Genome-wide survey of odorant-binding proteins in the dwarf honey bee *Apis florea*

Bhavika Mam [1,2,] Snehal Karpe[1#] and Ramanathan Sowdhamini [1]*

[1] National Centre for Biological Sciences, Tata Institute for Fundamental Research, GKVK Campus, Bangalore, Karnataka, India; bhavikam@ncbs.res.in

[2] The University of Trans-Disciplinary Health Sciences and Technology (TDU), Bangalore, Karnataka, India;

[#] Current address: Laboratory of Experimental Hematology, Institut Jules Bordet, Université Libre de Bruxelles, Brussels, Belgium; Unit of Animal Genomics, GIGA, University of Liège, Liège, Belgium

***Correspondence: mini@ncbs.res.in

Abstract

Odorant binding proteins (OBPs) in insects bind to volatile chemical cue and help in their binding to odorant receptors. The odor coding hypothesis states that OBPs may bind with specificity to certain volatiles and aid the insect in various behaviours. Honeybees are eusocial insects with complex behaviour that requires olfactory inputs. Here, we have identified and annotated odorant binding proteins from the genome of the dwarf honey bee, ***Apis florea*** using an exhaustive homology-based bioinformatic pipeline and analyzed the evolutionary relationships between the OBP subfamilies. Our study suggests that Minus-C subfamily may have diverged from the Classic subfamily of odorant binding proteins in insects.

1. Introduction

Insects are a diverse class of Arthropods with a highly sensitive olfactory system**.** Olfactory information helps in mate selection, oviposition while mating, foraging for food and social behaviour (Hildebrand and Shepherd, 1997). Odorant binding proteins (OBPs) abound in the sensillar lymph of insects and in the nasal mucus of many animal species with the presence of at least 50 *OBP* genes reported in some species (Hekmat-Scafe *et al*, 2002). Despite their abundance and diversity, the role of OBPs in olfactory coding is yet to be completely explored (Larter *et al*, 2016).

OBP proteins are small, soluble globular proteins, 10-30 kDa, that are further characterized by alpha-richness, and the presence of six highly conserved cysteine residues (C1-C6) with conserved disulphide spacing (Vogt et al, 1985; Pelosi and Maida, 1990) that stabilizes its tertiary structure. It has been hypothesized that OBPs bind to ligands and solubilize them to aid transport and delivery towards odorant receptors.

Genome-wide surveys to identify odorant-binding proteins in insect orders have been previously performed for various insect species in existing literature. Previous studies have predicted the presence of odorant binding proteins in various species including *Apis mellifera* (Order: Hymenoptera) (Forêt and Maleszka, 2006), *Drosophila melanogaster* (Order: Diptera) (Hekmat and Scafe, 2002; Graham and Davies, 2002), *Anopheles gambiae* (Order: Diptera) (Manoharan *et al*, 2013), *Periplaneta americana* (Order: Blattodea) (He *et al*, 2017) using homology-based bioinformatic approaches as a typical start-point.

Previous work in our laboratory (Karpe *et al*, 2016), has identified odorant receptors (ORs) in *Apis florea* using an exhaustive genomic pipeline. In order to complement the search of ORs

towards a better understanding of odor coding, this study investigated odorant binding proteins (OBPs) in *Apis florea*.

Apis florea or the red dwarf honey bee exhibits the complex behavior of eusociality, where there is reproductive division of labour within a colony that comprises a female queen, male drones and female worker bees. While worker bees perform important tasks such as foraging, guarding the colony hive, maintenance and other diverse tasks for the colony, the queen and drone perform reproductive roles (Page and Robinson, 1991).

Members of the species exhibit haplodiploidy (Halling *et al*, 2001) system of genetic inheritance, where the males in this species are haploid, possessing half the number of chromosomes as diploid females. *Apis florea* is geographically distributed with a preference for warm climate (Otis, 1991) in regions such as mainland Asia, southern border of the Himalayas, plateau of Iran, Oman and in Vietnam, southeast China and peninsular Malaysia (Hepburn *et al*, 2005; Oldroyd and Nanork, 2009; Moritz et al, 2010) and display open nesting typically on low-lying tree branches in shaded regions (Wongsiri *et al*, 1997; Hepburn *et al*, 2005). *Apis florea* are important pollinators of tropical and ornamental plants as well as agricultural crops. They primarily feed on pollen and nectar from flowering plants. Like other honey bees, the body of *Apis florea* is studded with various types of sensilla among which olfactory sensilla (sensilla basiconica and sensilla chaetica) are prominent structures (Gupta, 1992). The antenna of the insect is typically the main site for olfactory receptors (Wigglesworth, 1965). The antennae of *Apis florea* harbor hair-like sensillae trichodea types I, II, III, IV, sensilla basiconica, sensilla placodea and sensilla ampullaceal (Gupta, 1992; Kumar *et al*, 2014).

Insect OBPs, although highly divergent, are classified on the basis of conserved cysteine signature into Classic (six cysteines), Minus-C (loss of two conserved cysteines), Plus-C (additional cysteine residues and one proline) (Zhou et al., 2004) and Atypical (~ 10 cysteines and long C-terminus) (Hekmat-Scafe et al. 2002; Xu et al. 2003) and Dimer OBPs (two cysteine signatures). Rapid identification of repertoires of putative OBPs across various insect genomes has been suggestive of the idea that the ecological niche of an insect species may correlate with abundance of OBPs and social behaviour (Zhou et al, 2020). While reference Dipteran fruitfly, *Drosophila melanogaster* and Japanese encephalitis vector *Culex quinquefasciatus* have been found to have 51 and 110 putative OBPs respectively (Hekmat-Scafe et al., 2002; Manoharan et al, 2013), previous studies in Hymenopteran OBPs have also found species-specific differences in OBPs including 21 OBPs in eusocial *Apis mellifera* (Foret and Maleszka, 2006), 7 OBPs in fig wasp *Ceratosolen solmsi* (Wang et al, 2014) that living in closed spaces and 90 in *P. xylostella* (Vieira et al, 2012) that lives in open spaces. Using *Apis mellifera* as a closely related reference genome and a revised annotation of its OBPs, we thus investigated the identification, annotation and subfamily-based classification of putative OBPs from the genome of *Apis florea* and examined their evolutionary relationships using *in silico* approaches.

