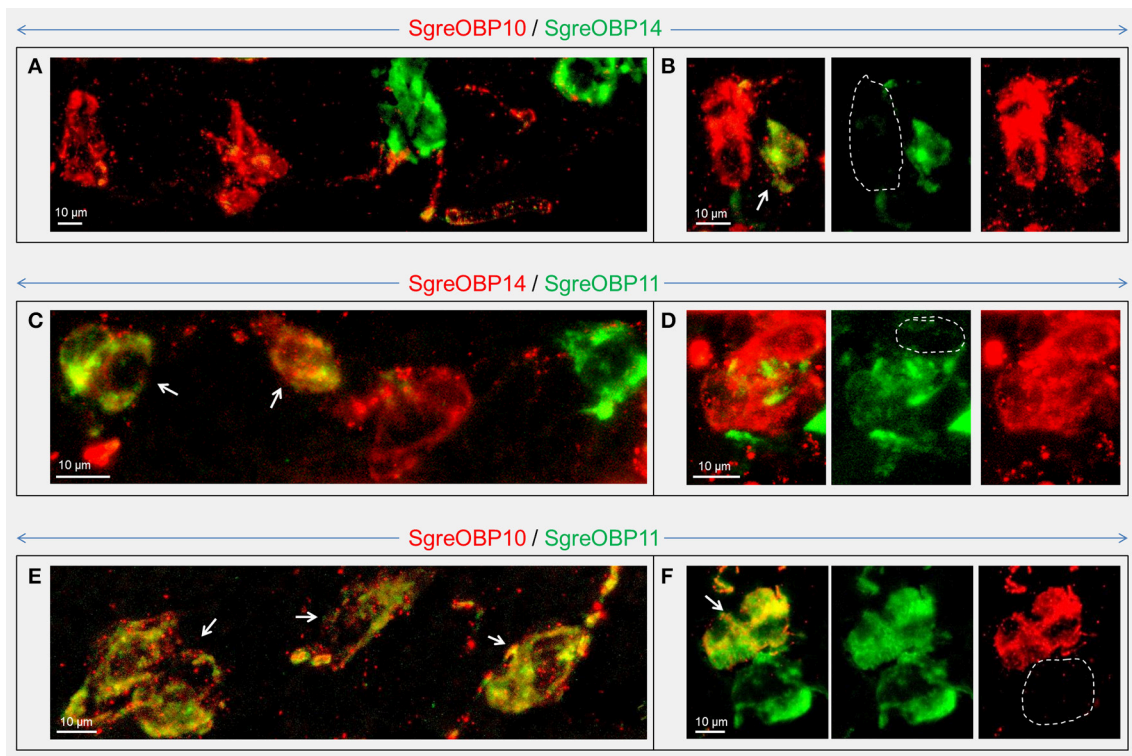


FIGURE 8 | Visualization of cells expressing distinct subtypes of subfamily II-A SgreOBPs. Cells expressing distinct subtypes of subfamily II-A SgreOBPs were visualized by two-color FISH employing combinations of subtype specific antisense riboprobes. The dash line indicates the absence of OBP labeling at the particular area. **(A)** A representative confocal image demonstrating expression of SgreOBP10 and SgreOBP14 in separate cells. **(B)** Rarely cells could be observed that co-expressed SgreOBP10 and SgreOBP14 (white arrows). **(C,D)** For the combination of SgreOBP14 and SgreOBP11 co-expression (white arrows) was observed at a similar rate as a separate expression of the two OBPs in different sensilla. **(E)** SgreOBP10 and SgreOBP11 were frequently found to be co-expressed in the same cells (white arrows). **(F)** Only few cells were detected that selectively expressed only one of the OBP subtypes. **(B,D,F)** Confocal images show the overlaid fluorescence channels (left) as well as the separated green and red fluorescence channels (middle and right) on the same magnification.

of locust, flies and moths have revealed a response spectrum confined to certain ecologically important odorants, including organic acid, amines and plant derived odorants (Pophof, 1997; Ochieng and Hansson, 1999; Yao, 2005). Thus, it will be of particular interest to unravel a potential role of locust subfamily II-A OBPs and their orthologs in other species for the detection of cognate odorants in sensilla coeloconica. While concentrating on OBPs of subfamily II-A, we are aware that sensilla coeloconica may also comprise OBPs of other phylogenetic clades.

