

Genome annotation and comparative analyses of the odorant-binding proteins and chemosensory proteins in the pea aphid *Acyrtosiphon pisum*

J.-J. Zhou^{*1}, F. G. Vieira^{†‡1}, X.-L. He^{*}, C. Smadja[§],
R. Liu^{*}, J. Rozas^{†‡} and L. M. Field^{*}

^{*}Department of Biological Chemistry, Rothamsted Research, Harpenden, Hertfordshire, UK; [†]Departament de Genètica, Universitat de Barcelona, Barcelona, Spain; [‡]Institut de Recerca de la Biodiversitat, Universitat de Barcelona, Barcelona, Spain; and [§]Department of Animal and Plant Sciences, University of Sheffield, Sheffield, UK

Abstract

Odorant-binding proteins (OBPs) and chemosensory proteins (CSPs) are two families of small water-soluble proteins, abundant in the aqueous fluid surrounding olfactory receptor neurons in insect antennae. OBPs are involved in the first step of olfactory signal transduction, carrying airborne semiochemicals to the odorant receptors and can be classified into three groups: Classic OBPs, Plus-C OBPs and Atypical OBPs. Here, we identified and annotated genes encoding putative OBPs and CSPs in the pea aphid *Acyrtosiphon pisum* using bioinformatics. This identified genes encoding 13 Classic and two Plus-C OBPs and 13 CSPs. Homologous OBP sequences were also identified in nine other aphid species, allowing us to compare OBPs across several aphid and non-aphid species. We show that, although OBP sequences are divergent within a species and between different orders, there is a high similarity between orthologs within a range of aphid species. Furthermore, the phylogenetic relationships between OBP orthologs reflect the divergence of aphid evolution lineages. Our results support the 'birth-and-death' model as the major mechanism explaining aphid OBP sequence evolution, with the

main force acting on the evolution being purifying selection.

Keywords: Aphids, olfaction, chemical ecology, gene evolution.

Introduction

Odorant-binding proteins (OBPs) are small, globular, water-soluble proteins which carry airborne semiochemicals to the chemoreceptors in insect antennae (Pophof, 2004; Pelosi *et al.*, 2006; Grosse-Wilde *et al.*, 2006) and in one case it has been suggested that the ligand-OBP complex may activate the receptor directly rather than the OBP releasing the ligand before receptor binding (Laughlin *et al.*, 2008). Most insect OBPs have six highly conserved cysteines (Breer *et al.*, 1990; Krieger *et al.*, 1993) which form disulphide bridges and stabilize their 3D structure (Leal *et al.*, 1999; Scaloni *et al.*, 1999; Sandler *et al.*, 2000; Tegoni *et al.*, 2004), and these have been designated as Classic OBPs. Other OBPs include Plus-C OBPs with eight conserved cysteines and one conserved proline and Atypical OBPs with nine to 10 conserved cysteines. Another protein family thought to be involved in chemoreception are the chemosensory proteins (CSPs) with four conserved cysteines. The sequence motifs of OBPs and CSPs have been used in genome-wide identification and annotation in a wide range of insect species (Li *et al.*, 2004; Zhou *et al.*, 2004, 2006, 2008; Vieira *et al.*, 2007) and detailed comparative analyses of the evolution of the OBP genes have been carried out (Vogt *et al.*, 2002; Sanchez-Gracia *et al.*, 2009).

The publication of the whole genome sequence of the pea aphid *Acyrtosiphon pisum* now provides a platform for the analysis of OBPs and CSPs in aphids in relation to their chemical ecology. Aphids communicate with each other and migrate between host plants using species-specific chemical signals (semiochemicals) such as pheromones and plant volatiles acting as attractants and repellents. The aphid sex pheromones are often similar amongst different species with discrimination relying on blends of two or three compounds in species-specific

Correspondence: Dr Linda M. Field, Department of Biological Chemistry, Rothamsted Research, Harpenden, Hertfordshire AL5 2JQ, UK. Tel.: +44 158276 3133; fax: +44 158276 2595; e-mail: lin.field@bbsrc.ac.uk

¹These authors contributed equally.

ratios (Pickett *et al.*, 1992; Hardie *et al.*, 1997). In addition, behavioural studies show that plant volatile components enhance successful mate location by male aphids (Pickett *et al.*, 1992; Guldmond *et al.*, 1993; Hardie *et al.*, 1994; Lösel *et al.*, 1996; Campbell *et al.*, 2003). Thus, there is considerable interest in understanding how aphids respond to semiochemicals. This is further motivated by the possibility that insight into the response of aphids to host odours could help to develop novel control strategies that interfere with the interactions between pest aphids and crop plants.

Here, we report the annotation of genes encoding putative OBPs and CSPs in the genome of *A. pisum* and the identification of orthologous genes in nine other aphid species from two families and four tribes. We compared aphid OBP sequences with each other and with other insect OBPs. We also address the molecular evolution of the gene family as a whole and the influence of natural selection on the evolution of insect OBP genes.

Results and discussion

OBP and CSP genes in *Acyrtosiphon pisum*

Analyses of the *A. pisum* genome sequence (27 798 scaffolds) and expressed sequence tag (EST; 169 599

sequences) databases identified 15 sequences encoding putative OBPs (four of them are truncated, likely due to incomplete DNA sequencing/assembly) and 13 encoding CSPs (three of them are also truncated) (Table 1). Some of these are very similar (*ApisOBP3*, *ApisOBP11* and *ApisOBP12*) with an identity of 45.2% at the amino acid level. The number of OBPs in *A. pisum* is small compared to that reported in *Drosophila melanogaster* (Hekmat-Scafe *et al.*, 2002; Vieira *et al.*, 2007), *Anopheles gambiae* and *Aedes aegypti* (Zhou *et al.*, 2008) but is comparable to *Apis mellifera* (Forêt & Maleszka, 2006) and *Bombyx mori* (Zhou *et al.*, 2009). Multiple factors are probably responsible for the differences in the OBP numbers. It has been observed that parasitic and symbiotic lifestyles lead to a genome reduction, either by redundancy of function, or as a result of a simpler and homogeneous host environment (Wernegreen, 2002; Moya *et al.*, 2008). *A. pisum*, with what can be considered as a parasitic lifestyle, may have relaxed selective constraint on genes related, for example, to avoidance of hazardous substances, digestive processes and food/mate location.

We have identified six putative OBP genes from 15 611 ESTs of the blood sucking bug *Rhodnius prolixus*, a hemimetabolous insect, and found *A. pisum* OBP orthologs.