2. Materials and Methods

2.1 Obtaining genome of honeybee *Apis florea*

Aflo_1.1 genome was obtained from the National Center for Biotechnology Information (NCBI) (https://www.ncbi.nlm.nih.gov).

2.2 Preparing query dataset from *Apis mellifera*

AmelOBPs were pooled from the NCBI non-redundant protein database (29 putative AmelOBPs) and a previous study (Foret and Malesczka, 2007; 21 AmelOBPs) to obtain a filtered set of query

protein sequences. Reciprocal homology was performed using the query set obtained and AmelOBPs from a recent study (Vieira *et al*, 2011; 21 AmelOBPs). An e-value cutoff of e^-10 was used. The resultant matches as well as unmatched OBPs (putative OBPs with no reciprocal hit; 10 protein sequences) resulted in a final dataset of annotated and unique AmelOBPs.

2.3 Query protein to subject genome alignments

Genomic alignments were obtained using Exonerate (Slater & Birney, 2005) with intron sizes of 500, 2000, 5000 and 10000 respectively with BLOSUM62 (Henikoff and Henikoff, 1992) as the substitution matrix.

The genomic alignments were processed as per the methodology in previous in-house study from lab (Karpe *et al*, 2016; 2017; 2020). The pipeline involves thoroughly scanning and scoring alignments to the genome based on length, degree of similarity and the best match of the scaffold location in the subject genome to the query sequence. The unique set of genomic alignments was then processed further to translate amino acids from corresponding in-frame codons. The resultant set of gene models and protein sequences were also manually corrected for missing start and stop codons, missing N-terminal and C-terminal amino acids and annotated as Complete, Partial or pseudogene.

2.4 Homology-based validation & nomenclature

The predicted *Apis florea* OBPs (AfloOBP) were subjected to reciprocal homology with our manually curated AmelOBP dataset, as explained above. The final dataset of predicted AfloOBPs comprised resultant matches as well as unique sequences with no corresponding hits found in the AmelOBP dataset. The AfloOBP predicted protein sequence dataset was thus annotated with respect to AmelOBP homolog, if present as well as its status as '**Complete**' or '**Partial**'.

2.5 Secondary structure prediction

Secondary structure of the protein sequences were predicted using neural network-based PSIPRED v3.2 (Conesa *et al*, 2005; Buchan *et al*, 2013).

2.6 Signal peptide detection

N-terminal signal peptide was detected using SignalP4.1 (Nielsen et al, 1997; Petersen et al, 2011). This algorithm uses neural networks and Hidden Markov Models to determine signal peptides in a given protein sequence. The predicted signal peptide for a given sequence was cleaved off and the "mature" sequence was used for multiple sequence alignment and phylogeny.

2.7 Preparing dataset of insect OBPs for rooted and unrooted phylogeny

In order to prepare an outgroup for the rooted phylogeny, annotated chemosensory proteins of ***Apis mellifera*** (AmelCSPs) were obtained from a previous study (Forêt, Wanner and Maleszka, 2015), namely, AmelCSP1, AmelCSP2, AmelCSP3, AmelCSP4, AmelCSP5 and AmelCSP6.

In order to construct the phylogeny (Vogt, Große-Wilde and Zhou, 2015; Missbach, Vogel, Hansson and Große-Wilde, 2015), protein sequences of OBPs from 11 insect orders from representative insect species were obtained from previous literature and UniProt (The UniProt Consortium, 2019) database. The insect orders, corresponding species and the number of species-specific OBPs have been tabulated as in **Table 1.**

2.8 Structure-based sequence alignment and phylogenetic analysis

A structure-based seed template was obtained from the PASS2.5 database (Gandhimathi *et al*, 2012) with the SCOP ID of the fold as 47565. The dataset of "mature" AfloOBP sequences was aligned against the seed template using MAFFT (Katoh *et al* 2002; 2013). A phylogenetic tree was

constructed using RaxML (Stamatakis *et al* 2006; 2014) with the maximum likelihood method with 100 bootstraps and the WAG evolutionary model (Whelan and Goldman, 2001). The phylogenetic tree was visualized and annotated using iTOL (Letunic and Bork, 2006; 2016).

3. Results and Discussion

We filtered and re-annotated OBPs from closely related reference genome *Apis mellifera* using a homology-based approach (See Methods). The final dataset (SI Table 1) comprised 25 AmelOBP protein sequences.

Genome-wide survey of ***Apis florea*** revealed 22 novel OBP protein sequences with 15 complete and 7 partial sequences either towards the N-terminus, C-terminus, or both with an average exon number of 5 (**Table 2; SI_Table2**). Secondary structure analysis revealed alpha-rich state of OBPs with high confidence. Typically, 6-7 alpha helices per complete AfloOBP sequence was predicted.

Out of 15 ***AfloOBP*** genes predicted as complete, manually corrected and annotated, 16 translated protein sequences were predicted to have signal peptide sequences. The average length of signal peptides predicted in our AfloOBP dataset was 19 amino acids. Cleavage position ranged from 16th to 24th amino acids in the sequence.

Sequences AfloOBP1-AfloOBP13 were found to display the conserved Cysteine signature of Classic and Minus-C subfamilies, as their orthologs in ***Apis mellifera***, and comparable to that of AgamOBP (**Figure 1)**. Multiple sequence alignment revealed conserved Cysteine profiles specific to Classic and MinusC subfamilies in the ***Apis florea*** genome (**Figure 2**). Sequences AfloOBP14-AfloOBP21 were found to show the conserved Minus-C cysteine signature where Cysteine residues in the conserved second and fifth positions are missing. Our analysis shows that the conserved cysteine signature for both subfamilies in ***Apis florea*** is similar to the representative

signature observed in a previous study (Xu *et al*, 2009). The conserved cysteine signature for Classic subfamily for the Hymenopteran insect order was determined as C1-X **23:35**-C2-X**3**-C3-X 27:45-C4-X 7:14-C5-X**8**-C6 (Xu *et al*, 2009). Our study has identified 13 Classic and 9 Minus-C OBPs in ***Apis florea***. We observe the Classic cysteine signature to be conserved similarly as C1-X **27:37**-C2-X**3:4**-C3-X 33:43-C4-X 9:13-C5-X**8-9**-C6.