Unlike DmelOBP84a, which is broadly expressed in almost all sensilla coeloconica (Larter et al., 2016), the OBPs of the locust subfamily II-A are expressed in sensilla coeloconica in a combinatorial mode (Figure 8). This is in line with the previous finding that different subsets of sensilla coeloconica in *S. gregaria* showed individual response spectra to a repertoire of odorants (Ochieng and Hansson, 1999), suggesting a sensilla-specific response spectrum and sensilla-specific repertoire of odorant sensing proteins. Thus, it is conceivable that a distinct combination of OBPs in a sensillum coeloconicum (Figure 8) may correlate with particular endogenous IR types.

Although amino acid sequences of OBPs can be highly divergent, the folding of proteins forming a hydrophobic pocket

is well conserved across insect species; in fact to date the structures of more than 20 OBPs have been solved by X-ray crystallography and/or nuclear magnetic resonance (NMR) spectroscopy (Pelosi et al., 2017). The results of these studies revealed that the C-terminal domain, especially the length of the C-terminus has important implications on the mechanism of ligand-binding (Tegoni et al., 2004). Long C-terminus apparently enter the binding pocket and determine the shape of the cavity (Sandler et al., 2000), medium-length C-terminus act as a lid covering the entrance to the binding pocket (Lartigue et al., 2004). In view of these findings, simulation of the putative tertiary structures of locust OBPs revealed some interesting features. The three ortholog groups of subfamily I-A significantly differ in their C-terminal domain. LmigOBP1 and its orthologs have a long (17 aa) C-terminus, long enough to form an extra α -helix and thus affecting the shape of the cavity (Figure 4, Figure S4); other two ortholog groups have both a medium size C-terminus (7 aa), however, significantly different in the amino acid sequence. These observations may suggest significant differences in the mechanisms of OBP/ligand interaction among the three ortholog groups in subfamily I-A.

The results of this study indicate that in a considerable number of sensilla at least two OBP subtypes are co-expressed (Figures 7, 8). This is of particular interest, since hetero- and homo-dimerization of OBPs have been reported *in vitro* (Andronopoulou et al., 2006), which is accompanied by a set of conformational changes (Wogulis et al., 2006; Mao et al., 2010). Although the underlying mechanisms are still elusive, there is evidence that electrostatic interaction at short range forming the salt bridges may contribute to specific protein-protein interaction (Sheinerman et al., 2000; Kumar and Nussinov, 2002). In locusts, the patch of charged residues buried on the OBP-interface (Figures 4, 5) is likely to provide hot spots for protein-protein interactions. In addition, changes of the OBP tertiary structure has been demonstrated as a consequence of pH changes in the environment (Zubkov et al., 2005; Pesenti et al., 2008). This notion may also fit for locust OBPs since an intermingled distribution of both negative and positive charged residues was observed by elucidating a map of electrostatic potentials (Figures 4, 5). The presence of multiple OBP subtypes and their possible interaction may have functional implications for the binding capacity of the olfactory system. In fact, recent binding assays have shown that in the presence of two OBPs the binding affinity to cognate ligands altered considerably compared to the binding characteristics of a single OBP type (Qiao et al., 2011; Sun et al., 2016). This notion may be particular relevant with respect to sensilla basiconica of locusts, which house up to 50 sensory neurons responding to a variety of different odorants (Ochieng et al., 1998; Ochieng and Hansson, 1999), and the fact that the number of OBP genes is much smaller than the size of the OR gene family in locusts, encoding more than 140 ORs in *L. migratoria* (Wang et al., 2014) and at least 120 ORs in *S. gregaria* (Pregitzer et al., 2017). The selective sensilla expression pattern implies that a small number of OBP

subtypes are present in the sensillum lymph (Figure 7, Figure S6). Assuming that each OBP subtype has distinct ligand specificity, the mixture may provide a much broader binding spectrum. A possible combinatorial mode of OBP participation in locust olfaction is an interesting aspect for future studies.