Table 1. Odorant-binding proteins and chemosensory proteins annotated in the *Acyrtosiphon pisum* genome

Name	EST redundancy	Signal peptide	Amino acids	Gene ID*	Genome annotation†	No. of intron
ApisOBP1	112	1–21 aa	159 aa	ACYPIG336037	EQ115283 (2092..7402, –)	6
ApisOBP2	29	1–19 aa	243 aa	ACYPIG179180	EQ125284 (8441..12325, +)	4
ApisOBP3	52	1–23 aa	141 aa	ACYPIG117886	EQ113858 (4821..13329, +)	5
ApisOBP4	22	1–22 aa	193 aa	ACYPIG478320	EQ127504 (11369..14618, –)	6
ApisOBP5	21	1–25 aa	221 aa	ACYPIG747500	EQ119281 (57020..67390, +)	8
ApisOBP6	3	1–19 aa	215 aa	ACYPIG250873	EQ122941 (3782..9126, +)	7
ApisOBP7	7	1–30 aa	155 aa	ACYPIG658982	EQ121833 (58105..66623, +)	6
ApisOBP8	2	1–18 aa	162 aa	ACYPIG938564	EQ124790 (36319..42743, +)	6
ApisOBP9	3	1–24 aa	165 aa	ACYPIG781430	EQ112785 (118532..124337, –)	6
ApisOBP10	ND	1–24 aa	143 aa	ACYPIG102270	EQ111016 (18155..20991, –)	6
ApisOBP11	ND	1–23 aa	112 aa	ACYPIG252504	EQ113328 (24063..39281, +)	6
ApisOBP12-CN	ND	No SP	112 aa	ACYPIG244867	EQ124790 (31961..35227, +)	NA
ApisOBP13-NT	NA	No SP	82 aa	ACYPIG620194	EQ121843 (21060..21949, –)	NA
ApisOBP14-CN	NA	NA	26 aa	ACYPIG570521	EQ117403 (522..1405, +)	NA
ApisOBP15-NT	NA	No SP	23 aa	ACYPIG803752	EQ118788 (55206..56130, –)	NA
ApisCSP1	7	1–16 aa	221 aa	ACYPIG722316	EQ110797 (7930..17292, –)	1
ApisCSP2	28	1–20 aa	131 aa	ACYPIG986248	EQ125317 (40244..41374, +)	1
ApisCSP3	10	1–20 aa	123 aa	ACYPIG242345	EQ126525 (52894..53537, +)	1
ApisCSP4	22	1–22 aa	145 aa	ACYPIG606098	EQ117790 (6019..7042, +)	1
ApisCSP5	12	1–19 aa	137 aa	ACYPIG633836	EQ121783 (61045..62836, +)	1
ApisCSP6	3	1–21 aa	131 aa	ACYPIG284520	EQ110797 (29846..31524, –)	1
ApisCSP7	5	No SP	155 aa	ACYPIG785909	EQ122410 (2584..10467, –)	2
ApisCSP8	10	1–37 aa	163 aa	ACYPIG640574	EQ121783 (40804..46054, +)	1
ApisCSP9	16	1–21 aa	176 aa	ACYPIG800452	EQ125317 (41584..44149, –)	1
ApisCSP10	ND	1–21 aa	150 aa	ACYPIG375309	EQ116326 (76913..80468, –)	1
ApisCSP11-NT	ND	1–19 aa	147 aa	ACYPIG632142	EQ110797 (20097..25202, –)	NA
ApisCSP12-CT	NA	NA	53 aa	ACYPIG215889	EQ126525 (61094..61257)	NA
ApisCSP13-CT	NA	NA	78 aa	ACYPIG819119	EQ116510 (25..818, +)	NA

*AphidBase gene identity.

†AphidBase scaffold ID (start.stop nt, orientation).

CN, both C- and N-terminus missing; CT, C-terminus missing; NA, not applied; ND, not detected; NT, N-terminus missing; SP, signal peptide.

For example, *ApisOBP2* has 23.2% and 26.3% identity to *RproOBP3* and *RproOBP6*, respectively. We also found orthologs in the body louse *Pediculus humanus*, a human parasitic insect (Vieira *et al.*, unpubl. data). However, only the hemimetabolous insects *A. pisum* and *R. prolixus* have paralogs with high amino acid identity. Thus there is 36.3% identity between *RproOBP3* and *RproOBP6* and 80.9% between *RproOBP4* and *RproOBP5* which is clustered with *RproOBP2* with an overall identity of 38.9%. There is 36.8% identity between *ApisOBP1* and *ApisOBP8*, and 66.4% between *ApisOBP11* and *ApisOBP12* with 60.3% and 61.0% identity to *ApisOBP3*, respectively. These OBP expansions by gene duplication are much less than observed in the Dipteran insects (Xu *et al.*, 2003; Zhou *et al.*, 2008). In contrast to the varying number of OBPs amongst different insect species there is a more consistent expansion of chemoreceptor genes (odorant and/or gustatory receptors) in *B. mori* (Wanner & Robertson, 2008), *D. melanogaster* (Clyne *et al.*, 2000; Robertson *et al.*, 2003), *Ae. aegypti* (Bohbot *et al.*, 2007; Kent *et al.*, 2008), *An. gambiae* (Fox *et al.*, 2001), *Apis mellifera* (Robertson & Wanner, 2006), *Tribolium castaneum* (Engsontia *et al.*, 2008) and *A. pisum* (Smadja *et al.*, 2009).

Genomic structure of *Acyrtosiphon pisum* OBP and CSP genes

The *A. pisum* OBP genes are generally sparsely distributed across 14 scaffolds (Fig. 1). However, the CSP genes are clustered with *ApisCSP1*, *ApisCSP6* and *ApisCSP11* on EQ110797 within a 23594 base pair (bp) region; *ApisCSP2* and *ApisCSP9* on EQ125317 within a 3905 bp region (210 bp apart); *ApisCSP5* and *ApisCSP8* on EQ121783 within a 22032 bp region and *ApisCSP3* and *ApisCSP12* on EQ126525 within a 8363 bp region.

The *A. pisum* OBP genes have more and longer introns than their counterparts in *D. melanogaster* with an average intron number of 6.0 introns per gene ($n = 11$ genes) for *A. pisum* and 1.5 per gene ($n = 51$ genes) for *D. melanogaster*. The average intron length is 6111.8 bp ($n = 77$ introns), ranging from 58 bp to 118532 bp for *A. pisum* and 93.2 bp ($n = 90$ introns), ranging from no intron to 638 bp for *D. melanogaster* (Fig. 1). These data are consistent with results showing that *D. melanogaster* has experienced a drastic reduction in non-coding DNA including introns (Petrov *et al.*, 1996; Zdobnov *et al.*, 2002). Two of the *A. pisum* OBP genes, *ApisOBP5* and *ApisOBP6* encode proteins predicted to be Plus-C OBPs and this is the first report of this type of OBP outside of dipteran insects, indicating that Plus-C OBPs must have evolved before the divergence of the aphids from the dipterans.