Phylogenetic inference revealed the clustering of Minus-C OBPs as a sub-clade of the Classic OBP subfamily comprising members of both ***Apis mellifera*** and ***Apis florea*** OBPs (**Figure 3A**). Moreover, conserved cysteine signature specific to the chemosensory protein (CSP) family was observed in the outgroup chemosensory proteins (AmelCSP) (**Figure 2**). AmelCSPs used as outgroup clustered distinctly (**Figure 3A**) from the odorant binding proteins input to the phylogeny with 100% bootstrap value. Minus-C OBPs were found to cluster together with 60% bootstrap value closest to AfloOBP9, annotated as a Classic OBP. OBPs of the Minus-C subfamily, AfloOBP 14-20 emerge closest to AfloOBP13, a Classic OBP with an observed six cysteine signature. Interestingly, all the other Classic OBPs cluster distinctly in a clade corresponding to the insect Classic subfamily, however, AfloOBP13 clusters closely with the Minus-C group in a distinct sub-clade suggesting an evolutionary ancestral link (**Figure 3B**). Interestingly, antennal OBP (MsexABP1) from Lepidopteran insect *Manduca sexta* clustered close to the Minus-C clade along with other bee species (Hymenopteran) with high bootstrap support of 97%. It is also observed that Classic OBPs in ***Apis florea*** are phylogenetically distant from Minus-C (bee OBPs) than clades representing Atypical OBPs in Dipterans and Plus-C insect OBPs. This suggests that Minus-C OBPs in honey bees may have evolved from a single ancestral Classic OBP (similar to AfloOBP13, AmelOBP13) of its species by deletion of second and fifth cysteines. The evolution and insect order-specific occurrence of Minus-C, Plus-C and Atypical subfamilies of insect OBPs may have functional roles and would be interesting to investigate.

Taken together our observations from a comprehensive bioinformatic analysis strongly suggest that Minus-C OBPs are likely to have evolved from a Classic OBP subfamily member in insects. It is possible that the evolution of a subfamily could be an adaptation to the local niche of the insect species for functional specificity (Zhou *et al*, 2020).

Conclusion

In a step towards understanding the role of OBPs in insects, a bioinformatics-based approach was used as the starting point here. A total of 22 OBPs **including isoforms** have been identified and annotated from the genome of eusocial Asian red dwarf honeybee, *Apis florea* using a modified in-house pipeline**.** Our results include AfloOBPs that have been previously identified by the automated pipeline of NCBI with a query coverage and identity of 100% each (AfloOBP9 and AfloOBP11) (SI_Table3). Our annotated data includes complete OBPs that were identified as having incomplete exons in N-termini and C-termini or/and labelled as uncharacterized by the automated pipeline of NCBI. We also observe that number of *OBP* genes in *Apis florea* (22) and the western honeybee, *Apis mellifera* (25) are similar despite the differences in respective ecological niche.

We have analyzed the characteristic conserved features of these OBPs using computational methods and phylogeny resulting in discovery of new gene models as well as improvement on existing gene models from NCBI. Presence of conserved cysteine pattern, disulphide spacing, domain analysis, size and predicted secondary structure further strengthen their identity as putative insect OBPs. Moreover, the use of structurally- guided multiple sequence alignment for phylogenetic inference has been suitable

The Classic OBP subfamily clade appears to have expanded to Minus-C OBPs in honeybee and few other insect orders suggesting that Minus-C may have evolved from the Classic subfamily

through strongly conserved deletions in positions corresponding to the two missing cysteines in a Minus-C OBP.

Tables and Figures

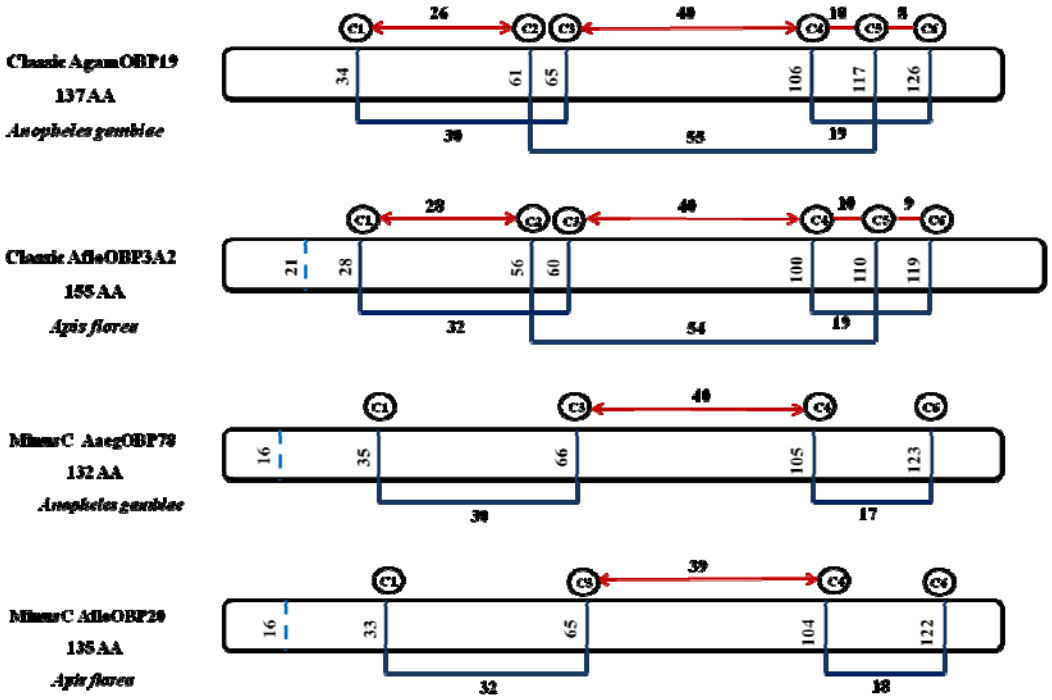


Figure 1: Cysteine signatures across insect orders are comparable. Cysteine residue positioning and inter-disulphide spacing is conserved in ***Apis florea*** (Hymenoptera) genome across Classic and Minus-C subfamilies. The cysteine signature of OBPs from ***Anopheles gambiae*** is given for reference.