AUTHOR CONTRIBUTIONS

HB, JK, XJ, and PP conceived the study. XJ conducted the experiments. HB, JK, XJ, and PP interpreted the results. XJ and PP drafted the preliminary manuscript. HB and JK refined and approved the final manuscript.

FUNDING

The author XJ is supported by a grant from Chinese Scholarship Council (CSC) with the award number 201406350032.

ACKNOWLEDGMENTS

We are grateful to Dr. Ewald Grosse-Wilde (Max Planck Institute for Chemical Ecology) for the help with generating the antennal transcriptome sequence data base and initial bioinformatics analysis. We thank Heidrun Froß for her excellent technical assistance, Rosolino Bumbalo for his suggestions for the data analysis and Christian Heidel for suggestions and comments on the structure modeling.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fphys.2017.00734/full#supplementary-material>

REFERENCES

- Andronopoulou, E., Labropoulou, V., Douris, V., Woods, D. F., Biessmann, H., and Iatrou, K. (2006). Specific interactions among odorant-binding proteins of the African malaria vector *Anopheles gambiae*. *Insect Mol. Biol.* 15, 797–811. doi: 10.1111/j.1365-2583.2006.00685.x
- Anton, S., Dufour, M. C., and Gadenne, C. (2007). Plasticity of olfactory-guided behaviour and its neurobiological basis: lessons from moths and locusts. *Entomol. Exp. Appl.* 123, 1–11. doi: 10.1111/j.1570-7458.2007.00516.x
- Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., et al. (2009). MEME Suite: tools for motif discovery and searching. *Nucleic Acids Res.* 37, W202–208. doi: 10.1093/nar/gkp335
- Ban, L., Scalon, A., D'Ambrosio, C., Zhang, L., Yahn, Y., and Pelosi, P. (2003). Biochemical characterization and bacterial expression of an odorant-binding protein from *Locusta migratoria*. *Cell. Mol. Life Sci.* 60, 390–400. doi: 10.1007/s000180300032
- Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* 17, 540–552. doi: 10.1093/oxfordjournals.molbev.a026334
- Cicconardi, F., Di Marino, D., Olimpieri, P. P., Arthofer, W., Schlick-Steiner, B. C., and Steiner, F. M. (2017). Chemosensory adaptations of the mountain fly *Drosophila nigrosarsa* (Insecta: Diptera) through genomics' and structural biology's lenses. *Sci. Rep.* 7:43770. doi: 10.1038/srep43770
- Damberger, F., Nikonova, L., Horst, R., Peng, G., Leal, W. S., and Wüthrich, K. (2000). NMR characterization of a pH-dependent equilibrium between two folded solution conformations of the pheromone-binding protein from *Bombyx mori*. *Protein Sci.* 9, 1038–1041. doi: 10.1110/ps.9.5.1038
- DeLano, W. (2002). Pymol: an open-source molecular graphics tool. *CCP4 Newsl. Protein Crystallogr.* 700.
- Foret, 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. doi: 10.1101/gr.5075706
- Guo, M., Krieger, J., Große-Wilde, E., Mißbach, C., Zhang, L., and Breer, H. (2013). Variant ionotropic receptors are expressed in olfactory sensory neurons of coeloconic sensilla on the antenna of the desert locust (*Schistocerca gregaria*). *Int. J. Biol. Sci.* 10, 1–14. doi: 10.7150/ijbs.7624
- Guo, W., Wang, X., Ma, Z., Xue, L., Han, J., Yu, D., et al. (2011). CSP and *takeout* genes modulate the switch between attraction and repulsion during behavioral phase change in the migratory locust. *PLoS Genet.* 7:e1001291. doi: 10.1371/journal.pgen.1001291
- Hansson, B. S., and Stensmyr, M. C. (2011). Evolution of insect olfaction. *Neuron* 72, 698–711. doi: 10.1016/j.neuron.2011.11.003
- Hassanali, A., Njagi, P. G. N., and Bashir, M. O. (2005). Chemical ecology of locust and related acridids. *Annu. Rev. Entomol.* 50, 223–245. doi: 10.1146/annurev.ento.50.071803.130345
- Heifetz, Y., Voet, H., and Applebaum, S. W. (1996). Factors affecting behavioral phase transition in the desert locust, *Schistocerca gregaria* (Forskål) (Orthoptera: Acrididae). *J. Chem. Ecol.* 22, 1717–1734. doi: 10.1007/BF02272410