Orthologous genes in other aphid species

Orthologs of the *A. pisum* OBP genes were identified in nine other aphid species using PCR with primers designed to the *A. pisum* sequences (Table S1). The gene encoding OBP2 is present in all species, except *Tubero-lachnus salignus*, and further analysis of aphid OBP2 shows that the sequences cluster into three groups, which correspond with the three aphid 'tribes', the Macrosiphini, the Pterocommatini and the Aphidini (Fig. 2 and Table S2). The average number of orthologous genes found was 6, 3, 5 and 1 for Macrosiphini, Aphidini, Pterocommatini and Lachnini, respectively. Only one orthologous OBP was found in *Tu. salignus*, which belongs to a different family from the other aphids. Thus, although we cannot exclude the presence of undetected OBP genes in these aphid species, the distribution of the OBP orthologs does reflect the life style and the host relationship of the aphid species within the tribes. The morphology of Pterocommatini and their simple life cycles on woody hosts are regarded as primitive and this tribe is placed as sister to Aphidini plus Macrosiphini (Blackman & Eastop, 2000), which is supported by our analysis of OBP genes (Fig. 2 and Table S2).

Phylogenetic analysis of aphid OBP genes

The phylogenetic relationships of the predicted OBPs in the nine aphids, and other insect species (*D. melanogaster*, *An. gambiae*, *B. mori*, *T. castaneum*, and *Apis mellifera*) are shown in Fig. 3. This reveals a divergent repertoire with only a few clear orthologous groups that include non-aphid species, possibly reflecting the OBP gene family's evolutionary process, dominated by a number of gene losses and lineage-specific expansions. The two orthologous groups with a clear member across all sequenced insects (apart from Hymenoptera) are those that include *ApisOBP4* and *ApisOBP13* (Figs 3 and 4). Interestingly, some members of the *ApisOBP4* group, for example *DmelOBP73a* have not been assigned as OBPs previously because of their divergence from other OBP members (Hekmat-Scafe *et al.*, 2002; Vieira *et al.*, 2007). The high conservation of OBPs in this group, across a large number of divergent species indicates a possible crucial function for these proteins.

The phylogenetic relationship of the OBPs from the aphid species is shown in Fig. 4 and this is largely consistent with the accepted species tree (von Dohlen *et al.*, 2006). The tree shows higher divergence times for paralogs compared with orthologs, long tree branches and a scattered phylogenetic distribution, indicating that the OBP gene family is quite old (with the MRCA tracing back to the origin of insects – 350–400 million years ago). In addition, the analysis including some *R. prolixus* OBPs

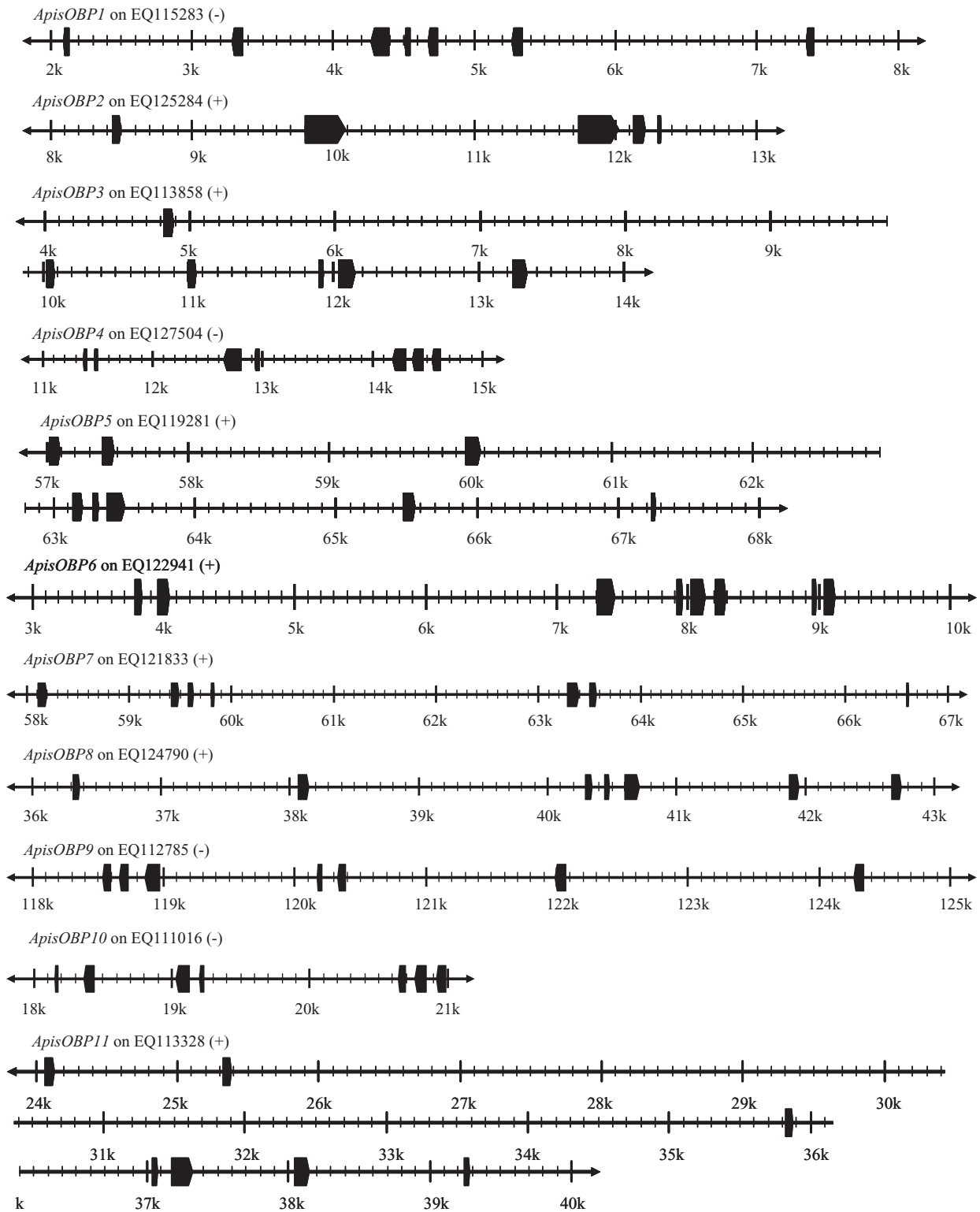


Figure 1. Structures of *OBP* genes in *Acyrthosiphon pisum*. The gene structures were drawn using the genomic coordinates of each *OBP* gene on its scaffold. The two-arrowed lines are the scaffold region which contains the *OBP* gene. The scales are for every 100 bp (minor marks) and 1000 (major marks) and are indicated below the scaffold line. The names of the *OBP* genes and scaffolds are indicated above the line. The transcriptional direction of each *OBP* gene is indicated in the parenthesis after the scaffold name: (+) for same direction as the scaffold and (-) for opposite direction (some very long genomic regions are presented in several lines with a continuous scale).