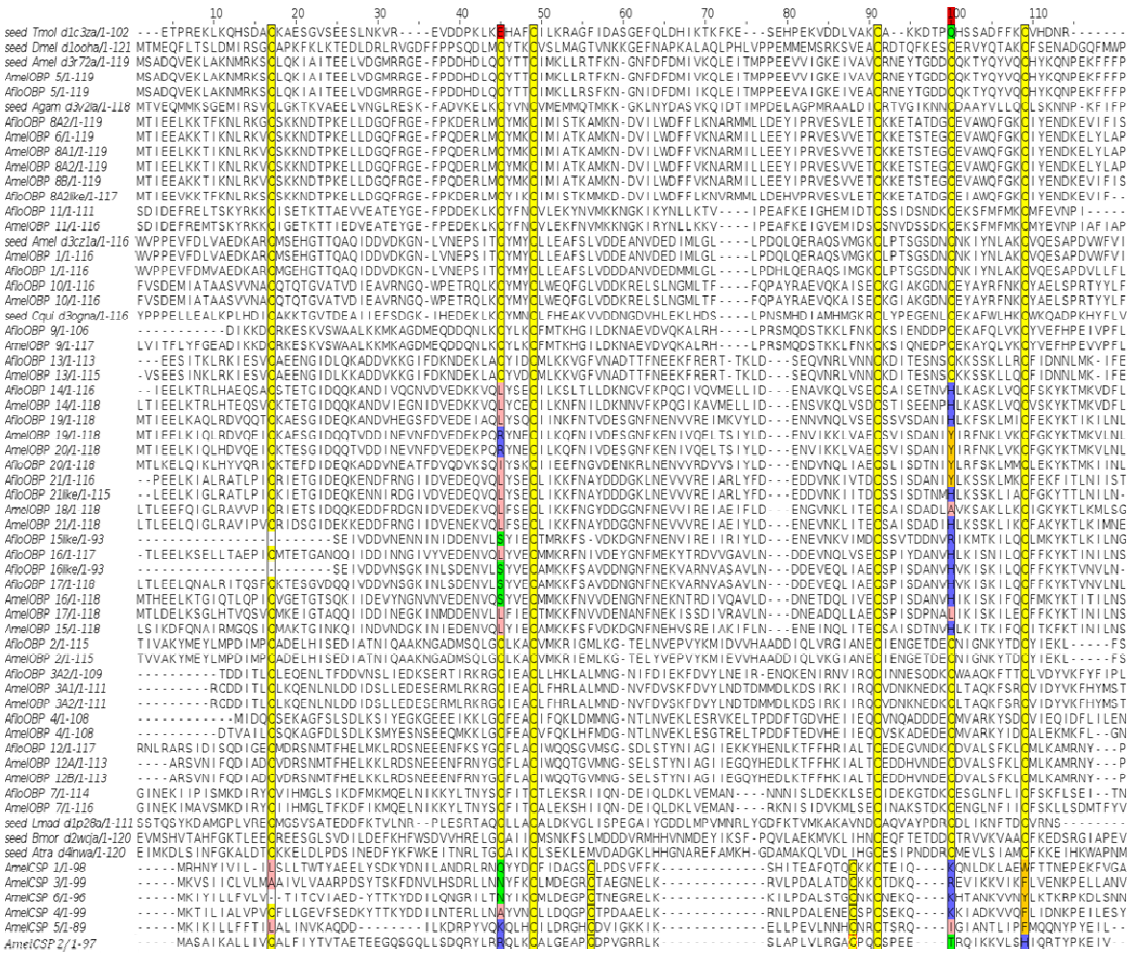
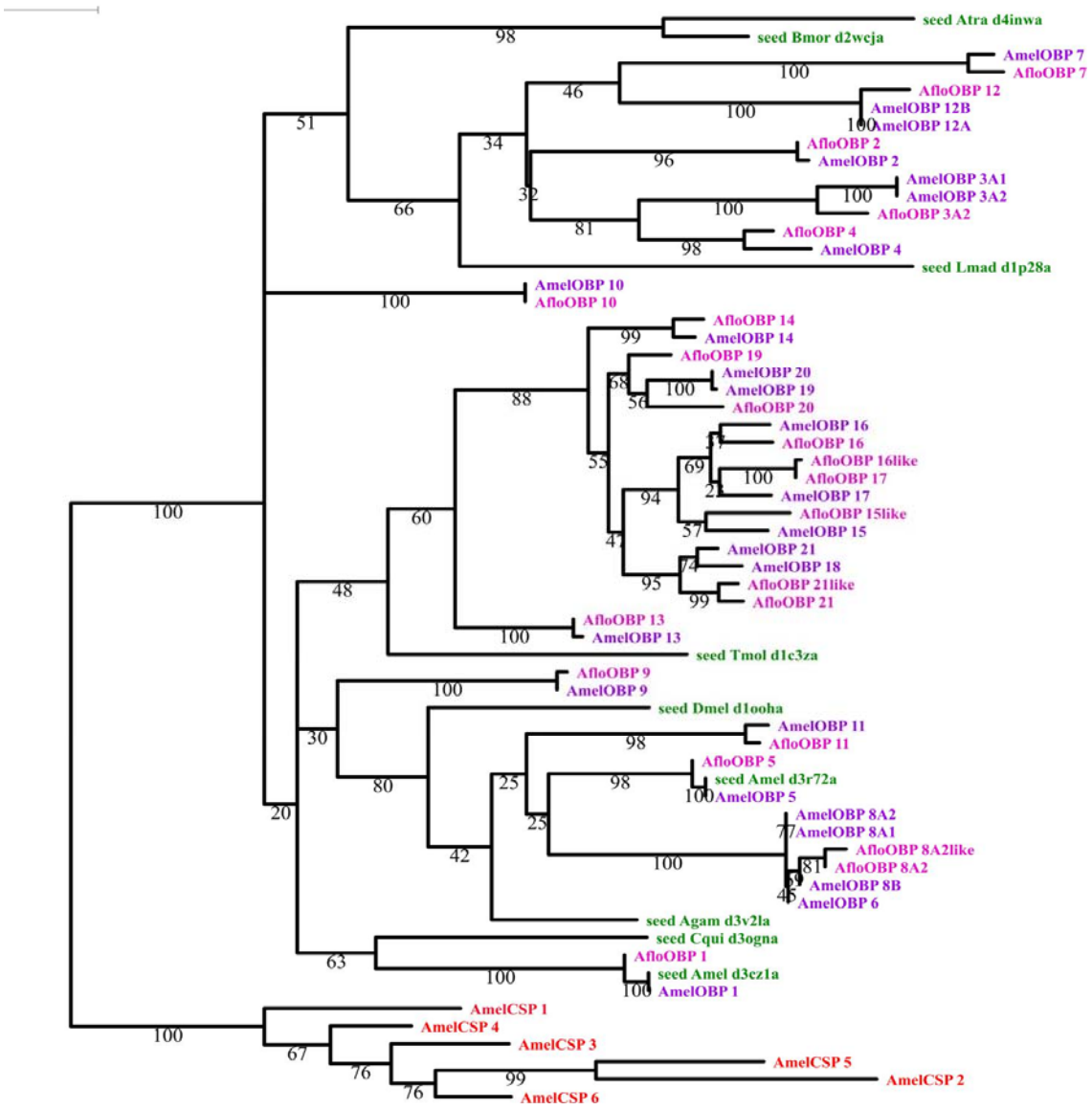