- Hekmat-Safe, D. S., Scape, 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. doi: 10.1101/gr.239402
- Horst, R., Damberger, F., Luginbühl, P., Güntert, P., Peng, G., Nikonova, L., et al. (2001). NMR structure reveals intramolecular regulation mechanism for pheromone binding and release. *Proc. Natl. Acad. Sci. U.S.A.* 98, 14374–14379. doi: 10.1073/pnas.251532998
- Huerta-Cepas, J., Szklarczyk, D., Forslund, K., Cook, H., Heller, D., Walter, M. C., et al. (2016). EGGNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Res.* 44, D286–D293. doi: 10.1093/nar/gkv1248
- Jiang, X., Pregitzer, P., Grosse-Wilde, E., Breer, H., and Krieger, J. (2016). Identification and characterization of two “sensory neuron membrane proteins” (SNMPs) of the desert locust, *Schistocerca gregaria* (Orthoptera: Acrididae). *J. Insect Sci.* 16:33. doi: 10.1093/jisesa/iw015
- Jin, X., Brandazza, A., Navarrini, A., Ban, L., Zhang, S., Steinbrecht, R. A., et al. (2005). Expression and immunolocalisation of odorant-binding and chemosensory proteins in locusts. *Cell. Mol. Life Sci.* 62, 1156–1166. doi: 10.1007/s00018-005-5014-6
- Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010
- Kosakovsky Pond, S. L., Frost, S. D. W., and Muse, S. V. (2005). HyPhy: hypothesis testing using phylogenies. *Bioinformatics* 21, 676–679. doi: 10.1093/bioinformatics/bti079
- Kosakovsky Pond, S. L., Poon, A. F. Y., and Frost, S. D. W. (2009). “Estimating selection pressures on alignments of coding sequences,” in *The Phylogenetic Handbook: A Practical Approach To Phylogenetic Analysis And Hypothesis Testing*, eds P. Lemey, M. Slaemi, and A. M. Vandamme (Cambridge: Cambridge University Press), 419–490.
- Kumar, S., and Nussinov, R. (2002). Close-range electrostatic interactions in proteins. *ChemBiochem.* 3, 604–617. doi: 10.1002/1439-7633(20020703)3:7<604::AID-CBIC604>3.0.CO;2-X
- Lagarde, A., Spinelli, S., Tegoni, M., He, X., Field, L., Zhou, J. J., et al. (2011). The crystal structure of odorant binding protein 7 from *Anopheles gambiae* exhibits an outstanding adaptability of its binding site. *J. Mol. Biol.* 414, 401–412. doi: 10.1016/j.jmb.2011.10.005
- Larter, N. K., Sun, J. S., and Carlson, J. R. (2016). Organization and function of *Drosophila* odorant binding proteins. *Elife* 5:e20242. doi: 10.7554/eLife.20242
- Lartigue, A., Gruez, A., Briand, L., Blon, F., Bézirard, V., Walsh, M., et al. (2004). Sulfur single-wavelength anomalous diffraction crystal structure of a pheromone-binding protein from the honeybee *Apis mellifera* L. *J. Biol. Chem.* 279, 4459–4464. doi: 10.1074/jbc.M311212200
- Leal, W. S. (2013). Odorant reception in insects: roles of receptors, binding proteins, and degrading enzymes. *Annu. Rev. Entomol.* 58, 373–391. doi: 10.1146/annurev-ento-120811-153635
- Leal, W. S., Nikonova, L., and Peng, G. (1999). Disulfide structure of the pheromone binding protein from the silkworm moth, *Bombyx mori*. *FEBS Lett.* 464, 85–90. doi: 10.1016/S0014-5793(99)01683-X
- Maida, R., Mameli, M., Müller, B., Krieger, J., and Steinbrecht, R. A. (2005). The expression pattern of four odorant-binding proteins in male and female silk moths, *Bombyx mori*. *J. Neurocytol.* 34, 149–163. doi: 10.1007/s11068-005-5054-8
- Mao, Y., Xu, X., Xu, W., Ishida, Y., Leal, W. S., Ames, J. B., et al. (2010). Crystal and solution structures of an odorant-binding protein from the southern house mosquito complexed with an oviposition pheromone. *Proc. Natl. Acad. Sci. U.S.A.* 107, 19102–19107. doi: 10.1073/pnas.1012274107
- Miller, M. A., Pfeiffer, W., and Schwartz, T. (2012). “The CIPRES science gateway: enabling high-impact science for phylogenetics researchers with limited resources,” in *Proceedings of the 1st Conference of the Extreme Science and Engineering Discovery Environment: Bridging from the Extreme to the Campus and Beyond* (Chicago), 1–8.
- Nardi, J. B., Miller, L. A., Walden, K. K. O., Rovelstad, S., Wang, L., Frye, J. C., et al. (2003). Expression patterns of odorant-binding proteins in antennae of the moth *Manduca sexta*. *Cell Tissue Res.* 313, 321–333. doi: 10.1007/s00441-003-0766-5