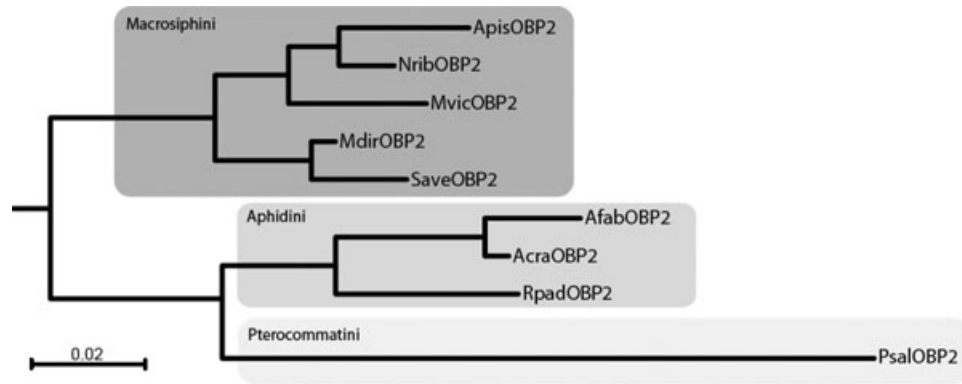


Figure 2. Relationship between orthologs of *OBP2* in nine aphid species. The species name is represented with a four-letter abbreviation: Apis for *Acyrtosiphon pisum*, Mdir for *Metopolophium dirhodum*, Save for *Sitobium avenae*, Nrib for *Nasonovia ribis-nigri*, Mvic for *Megoura viciae*, Psal for *Pterocomma salicis*, Rpad for *Rhopalosiphon padi*, Afab for *Aphis fabae* and Acra for *Aphis craccivora*. The tree was displayed using FigTree v1.2.1 (<http://tree.bio.ed.ac.uk/software/figtree>). The scale bar represents the number of substitutions per site. All accession numbers of the *OBP* genes used are provided in Table S4.

also supports a highly dynamic evolutionary process for this family. In fact, and in spite of the close relationship with *R. prolixus*, we were able to detect several lineage specific expansions and few common orthologous groups.

Overall, the analysis of the aphid OBPs further supports the ‘birth-and-death’ evolutionary model (Nei & Rooney, 2005) as the major mechanism for the evolution of this gene family. That is, in aphids, OBPs would originate by tandem gene duplication and gradually diverge from each other in sequence (and presumably also in function) while others could eventually be lost (transiently by a pseudogenization event). This would lead to the identification of more orthologous groups at short time scales (among aphids) with better phylogeny coordination and to inferring several gene duplications and putative non-functional members (pseudogenes).

Impact of natural selection

We determined the impact of natural selection on the evolution of the *OBP* coding regions from the nine aphid species using only those groups with at least five sequences: *OBP2*, *OBP3*, *OBP4*, *OBP5*, *OBP8* and *OBP10* groups (the *OBP1* group was not analysed due to the low level of sequence variability). To avoid the problem of the saturation of substitutions we analysed each orthologous group separately.

Estimates of the ω values ranged from 0.11 to 0.30 (Table S3a); which is similar to that obtained in *D. melanogaster* (average $\omega = 0.153$; Vieira *et al.*, 2007) and point to purifying selection as a major selective force. The comparative analyses of the M0 and FR models reveal that, in general, the M0 model fits the data better than does the FR model, with the only exception being the *OBP2* orthologous group (LRT; $P = 0.0286$). There is a slight

indication of positive selection but the branch is too small (Table S3b). Nevertheless, the statistical power seems to be low since the only significant group (the *OBP2* group) is the one with more sequences ($n = 9$) (Fig. 4) and the remaining orthologous groups have a relatively low number of sequences.

For the analysis of the putative heterogeneity in the distribution of the ω rates along the coding region we contrasted the M0 and the M3 ($k = 2$) models. We found that in all cases the M3 model was the best fit to the data (LRT; $P < 0.001$). We then tested whether the heterogeneity results from some form of positive selection. After contrasting the M7 and M8 models, the null hypothesis (M7 model) was again only rejected in the large *OBP2* orthologous group (LRT; $P = 0.0221$). In addition, the LRT between the more conservative M8a and M8 models is also significant (LRT; $P = 0.0407$), further supporting the positive selection analysis. More specifically, the analysis allows us to identify a single amino acid, Ser88 in *ApisOBP2* with the positive selection hallmark ($\omega = 1.851$; $PP_{(\omega>1)} = 0.977$). It is worth emphasizing that, despite the observed high functional constraint levels, the large time-scale analysed may obscure short and/or recent/ongoing episodes of molecular adaptation. In similar analyses of the *Drosophila* *OBP* gene family (Vieira *et al.*, 2007), the fingerprint of positive selection was only identified using other approaches (Sánchez-Gracia & Rozas, 2008).