Figure 2: Multiple sequence alignment of OBPs from ***Apis florea***. AfloOBPs. The alignment also contains OBPs from its phylogenetic neighbour ***Apis mellifera*** (AmelOBP). Chemosensory proteins from ***Apis mellifera*** (AmelCSPs) are present as an outgroup.

A


seed Atra d4inwa
seed Bmor d2wcja
AmelOBP 7
AfloOBP 7
AfloOBP 12
AmelOBP 12B
AmelOBP 12A
AfloOBP 2
AmelOBP 2
AmelOBP 3A1
AmelOBP 3A2
AfloOBP 3A2
AfloOBP 4
AmelOBP 4
seed Lmad d1p28a
AmelOBP 10
AfloOBP 10
AfloOBP 14
AmelOBP 14
AfloOBP 19
AmelOBP 20
AmelOBP 19
AfloOBP 20
AmelOBP 16
AfloOBP 16
AfloOBP 16like
AfloOBP 17
AmelOBP 17
AfloOBP 15like
AmelOBP 15
AmelOBP 21
AmelOBP 18
AfloOBP 21like
AfloOBP 21
AfloOBP 13
AmelOBP 13
seed Tmol d1c3za
AfloOBP 9
AmelOBP 9
seed Dmel d1ooha
AmelOBP 11
AfloOBP 11
AfloOBP 5
seed Amel d3r72a
AmelOBP 5
AmelOBP 8A2
AmelOBP 8A1
AfloOBP 8A2like
AfloOBP 8A2
AmelOBP 8B
AmelOBP 6
seed Agam d3v2la
seed Cqui d3ogna
AfloOBP 1
seed Amel d3cz1a
AmelOBP 1
AmelCSP 1
AmelCSP 4
AmelCSP 3
AmelCSP 5
AmelCSP 2
AmelCSP 6