- Niu, D. J., Liu, Y., Dong, X. T., and Dong, S. L. (2016). Transcriptome based identification and tissue expression profiles of chemosensory genes in *Blattella germanica* (Blattaria: Blattellidae). *Comp. Biochem. Physiol. D Genomics Proteomics* 18, 30–43. doi: 10.1016/j.cbd.2016.03.002
- Ochieng, S. A., Hallberg, E., and Hansson, B. S. (1998). Fine structure and distribution of antennal sensilla of the desert locust, *Schistocerca gregaria* (Orthoptera: Acrididae). *Cell Tissue Res.* 291, 525–536. doi: 10.1007/s004410051022
- Ochieng, S. A., and Hansson, B. S. (1999). Responses of olfactory receptor neurons to behaviourally important odours in gregarious and solitary desert locust, *Schistocerca gregaria*. *Physiol. Entomol.* 24, 28–36. doi: 10.1046/j.1365-3032.1999.00107.x
- Pelosi, P., Iovinella, I., Zhu, J., Wang, G., and Dani, F. R. (2017). Beyond chemoreception: diverse tasks of soluble olfactory proteins in insects. *Biol. Rev. Camb. Philos. Soc.* doi: 10.1111/brv.12339. [Epub ahead of print].
- 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. doi: 10.1007/s00018-005-5607-0
- Pener, M. P., and Yerushalmi, Y. (1998). The physiology of locust phase polymorphism: an update. *J. Insect Physiol.* 44, 365–377. doi: 10.1016/S0022-1910(97)00169-8
- Pesenti, M. E., Spinelli, S., Bezirard, V., Briand, L., Pernellet, J. C., Tegoni, M., et al. (2008). Structural basis of the honey bee PBP pheromone and pH-induced conformational change. *J. Mol. Biol.* 380, 158–169. doi: 10.1016/j.jmb.2008.04.048
- 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. doi: 10.1038/nmeth.1701
- Policy, S. D., and Conway, D. J. (2001). Strong diversifying selection on domains of the *Plasmodium falciparum* apical membrane antigen 1 gene. *Genetics* 158, 1505–1512.
- Pophof, B. (1997). Olfactory responses recorded from sensilla coeloconica of the silkworm *Bombyx mori*. *Physiol. Entomol.* 22, 239–248. doi: 10.1111/j.1365-3032.1997.tb01164.x
- Pregitzer, P., Jiang, X., Grosse-Wilde, E., Breer, H., Krieger, J., and Fleischer, J. (2017). In search for pheromone receptors: certain members of the odorant receptor family in the desert locust *Schistocerca gregaria* (Orthoptera: Acrididae) are co-expressed with SNMP1. *Int. J. Biol. Sci.* 13, 911–922. doi: 10.7150/ijbs.18402
- Qiao, H., He, X., Schymura, D., Ban, L., Field, L., Dani, F. R., et al. (2011). Cooperative interactions between odorant-binding proteins of *Anopheles gambiae*. *Cell. Mol. Life Sci.* 68, 1799–1813. doi: 10.1007/s00018-010-0539-8
- Roy, A., Kucukural, A., and Zhang, Y. (2010). I-TASSER: a unified platform for automated protein structure and function prediction. *Nat. Protoc.* 5, 725–738. doi: 10.1038/nprot.2010.5
- Sandler, B. H., Nikonova, L., Leal, W. S., and Clardy, J. (2000). Sexual attraction in the silkworm moth: structure of the pheromone-binding-protein-bombykol complex. *Chem. Biol.* 7, 143–151. doi: 10.1016/S1074-5521(00)00078-8
- Sawyer, S. L., Wu, L. I., Emerman, M., and Malik, H. S. (2005). Positive selection of primate *TRIM5α* identifies a critical species-specific retroviral restriction domain. *Proc. Natl. Acad. Sci. U.S.A.* 102, 2832–2837. doi: 10.1073/pnas.0409853102
- Scaloni, A., Monti, M., Angeli, S., and Pelosi, P. (1999). Structural analysis and disulfide-bridge pairing of two odorant-binding proteins from *Bombyx mori*. *Biochem. Biophys. Res. Commun.* 266, 386–391. doi: 10.1006/bbrc.1999.1791
- Sheinerman, F. B., Norel, R., and Honig, B. (2000). Electrostatic aspects of protein-protein interactions. *Curr. Opin. Struct. Biol.* 10, 153–159. doi: 10.1016/S0959-440X(00)00065-8
- Stamatakis, A. (2014). RAXML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313. doi: 10.1093/bioinformatics/btu033
- Steinbrecht, R. A. (1996). Structure and function of insect olfactory sensilla. *Ciba Found. Symp.* 200, 158–174.
- Suh, E., Bohbot, J. D., and Zwiebel, L. J. (2014). Peripheral olfactory signaling in insects. *Curr. Opin. Insect Sci.* 6, 86–92. doi: 10.1016/j.cois.2014.10.006
- Sun, X., Zeng, F. F., Yan, M. J., Zhang, A., Lu, Z. X., and Wang, M. Q. (2016). Interactions of two odorant-binding proteins influence insect chemoreception. *Insect Mol. Biol.* 25, 712–723. doi: 10.1111/imb.12256