Conclusions

The genome sequence of *A. pisum*, has allowed us to identify *OBP* and *CSP* genes in a range of aphid species. This had not been possible previously using homology cloning (Jacobs *et al.*, 2005). The small number of OBPs

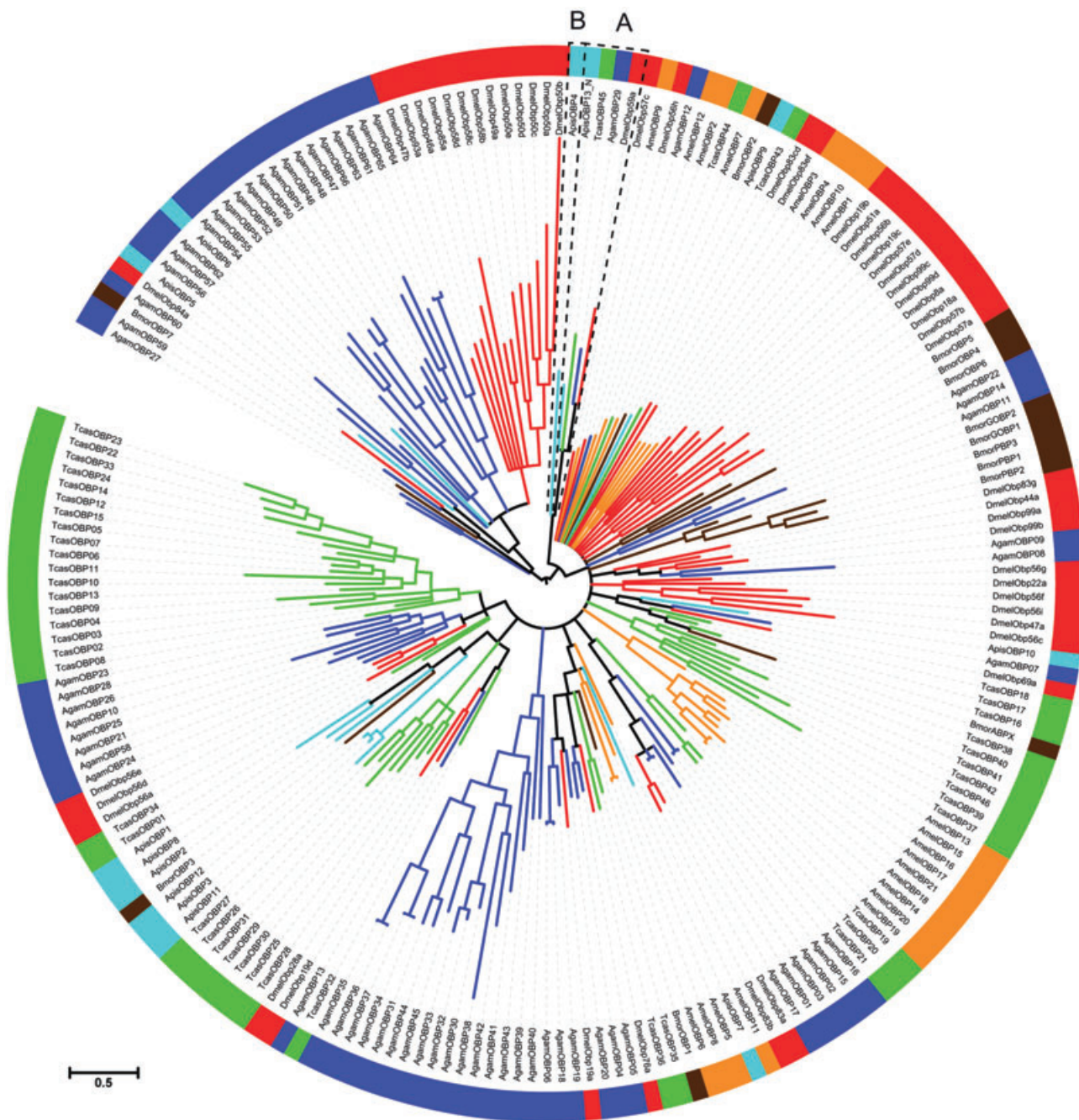


Figure 3. Odorant-binding protein (OBP) phylogenetic relationships from several insect species. The branches are colour coded for each insect species: *Acyrthosiphon pisum* (*Apis*) OBPs in cyan, *Drosophila melanogaster* (*Dmel*) in red, *Anopheles gambiae* (*Agam*) in blue, *Bombyx mori* (*Bmor*) in brown, *Tribolium castaneum* (*Tcas*) in green and *Apis mellifera* (*Amel*) in orange. Only published and complete OBP sequences are included, except for *Apis*OBP13_N. A and B represent the two conserved ortholog groups (see Fig. 4). The tree was obtained using software MrBayes V.3.1.2 and displayed using the iTOL web server (Letunic & Bork, 2007) with the scale bar representing the number of amino acid substitutions per site. All accession numbers of the OBP genes used are provided in Table S4.

in *A. pisum* relative to dipteran insects may be due to the highly specialized ecology of the aphid (host plant specialization) and its parasitic lifestyle. The high similarity of *OBP* genes in different aphid species indicates that *OBP*

genes become divergent before aphid speciation through the mechanisms proposed by the birth-and-death model. Furthermore lifestyle and environmental factors may be the main forces driving the expansion of insect OBPs.