B

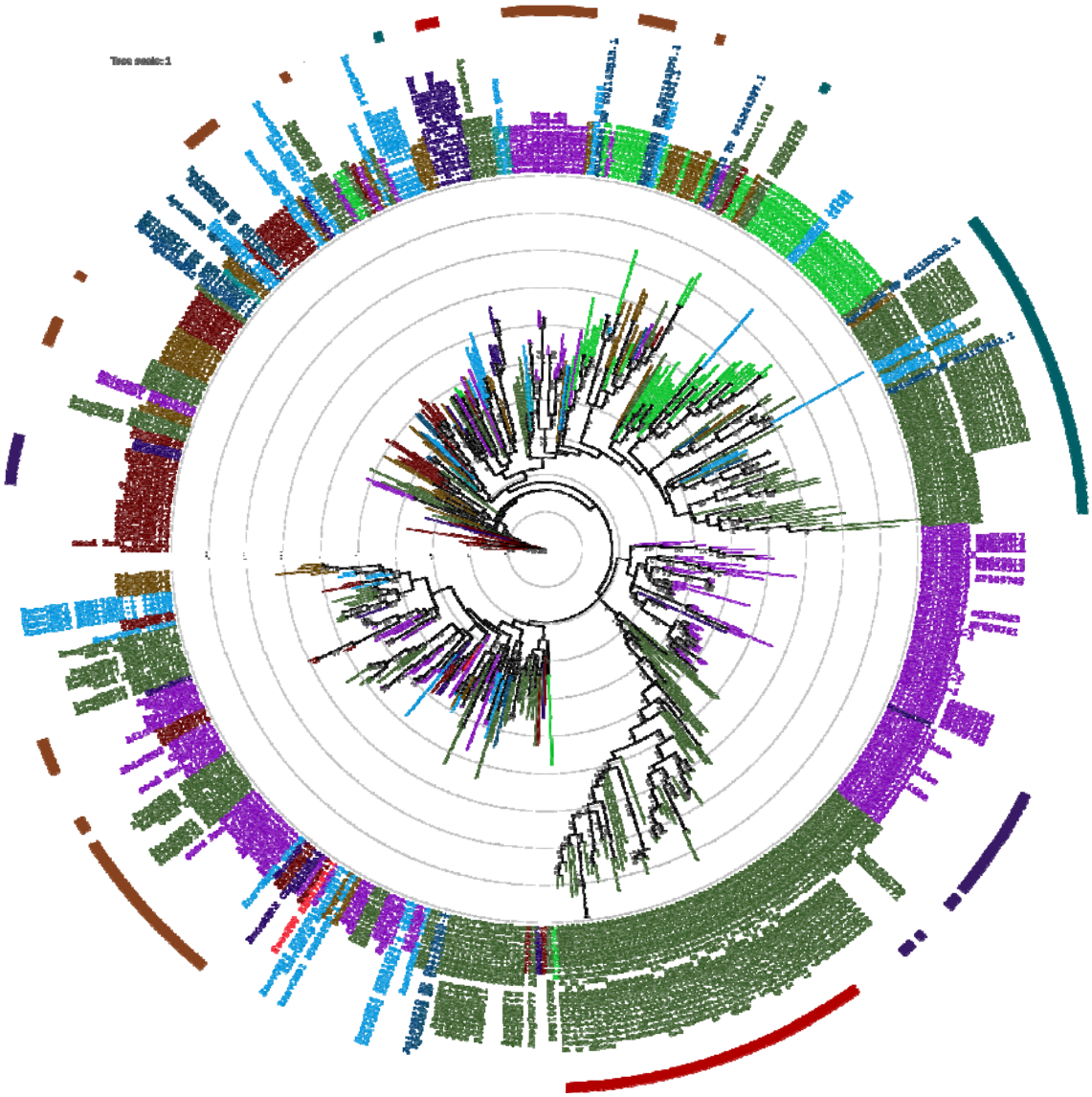


Figure 3: Phylogenetic tree of *Apis florea* OBPs

Rooted phylogeny (A) of OBPs in sister species, ***Apis florea*** *(*Aflo; in pink*)* and ***Apis mellifera*** (Amel; in purple). Members of the alignment template are colored in green, whereas the outgroup consisting of ***Apis mellifera*** chemosensory proteins (AmelCSP) is colored in red. The bootstrap values of the branches are indicated on the nodes in percentage values. **Unrooted phylogeny (B)** of OBPs from representative members of **11 insect orders** represent clade denoting Classic subfamily is colored in **brown,** Atypical in **red**, Minus-C in **violet** and Plus-C in **cyan**. The outer

circle denotes members clades. The inner branch colors and label colors are colored as per order. Hymenoptera is denoted in **violet**. The bootstrap values of the branches are indicated on the nodes in percentage values.

Tables

Sr.no.	Order	Reference
1.	Archaeognatha	Missbach *et al*, 2015
2.	Blattaria	Xu *et al*, 2009
3.	Coleoptera	Xu *et al*, 2009; Gu *et al*, 2015
4.	Diptera	Manoharan *et al*, 2013; Hekmat-Scafe *et al*, 2002, NCBI
5.	Hemiptera	Xu *et al*, 2009; Zhou *et al*, 2010
6.	Hymenoptera	Xu *et al*, 2009; Donnell *et al*, 2013; Gress *et al*, 2014; Li *et al*, 2015
7.	Isoptera	Terrapon *et al*, 2014; NCBI
8.	Lepidoptera	Gong *et al*, 2009; Xu *et al*, 2009; Li *et al*, 2015
9.	Orthoptera	Xu *et al*, 2009
10.	Thysanoptera	NCBI
11.	Zygentoma	Missbach *et al*, 2015

Table 1: List of insect orders represented in the phylogenetic tree of insect OBPs: The protein sequences of OBPs were obtained from an exhaustive literature survey referenced alongside.

AfloOBP	Scaffold ID	Gene model	Complete /Partial/ Pseudo	Length	Ortholog AmelOBP
AfloOBP1_NP_001011590.1	NW_003791204.1-3214074 - 3213446	complement(join(3214077..3213993,3213864..3213776,3213713..3213611,3213408..3213408))	Partial	127	AmelOBP1
AfloOBP2_NP_001011591.1	NW_003789385.1-375403-374446	complement(join(374444..374517,374689..374806,374889..374980,375124..375226,375362..375403))	Complete	143	AmelOBP2
AfloOBP3_l2_XP_006567396.1	NW_003789385.1-369653-368622	complement(join(369650..369653,369330..369393,369053..369144,368830..368938,368623..368599))	Complete	156	AmelOBP3_I2 ABD92639.1
AfloOBP4_NP_001011589.1	NW_003789385.1-368133-366967	complement(join(366944..367050,367140..367251,367460..367551,367726..367780,368089..368133))	Complete	137	AmelOBP4 AF393495 ASP4
AfloOBP5_NP_001011588.1	NW_003789385.1-399844 - 398926	complement(join(398924..399009,399126..399237,399337..399425,399513..399588,399776..399844))	Complete	144	AmelOBP5 AF393497 ASP5
AfloOBP7_NP_	NW_003789385.1-	complement(join(371156	Complete	151	AmelOBP7

001035310.1	372608-371170	..371244,371730..371829,371911..372029,372168..372267,372564..372608))			ABD92640.1
AfloOBP8_2_NP_001164515.1	NW_003789385.1-378413-377643	complement(join(377590..377714,377823..377934,378006..378094,378338..378413))	Complete	134	AmelOBP8_2 AF339140 ASP8
AfloOBP8_2like_XP_006567383.1	NW_003789385.1-398107-397133	complement(join(398113..398042,397868..397949,397534..397622,397343..397454,397134..397155))	Complete	145	Amel OBP8_2
AfloOBP9_NP_001035315.1	NW_003791127.1-6189754 - 6190511	join(6189754..6189807,6189886..6189943,6190062..6190256,6190355..6190419,6190488..6190514)	Complete	133	AmelOBP9 ABD92641.1
AfloOBP10_XP_006566010.1	NW_003790158.1-2089723 - 2091001	join(2089724..2089792,2090128..2090215,2090297..2090349,2090443..2090599,2090829..2090893,2090987..2091004)	Complete	150	AmelOBP10 ABD92642.1
AfloOBP11_NP_001035316.1	NW_003790158.1-2092095 - 2092843	join(2092096..2092170,2092240..2092315,2092383..2092471,2092568..2092673,2092779..2092843)	Partial	137	AmelOBP11 ABD92643.1
AfloOBP12_NP_001035319.1	NW_003791605.1-886575 - 887634	join(886573..886608,886884..886977,887165..887	Partial	153	AmelOBP12_2 ABD92644.1