- Tamura, K., Stecher, G., Peterson, D., Filipiński, A., and Kumar, S. (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* 30, 2725–2729. doi: 10.1093/molbev/mst197
- Tegoni, M., Campanacci, V., and Cambillau, C. (2004). Structural aspects of sexual attraction and chemical communication in insects. *Trends Biochem. Sci.* 29, 257–264. doi: 10.1016/j.tibs.2004.03.003
- Terrapon, N., Li, C., Robertson, H. M., Ji, L., Meng, X., Booth, W., et al. (2014). Molecular traces of alternative social organization in a termite genome. *Nat. Commun.* 5:3636. doi: 10.1038/ncomms4636
- Theißen, G. (2002). Orthology: secret life of genes. *Nature* 415:741. doi: 10.1038/415741a
- Unni, S., Huang, Y., Hanson, R. M., Tobias, M., Krishnan, S., Li, W. W., et al. (2011). Web servers and services for electrostatics calculations with APBS and PDB2PQR. *J. Comput. Chem.* 32, 1488–1491. doi: 10.1002/jcc.21720
- 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. doi: 10.1093/gbe/evr033
- Vieira, F. G., Sánchez-Gracia, A., and Rozas, J. (2007). Comparative genomic analysis of the odorant-binding protein family in 12 *Drosophila* genomes: purifying selection and birth-and-death evolution. *Genome Biol.* 8:R235. doi: 10.1186/gb-2007-8-11-r235
- Vogt, R. G., Callahan, F. E., Rogers, M. E., and Dickens, J. C. (1999). Odorant binding protein diversity and distribution among the insect orders, as indicated by LAP, an OBP-related protein of the true bug *Lygus lineolaris* (Hemiptera, heteroptera). *Chem. Senses* 24, 481–495. doi: 10.1093/chemse/24.5.481
- Vogt, R. G., Große-Wilde, E., and Zhou, J. J. (2015). The Lepidoptera odorant binding protein gene family: gene gain and loss within the GOBP/PBP complex of moths and butterflies. *Insect Biochem. Mol. Biol.* 62, 142–153. doi: 10.1016/j.ibmb.2015.03.003
- Vogt, R. G., Rogers, M. E., Franco, M., and Sun, M. (2002). A comparative study of odorant binding protein genes: differential expression of the PBP1-GOBP2 gene cluster in *Manduca sexta* (Lepidoptera) and the organization of OBP genes in *Drosophila melanogaster* (Diptera). *J. Exp. Biol.* 205, 719–744.
- Wang, X., Fang, X., Yang, P., Jiang, X., Jiang, F., Zhao, D., et al. (2014). The locust genome provides insight into swarm formation and long-distance flight. *Nat. Commun.* 5:2957. doi: 10.1038/ncomms3957
- Wang, X., and Kang, L. (2014). Molecular mechanisms of phase change in Locusts. *Annu. Rev. Entomol.* 59, 225–244. doi: 10.1146/annurev-ento-011613-162019
- Wang, Z., Yang, P., Chen, D., Jiang, F., Li, Y., Wang, X., et al. (2015). Identification and functional analysis of olfactory receptor family reveal unusual characteristics of the olfactory system in the migratory locust. *Cell. Mol. Life Sci.* 72, 4429–4443. doi: 10.1007/s00018-015-2009-9