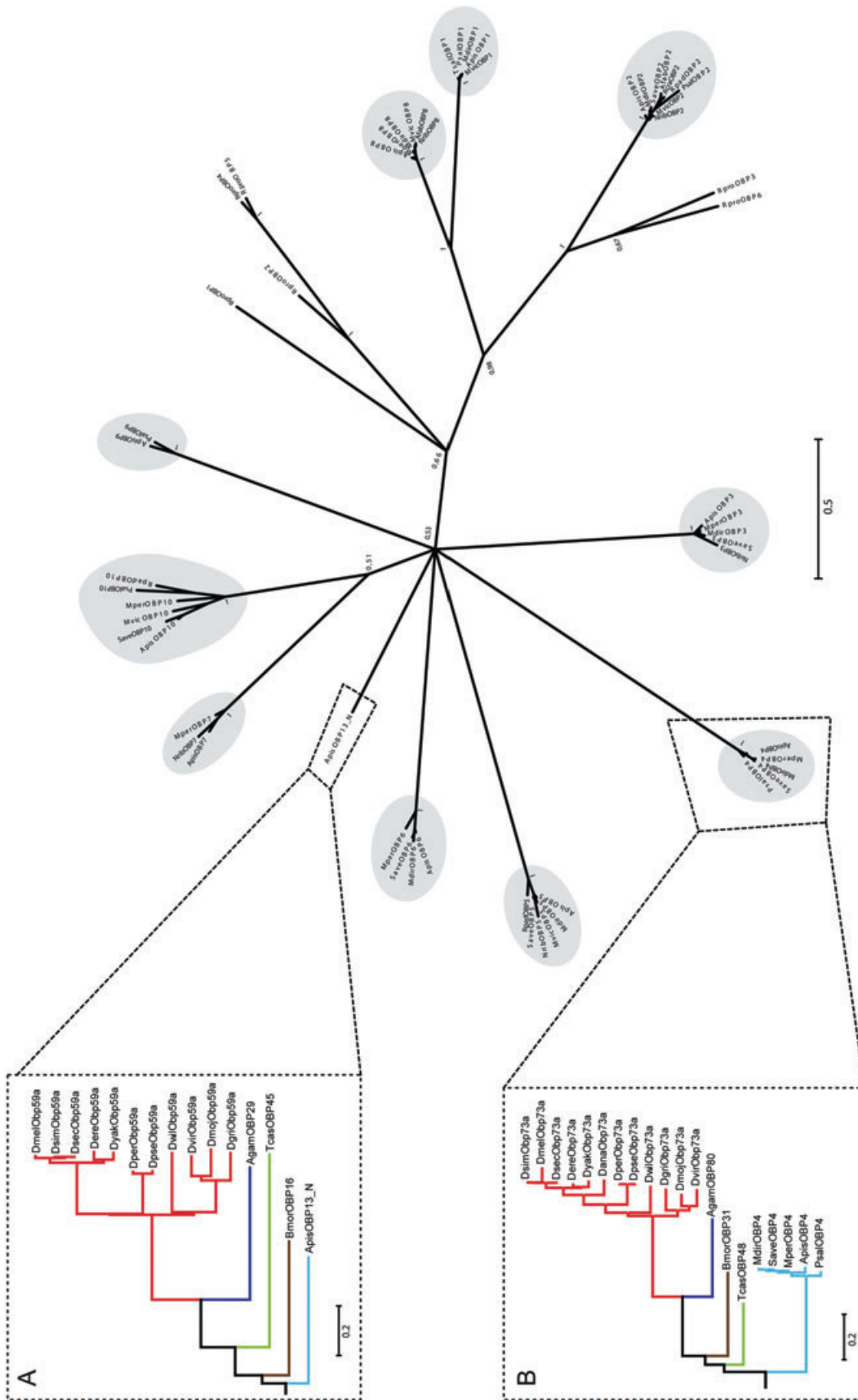


Figure 4. Phylogenetic relationships of the OBP proteins from nine aphid species and *Rhodnius prolixus* (Rpro). The tree was displayed using FigTree v1.2.1 (<http://tree.bio.ed.ac.uk/software/figtree/>). Aphid orthologous groups are highlighted in grey. Truncated sequences are not included, except for the ApisOBP13_N. Left part: detailed view of the two highly conserved orthologous groups (A and B): *Drosophila simulans*, *Drosophila sechellia*, *Drosophila yakuba*, *Drosophila erecta*, *Drosophila pseudoobscura*, *Drosophila persimilis*, *Drosophila virilis*, *Drosophila grimshawi*, *Drosophila ananassae*, *Drosophila willstoni*. In these two groups we also include some newly identified OBP members. The numbers in the nodes represent Bayesian posterior probabilities. The scale bars represent the number of amino acid substitutions per site. All accession numbers of the OBP genes used are provided in Table S4.

Experimental procedures

Identification sequences encoding putative OBPs and CSPs in the *Acyrtosiphon pisum* genome

The whole genome sequences and predicted gene model sets of *A. pisum* were downloaded from Aphidbase (<http://genouest.org/AphidBase/>) and the EST sequences retrieved from the NCBI EST database (<http://www.ncbi.nlm.nih.gov/dbEST/>). The genome sequences were searched using (1) an OBP 'MotifSearch' algorithm to identify the conserved cysteine motif C1-X8-41-C2-X3-C3-X21-47-C4-X7-15-C5-X8-C6 in the 6-frame translated sequences (Zhou *et al.*, 2004, 2006, 2008); (2) rps-BLAST with the PBP/GOBP (pfam01395) and CSP (pfam03392) conserved domains and 3) tBLASTn and PSI-BLAST using, as 'query', known OBPs from other insects. For the predicted gene set, the same basic methodology was used but with (1) BLASTp and (2) HMMER.

Aphid material

Aphids were reared as parthenogenic clones at 22 °C in a 16 h light: 8 h dark regime. Wingless morphs of mixed ages were collected, frozen in liquid nitrogen and stored at -80 °C prior to use.

Cloning and sequencing of OBP genes

Whole insects were ground in liquid nitrogen and total RNA extracted using RNAqueous kit (Ambion, Huntingdon, UK) and treated with DNaseI (Sigma, St Louis, MO, USA). RT-PCRs were done with each primer pair using Hotstart Taq DNA polymerase (Qiagen, Valencia, CA, USA). The PCR primers were designed on the pea aphid OBP sequences with Primer3 (<http://www-genome.wi.mit.edu/cgi-bin/primer>) (Table S1). The PCR products were run on 1% agarose gels and stained with ethidium bromide to check that the correct products were being amplified. They were then purified using a Qiagen kit and sequenced in both directions with the ABI BigDye™ Terminator Cycle Sequencing Ready Reaction Kit (PE Applied Biosystems, Foster City, CA, USA).

Multiple sequence alignments

Two types of multiple sequence alignment (MSA) were generated, one with amino acid and one with nucleotide coding sequences (CDS). Protein sequences were aligned using MAFFT (Kato *et al.*, 2005), with the following settings: E-INS-i with BLO-SUM30 matrix, maximum 10 000 iterations, gap opening penalty '1.53' (default) and offset (equivalent to gap extension penalty '0'). The OBP peptide signal was removed (using PrediSi software; Hiller *et al.*, 2004) prior to the alignment. The MSA for the CDS orthologous regions was done by first aligning the amino acid sequences and then using this alignment to guide the nucleotide CDS alignment.

Phylogenetic analysis

Phylogenetic relationships between homologous OBPs (both orthologs and paralogs) were obtained using the software MrBayes v3.1.2 (Ronquist & Huelsenbeck, 2003), under the WAG

evolutionary model of amino-acid evolution (Whelan & Goldman, 2001). The analysis used the default parameters except: 'stoprule = yes', 'stopval = 0.005', 'samplefreq = 1000' and 'burnin = 20%'.