		280,887359..887488,887555..887637)			
AfloOBP13_NP_001035314.1	NW_003791127.1-1792025 - 1793722	join(1792026..1792073,1793049..1793124,1793206..1793297,1793406..1793508,1793646..1793725)	Complete	133	AmelOBP13 ABD92645.1
AfloOBP14_NP_001035313.1	NW_003791127.1-1798992 - 1800750	join(1798993..1799043,1800073..1800145,1800223..1800314,1800408..1800513,1800668..1800753)	Complete	136	AmelOBP14 ABD92646.1
AfloOBP15-like_NP_001035298.1	NW_003791127.1-1814859 - 1815150	join(1814860..1814953,1815044..1815150,1815049..1815143,1815267..1815353)	Partial	95	AmelOBP15-like
AfloOBP16_NP_001035297.1	NW_003791127.1-1810606 - 1811267	join(1810607..1810679,1810752..1810843,1810942..1811047,1811185..1811270)	Partial	119	AmelOBP16 ABD92648.1
AfloOBP17_NP_001035296.1	NW_003791127.1-1816636 - 1817724	join(1816637..1816687, 1817647..1817719, 1817792..1817883, 1817990..1818095, 1818206..1818291)	Complete	136	AmelOBP17
AfloOBP17-like	NW_003791127.1-1817787 - 1818285	join(1817764..1817883,1817990..1818095,1818206..1818291)	Partial	94	AmelOBP17-like
AfloOBP19_NP_001035299.1	NW_003791127.1-1831794 - 1834099	join(1831795..1831845,1833447..1833519,1833613..1833704,1833789..183	Complete	136	AmelOBP19 ABD92651.1

		3894,1834019..1834102)			
AfloOBP19-like_NP_001035295.1_B	NW_003791127.1-1834614 - 1834698	join(1834615..1834704) possible isoform of previous	Partial	37	AmelOBP19
AfloOBP21-like_NP_001035296.1	NW_003791127.1-1821377 - 1823025	join(1821378..1821428,1822439..1822511,1822582..1822674,1822765..1822870,1822946..1823031)	Complete	137	AmelOBP21
AfloOBP21_NP_001035296.1	NW_003791127.1-1837383 - 1839033	join(1837384..1837434,1838450..1838522,1838593..1838684,1838771..1838876,1838954..1839039)	Complete	136	AmelOBP21 ABD92653.1

Table 2: OBPs from ***Apis florea*** annotated in our study have been listed with the scaffold identity, coding exons, complete or partial status of predicted protein sequence, length of protein sequence and its ortholog in ***Apis mellifera***.

Supplementary Materials:

a. Supplementary File 1- SI_Table1

   Table of OBPs from reference organism *Apis mellifera* derived from various literature sources and re-annotated to obtain a standard dataset

b. Supplementary File 2- SI_Table2

   Table of predicted OBPs from *Apis florea* with annotation

c. Supplementary_File 3- SI_Table3

   Table of blastp alignment hits obtained with query as our annotated set of AfloOBPs against Non-Redundant database.

Author ORCIDs

1. Bhavika Mam: https://orcid.org/0000-0002-3130-0925
2. Snehal Karpe: https://orcid.org/0000-0001-6062-7786
3. Ramanathan Sowdhamini: https://orcid.org/0000-0002-6642-2367

Funding:

BM acknowledges support from NCBS-TIFR and the Tata Education and Development Trust. SDK acknowledges support from Shyama Prasad Mukherjee Fellowship from Council of Scientific and Industrial Research (CSIR) and Bridging Postdoctoral Fellowship from NCBS-TIFR. RS acknowledges support and funding from JC Bose fellowship and from the DBT-Bioinformatics Centre (BT/PR40187/BTIS/137/9/2021).

Conflicts of Interest:

The authors declare no conflict of interest.

Ethics Approval/ declarations:

Not applicable

Data Availability Statement:

All the main data generated or analysed during this study are included in this published article [and its supplementary information files]. Related datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Code availability: Not applicable

Authors' contributions:

RS and SK conceived this research and designed experiments; BM co-designed experiments for protein sequence annotation, performed experiments, analysis, and wrote first draft of the paper. All authors read and approved the final manuscript.

Acknowledgments

BM acknowledges Prof. (Dr.) Axel Brockmann and Prof. (Dr.) Shannon Olsson for valuable comments during the course of the study. All authors acknowledge NCBS for infrastructural support.

References

1. Bhavika Mam, Ramanathan Sowdhamini. SoCCer: A pipeline to identify classes of soluble proteins in chemical communication in insect genomes, 03 September 2020, PREPRINT (Version 1) available at Research Square. DOI: 10.21203/rs.3.pex-1095/v1
2. Forêt S, Maleszka R. Function and evolution of a gene family encoding odorant binding-like proteins in a social insect, the honey bee (***Apis mellifera***) 2006. *Genome Res*. 16(11):1404-1413. DOI:10.1101/gr.5075706
3. Forêt S, Wanner KW, Maleszka R, 2007. Chemosensory proteins in the honey bee: Insights from the annotated genome, comparative analyses and expressional profiling. Insect Biochem Mol Biol. 37(1):19-28. DOI: 10.1016/j.ibmb.2006.09.009. Erratum in: Insect Biochem Mol Biol. 2015 Oct; 65:107. PMID: 17175443.
4. Gandhimathi A, Nair AG, Sowdhamini R, 2012. PASS2 version 4: an update to the database of structure-based sequence alignments of structural domain superfamilies. Nucleic Acids Res: D531-4. DOI: 10.1093/nar/gkr1096.
5. Graham LA, Davies PL. The odorant-binding proteins of ***Drosophila melanogaster***: annotation and characterization of a divergent gene family *2002*. Gene. 292(1-2):43-55. DOI: 10.1016/s0378-1119(02)00672-8.
6. Gupta M 1986. A quantitative study and ultrastructure of flagellar sensillae of ***Apis florea*** F. (Hymenoptera: Apidae). *Proc. Indian Acad. Sci. (Anim. Sci.)*, 95: 5: 595-603.
7. Gupta M, 1992. Scanning electron microscopic studies of antennal sensilla of adult worker ***Apis florea*** F (Hymenoptera: Apidae). *Apidologie*, Springer Verlag, 23 (1), 47-56.
8. Halling L, Oldroyd B, Wattanachaiyingcharoen, W *et al*. 2001. Worker policing in the bee ***Apis florea***. *Behav Ecol Sociobiol* 49, 509–513. DOI: 10.1007/s002650100325

9. He P, Li ZQ, Zhang YF, Chen L, Wang J, Xu L, Zhang YN and He M. 2017. Identification of odorant□binding and chemosensory protein genes and the ligand affinity of two of the encoded proteins suggest a complex olfactory perception system in ***Periplaneta americana***. *Insect Mol Biol*, 26: 687-701. DOI: 10.1111/imb.12328

10. Hekmat-Scafe DS, Scafe CR, McKinney AJ, Tanouye MA 2002. Genome-wide analysis of the odorant-binding protein gene family in ***Drosophila melanogaster***. *Genome Res*. 12(9):1357-1369. DOI:10.1101/gr.239402

11. Henikoff S, Henikoff, J.G, 1992. Amino Acid Substitution Matrices from Protein Blocks. *PNAS*. 89 (22): 10915–1091910919. DOI:10.1073/pnas.89.22.10915

12. Hepburn, RH, Radloff S, Otis G, Fuchs S, Verma L, Tan K, *et al*., 2005. *Apis florea*: morphometrics, classification and biogeography. *Apidologie*, 36(3), 359-376.