- Wheeler, W. C., Whiting, M., Wheeler, Q. D., and Carpenter, J. M. (2001). The phylogeny of the extant hexapod orders. *Cladistics* 17, 113–169. doi: 10.1111/j.1096-0031.2001.tb00115.x
- Wogulis, M., Morgan, T., Ishida, Y., Leal, W. S., and Wilson, D. K. (2006). The crystal structure of an odorant binding protein from *Anopheles gambiae*: evidence for a common ligand release mechanism. *Biochem. Biophys. Res. Commun.* 339, 157–164. doi: 10.1016/j.bbrc.2005.10.191
- 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. doi: 10.1046/j.1365-2583.2003.00440.x
- 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. doi: 10.1186/1471-2164-10-632
- Yang, Y., Krieger, J., Zhang, L., and Breer, H. (2012). The olfactory co-receptor Orco from the migratory locust (*Locusta migratoria*) and the desert locust (*Schistocerca gregaria*): identification and expression pattern. *Int. J. Biol. Sci.* 8, 159–170. doi: 10.7150/ijbs.8.159
- Yao, C. A. (2005). Chemosensory coding by neurons in the coeloconic sensilla of the *Drosophila* antenna. *J. Neurosci.* 25, 8359–8367. doi: 10.1523/JNEUROSCI.2432-05.2005
- Yu, F., Zhang, S., Zhang, L., and Pelosi, P. (2009). Intriguing similarities between two novel odorant-binding proteins of locusts. *Biochem. Biophys. Res. Commun.* 385, 369–374. doi: 10.1016/j.bbrc.2009.05.074
- Zhang, S., Pang, B., and Zhang, L. (2015). Novel odorant-binding proteins and their expression patterns in grasshopper, *Oedaleus asiaticus*. *Biochem. Biophys. Res. Commun.* 460, 274–280. doi: 10.1016/j.bbrc.2015.03.024
- Zheng, J., Li, J., Han, L., Wang, Y., Wu, W., Qi, X., et al. (2015). Crystal structure of the *Locusta migratoria* odorant binding protein. *Biochem. Biophys. Res. Commun.* 456, 737–742. doi: 10.1016/j.bbrc.2014.12.048
- Zhou, J. J., Huang, W., 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. doi: 10.1016/j.gene.2003.11.007
- Zhou, J. J., Vieira, F. G., He, X. L., Smadja, C., Liu, R., Rozas, J., et al. (2010). Genome annotation and comparative analyses of the odorant-binding proteins and chemosensory proteins in the pea aphid *Acyrtosiphon pisum*. *Insect Mol. Biol.* 19, 113–122. doi: 10.1111/j.1365-2583.2009.00919.x
- Zubkov, S., Gronenborn, A. M., Byeon, I. J. L., and Mohanty, S. (2005). Structural consequences of the pH-induced conformational switch in A. polyphemus pheromone-binding protein: mechanisms of ligand release. *J. Mol. Biol.* 354, 1081–1090. doi: 10.1016/j.jmb.2005.10.015

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Jiang, Krieger, Breer and Pregitzer. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.