The impact of natural selection on the CDS of the OBP genes was deduced by analysing the non-synonymous to synonymous divergence ratio ($\omega = d_n/d_s$) using the program 'codeml' of the software package PAML v4.1 (Yang, 2007) (this estimates by maximum likelihood the ω parameter under several evolutionary scenarios) and the phylogenetic relationships of von Dohlen *et al.* (2006). To assess for heterogeneity across branches the M0 (a single ω ratio for all lineages and sites) and the free ratios (FR; allows for different ω rates across branches) models were compared. To analyse the ω rate heterogeneity across sites M0 was compared with M3 ($k = 2$) (one ω ratio for all lineages; two ω categories of sites) models. To determine the presence of positive selection the M7 (one ω ratio for all lineages; 10 ω categories of sites following a beta distribution) and M8 (one ω rate for all lineages; 10 ω categories of sites following a beta distribution plus one extra site with $\omega > 1$) models were compared and, in order to be conservative, the M8a (one ω rate for all lineages; 10 ω categories of sites following a beta distribution plus one extra site with $\omega = 1$) with the M8 models (Swanson *et al.*, 2003; Wong *et al.*, 2004). The comparison between models was assessed using likelihood-ratio tests (LRTs) for hierarchical models (Anisimova *et al.*, 2001), a significantly higher likelihood of the alternative model than that of the null model indicating positive selection in the dataset examined. The posterior probability (PP) that a given site evolves under positive selection was estimated applying the BEB method (implemented in PAML software).

Acknowledgements

Rothamsted Research receives grant-aided support from the BBSRC of the UK. We thank Janet Martin and Lesley Smart at Rothamsted Research who provided us with *A. pisum*, *Myzus persicae*, *Metopolophium dirhodum*, *Megoura viciae*, *Nasonovia ribis-nigri*, *Sitobium avenae*, *Rhopalosiphon padi*, *Aphis fabae* and *Gia Aradottir* for supplying *Pterocomma salicis* and *Tu. salignus*. FGV was supported by the predoctoral fellowship SFRH/BD/22360/2005 from the 'Fundação para a Ciência e a Tecnologia' (Portugal). This work was funded by grant BFU2007-62927 from the 'Dirección General de Investigación Científica y Técnica' (Spain) to JR. We thank the International Aphid Genomics Consortium and the Baylor College of Medicine Human Genome Sequencing Centre for making the *A. pisum* genome sequences publicly available prior to publication.

References

- Anisimova, M., Bielawski, J.P. and Yang, Z. (2001) Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. *Mol Biol Evol* **18**: 1585–1592.
- Blackman, R.L. and Eastop, V.F. (2000) *Aphids on the World's Crops: an Identification and Information Guide*, 2nd edn. John Wiley & Sons Ltd, Chichester.