13. Hildebrand JG, and Shepherd, GM, 1997. Mechanisms of olfactory discrimination: Converging evidence for common principles across phyla. *Annu. Rev. Neurosci*. 20, 595–631. DOI: 10.1146/annurev.neuro.20.1.595

14. Karpe SD, Dhingra S, Brockmann A. and Sowdhamini R. 2017. Computational genome-wide survey of odorant receptors from two solitary bees ***Dufourea novaeangliae*** (Hymenoptera: Halictidae) and ***Habropoda laboriosa*** (Hymenoptera: Apidae). *Sci. Rep*, 7, 10823.

15. Karpe S.D., Jain R., Brockmann A. and Sowdhamini R. 2016. Identification of Complete Repertoire of ***Apis florea*** Odorant Receptors Reveals Complex Orthologous Relationships with ***Apis mellifera***. Genome Biology and Evolution, 8, 2879–2895.

16. Karpe SD, Tiwari V, Ramanathan S. 2021. InsectOR—Webserver for sensitive identification of insect olfactory receptor genes from non-model genomes. PLOS ONE 16(1): e0245324. DOI: 10.1371/journal.pone.0245324
17. Katoh K, Misawa K, Kuma K, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res*. 30:3059–3066.
18. Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30:772–80.
19. Kumar N, Nayyar K, Sharma R and Anudeep 2014. Ultramorphology of antennal sensilla of open-nesting honey bees ***Apis florea F.*** and ***Apis dorsata F.*** (Hymenoptera: Apidae*) J. Appl. & Nat. Sci.* 6 (1): 315-319.
20. Larter NK, Sun JS, Carlson JR, 2016. Organization and function of ***Drosophila*** odorant binding proteins. Elife. 5:e20242. DOI: 10.7554/eLife.20242
21. Letunic I and Bork P, 2006. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* 23(1):127-8
22. Letunic I and Bork P 2016. Interactive Tree Of Life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res* DOI: 10.1093/nar/gkw290
23. Manoharan M, Ng Fuk Chong M, Vaïtinadapoulé A, Frumence E, Sowdhamini R, Offmann B. 2013 Comparative genomics of odorant binding proteins in ***Anopheles gambiae***, ***Aedes aegypti***, and ***Culex quinquefasciatus***. *Genome Biol Evol*. 5(1):163-180. DOI:10.1093/gbe/evs131
24. Moritz, RF et al. 2010. Invasion of the Dwarf Honeybee ***Apis Florea*** into the Near East. *Biological Invasions* 12.5: 1093-9.

25. Nielsen H, Engelbrecht J, Brunak S and von Heijne G 1997. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Protein Engineering, 10:1-6.

26. Oldroyd B, and Nanork P 2009. Conservation of Asian Honey Bees. *Apidologie* 40.3: 296–312.

27. Otis GW 1991. Revised distribution of three recently recognised species of honey bees Asia. *Honeybee Sci* 15:167-170.

28. Page RE, and Robinson G 1991. The genetics of division of labour in honey bee colonies. *Adv insect physiol* 23: 117-169.

29. Petersen TN, Brunak S, von Heijne G and Nielsen H 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods*, 8:785:786. DOI: 10.1038/nmeth.1701

30. Slater GSC, Birney E. 2005. Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics* 6:31.

31. Stamatakis A 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*. 22:2688–90.

32. Stamatakis A 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. 30:1312–3.

33. Suwannapong G, Seanbualuang P & Benbow ME 2011. Using sensillum potential analysis to quantify pheromone sensing of the antennal sensilla of ***Apis florea Fabricius*** (1787), foragers and guards. *Journal of Asia-Pacific Entomology*, 14(1), 7–10.

34. The UniProt Consortium 2019. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res*. 47: D506-515.

35. Vogt GR, Große-Wilde E, Zhou JJ, 2015. The Lepidoptera Odorant Binding Protein gene family: Gene gain and loss within the GOBP/PBP complex of moths and butterflies, Insect Biochemistry and Molecular Biology: 62: 142-153. DOI:10.1016/j.ibmb.2015.03.003.

36. Missbach C, Vogel H, Hansson BS, Groβe-Wilde E, 2015. Identification of Odorant Binding Proteins and Chemosensory Proteins in Antennal Transcriptomes of the Jumping Bristletail *Lepismachilis y-signata* and the Firebrat *Thermobia domestica:* Evidence for an Independent OBP–OR Origin, *Chemical Senses*: 40:9, 615: 626. DOI: 10.1093/chemse/bjv050

37. Whelan S, Goldman N 2001. A General Empirical Model of Protein Evolution Derived from Multiple Protein Families Using a Maximum-Likelihood Approach, *Molecular Biology and Evolution*, 18:5: 691–699. DOI: 10.1093/oxfordjournals.molbev.a003851

38. Wigglesworth VB 1965. The Principles of Insect Physiology. Methuen Co Ltd, London, 6th edn

39. Wongsiri S, et al. 1997. Comparative biology of ***Apis andreniformis*** and ***Apis florea*** in Thailand. *Bee World* 78.1: 23-35.

40. Wu YR and Kuang B, 1987. Two species of small honeybee - a study of the genus ***Micrapis***. *Bee World*. 68: 153–55. DOI: 10.1080/0005772X.1987.11098924

41. Xu, Y. L., He, P., Zhang, L., Fang, S. Q., Dong, S. L., Zhang, Y. J., & Li, F. 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

42. Zhou YN, Xie S, Chen JN, et al. 2020. Expression and functional characterization of odorant-binding protein genes in the endoparasitic wasp Cotesia vestalis. Insect Science. DOI: 10.1111/1744-7917.12861.