- Bohbot, J., Pitts, R.J., Kwon, H.W., Rützler, M., Robertson, H.M. and Zwiebel, L.J. (2007) Molecular characterization of the *Aedes aegypti* odorant receptor gene family. *Insect Mol Biol* **16**: 525–537.
- Breer, H., Krieger, J. and Raming, K. (1990) A novel class of binding proteins in the antennae of the silkworm *Antheraea pernyi*. *Insect Biochem* **20**: 735–740.
- Campbell, C.A., Cook, F.J., Pickett, J.A., Pope, T.W., Wadhams, L.J. and Woodcock, C.M. (2003) Responses of the aphids *Phorodon humuli* and *Rhopalosiphum padi* to sex pheromone stereochemistry in the field. *J Chem Ecol* **29**: 2225–2234.
- Clyne, P.J., Warr, C.G. and Carlson, J.R. (2000) Candidate taste receptors in *Drosophila*. *Science* **287**: 1830–1834.
- von Dohlen, C.D., Rowe, C.A. and Heie, O.E. (2006) A test of morphological hypotheses for tribal and subtribal relationships of Aphidinae (Insecta: Hemiptera: Aphididae) using DNA sequences. *Mol Phylogenet Evol* **38**: 316–329.
- Engsontia, P., Sanderson, A.P., Cobb, M., Walden, K.K., Robertson, H.M. and Brown, S. (2008) The red flour beetle's large nose: an expanded odorant receptor gene family in *Tribolium castaneum*. *Insect Biochem Mol Biol* **38**: 387–397.
- Forêt, S. and Maleszka, R. (2006) Function and evolution of a gene family encoding odorant binding-like proteins in a social insect, the honey bee (*Apis mellifera*). *Genome Res* **16**: 1404–1413.
- Fox, A.N., Pitts, R.J., Robertson, H.M., Carlson, J.R., and Zwiebel, L.J. (2001) Candidate odorant receptors from the malaria vector mosquito *Anopheles gambiae* and evidence of down-regulation in response to blood feeding. *Proc Natl Acad Sci USA* **98**: 14693–14697.
- Grosse-Wilde, E., Svatos, A. and Krieger, J. (2006) A pheromone-binding protein mediates the bombykol-induced activation of a pheromone receptor *in vitro*. *Chem Senses* **31**: 547–555.
- Guldemond, J.A., Dixon, A.F.G., Pickett, J.A., Wadhams, L.J. and Woodcock, C.M. (1993) Specificity of sex pheromones, the role of host-plant odour in the olfactory attraction of males, and mate recognition in the aphid *Cryptomyzus*. *Physiol Entomol* **18**: 137–143.
- Hardie, J., Visser, J.H. and Piron, P.G.M. (1994) Perception of volatiles associated with sex and food by different adult forms of the black bean aphid, *Aphis fabae*. *Physiol Entomol* **19**: 278–284.
- Hardie, J., Peace, L., Pickett, J.A., Smiley, D.W.M., Storer, J.R. and Wadhams, L.J. (1997) Sex pheromone stereochemistry and purity affect field catches of male aphids. *J Chem Ecol* **23**: 2547–2554.
- Hekmat-Scafe, D.S., Scafe, C.R., McKinney, A.J. and Tanouye, M.A. (2002) Genomewide analysis of the odorant-binding protein gene family in *Drosophila melanogaster*. *Genome Res* **12**: 1357–1369.
- Hiller, K., Grote, A., Scheer, M., Munch, R. and Jahn, D. (2004) PrediSi: prediction of signal peptides and their cleavage positions. *Nucleic Acids Res* **32**: 375–379.
- Jacobs, S.P., Liggins, A.P., Zhou, J-J, Pickett, J.A., Pelosi, P. and Field, L.M. (2005) OS-D-like genes and their expression in aphids (Hemiptera:Aphididae). *Insect Mol Biol* **14**: 423–432.
- Katoh, K., Kuma, K., Toh, H. and Miyata, T. (2005) MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res* **33**: 511–518.
- Kent, L.B., Walden, K.K. and Robertson, H.M. (2008) The Gr family of candidate gustatory and olfactory receptors in the yellow-fever mosquito *Aedes aegypti*. *Chem Senses* **33**: 79–93.
- Krieger, J., Gänssle, H., Raming, K. and Breer, H. (1993) Odorant binding proteins of *Heliothis virescens*. *Insect Biochem Mol Biol* **23**: 449–456.
- Laughlin, J.D., Ha, T.S., Jones, D.N. and Smith, D.P. (2008) Activation of pheromone-sensitive neurons is mediated by conformational activation of pheromone-binding protein. *Cell* **133**: 1255–1265.
- Leal, W.S., Nikonova, L. and Peng, G. (1999) Disulfide structure of the pheromone binding protein from the silkworm moth, *Bombyx mori*. *FEBS Lett* **24**: 85–90.
- Letunic, I. and Bork, P. (2007) Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* **23**: 127–128.
- Li, Z.X., Pickett, J.A., Field, L.M. and Zhou, J-J. (2004) Identification and expression of odorant-binding proteins of the malaria-carrying mosquitoes *Anopheles gambiae* and *Anopheles arabiensis*. *Arch Insect Biochem Physiol* **58**: 175–189.
- Lösel, P.M., Lindemann, M., Scherkenbeck, J., Maier, J., Engelhard, B., Campbell, C.A.M. et al. (1996) The potential of semiochemicals for control of *Phorodon humuli* (Homoptera: Aphididae). *Pesticide Sci* **48**: 293–303.
- Moya, A., Peretó, J., Gil, R. and Latorre, A. (2008) Learning how to live together: genomic insights into prokaryote–animal symbioses. *Nat Rev Genet* **9**: 218–229.
- Nei, M. and Rooney, A.P. (2005) Concerted and birth-and-death evolution of multigene families. *Annu Rev Genet* **39**: 121–152.
- 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.
- Petrov, D.A., Lozovskaya, E.R. and Hartl, D.L. (1996) High intrinsic rate of DNA loss in *Drosophila*. *Nature* **384**: 346–349.
- Pickett, J.A., Wadhams, L.J., Woodcock, C.M. and Hardie, J. (1992) The chemical ecology of aphids. *Ann Rev Ent* **37**: 67–90.
- Pophof, B. (2004) Pheromone-binding proteins contribute to the activation of olfactory receptor neurons in the silkmoths *Antheraea polyphemus* and *Bombyx mori*. *Chem Senses* **29**: 117–125.
- Robertson, H.M. and Wanner, K.W. (2006) The chemoreceptor superfamily in the honey bee, *Apis mellifera*: expansion of the odorant, but not gustatory, receptor family. *Genome Res* **16**: 1395–1403.
- Robertson, H.M., Warr, C.G. and Carlson, J.R. (2003) Molecular evolution of the insect chemoreceptor gene superfamily in *Drosophila melanogaster*. *Proc Natl Acad Sci USA* **100** (Suppl 2): 14537–14542.
- Ronquist, F. and Huelsenbeck, J.P. (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**: 1572–1574.
- Sanchez-Gracia, A. and Rozas, J. (2008) Divergent evolution and molecular adaptation in the *Drosophila* odorant-binding protein family: inferences from sequence variation at the OS-E and OS-F genes. *BMC Evol Biol* **8**: 323.
- Sanchez-Gracia, A., Vieira, F.G. and Rozas, J. (2009) Molecular evolution of the major chemosensory gene families in insects. *Heredity* **103**: 208–216.
- 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.
- 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 Comm* **266**: 386–391.
- Smadja, C., Shi, P., Butlin, R.K. and Robertson, H.M. (2009) Large gene expansions and adaptive evolution of odorant and gustatory receptor genes in the pea aphid, *Acyrtosiphon pisum*. *Mol Biol Evol* **26**: 2073–2086.
- Swanson, W.J., Nielsen, R. and Yang, Q. (2003) Pervasive adaptive evolution in mammalian fertilization proteins. *Mol Biol Evol* **20**: 18–20.
- Tegoni, M., Campanacci, V. and Cambillau, C. (2004) Structural aspects of sexual attraction and chemical communication in insects. *Trends Biochem Sci* **29**: 257–264.
- 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.
- Vogt, R.G., Rogers, M.E., Franco, M.D. 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.
- Wanner, K.W. and Robertson, H.M. (2008) The gustatory receptor family in the silkworm moth *Bombyx mori* is characterized by a large expansion of a single lineage of putative bitter receptors. *Insect Mol Biol* **17**: 621–629.
- Wernegreen, J.J. (2002) Genome evolution in bacterial endosymbionts of insects. *Nat Rev Genet* **3**: 850–861.
- Whelan, S. and Goldman, N. (2001) A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol Biol Evol* **18**: 691–699.
- Wong, W.S., Yang, Z., Goldman, N. and Nielsen, R. (2004) Accuracy and power of statistical methods for detecting adaptive evolution in protein coding sequences and for identifying positively selected sites. *Genet* **168**: 1041–1051.
- 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.
- Yang, Z. (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* **24**: 1586–1591.
- Zdobnov, E.M. et al. (2002) Comparative genome and proteome analysis of *Anopheles gambiae* and *Drosophila melanogaster*. *Science* **298**: 149–159.
- Zhou, J.-J., Huang, W.-S., Zhang, G.A., Pickett, J.A. and Field, L.M. (2004) 'Plus-C' odorant-binding protein genes in two *Drosophila* species and the malaria mosquito *An. gambiae*. *Gene* **327**: 117–129.
- Zhou, J.-J., Kan, Y., Antoniw, J., Pickett, J.A. and Field, L.M. (2006) Genome and EST analyses and expression of a gene family with putative functions in insect chemoreception. *Chem Senses* **31**: 453–465.
- Zhou, J.-J., He, X.L., Pickett, J.A. and Field, L.M. (2008) Identification of odorant-binding proteins of the yellow fever mosquito *Aedes aegypti*: genome annotation and comparative analyses. *Insect Mol Biol* **17**: 147–163.
- Zhou, J.-J., Robertson, G., He, X., Dufour, S., Hooper, A.M., Pickett, J.A. et al. (2009) Characterisation of *Bombyx mori* odorant-binding proteins reveals that a 'General Odorant-Binding Protein' discriminates between sex pheromone components. *J Mol Biol* **389**: 529–545.

Supporting Information

Additional Supporting Information may be found in the online version of this article under the DOI reference: DOI 10.1111/j.1365-2583.2009.00919.x

Table S1. Primers for homologous identification of odorant-binding proteins in other aphids.

Table S2. Orthologs out of 10 *Acyrtosiphon pisum* odorant-binding proteins in other aphid species as detected by RT-PCR with ApisOBP primers.

Table S3. (a) Estimates of the omega values for each aphid orthologous group; (b) FR model results for the AphidOBP2 orthologous group.

Table S4. The GenBank accession numbers of odorant-binding proteins used in this study.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.