Chem. Senses 31: 453–465, 2006

doi:10.1093/chemse/bjj050
Advance Access publication March 31, 2006

Genome and EST Analyses and Expression of a Gene Family with Putative Functions in Insect Chemoreception

Jing-Jiang Zhou[1], Yunchao Kan[2], John Antoniw[1], John A. Pickett[1] and Linda M. Field[1]

[1]Rothamsted Research, Harpenden, AL5 2JQ, UK and [2]Nanyang-Rothamsted Joint Laboratory for Insect Biology, Nanyang Normal University, Nanyang, People's Republic of China

Correspondence to be sent to: Jing-Jiang Zhou, Biological Chemistry Division, Rothamsted Research, Harpenden, AL5 2JQ, UK. e-mail: jing-jiang.zhou@bbsrc.ac.uk

Abstract

Odorant-binding proteins (OBPs) are thought to be responsible for the transport of semiochemicals across hydrophobic interfaces to olfactory receptors. In insects, a second class of OBPs with four conserved cysteines has been variously named as sensory appendage proteins, olfactory segment-D proteins, and chemosensory proteins (CSPs). The physiological functions of these proteins have remained elusive. Here we report a comprehensive survey of both genome and expressed sequence tags (EST) databases. This showed that CSPs are apparently only present in the phylum, Arthropoda, and in two subphyla, Crustacea and Uniramia. This is the first report of a putative CSP in Crustacea and suggests that the origin of these genes predates the divergence of Uniramia and Crustacea. For the Uniramia, we identified 74 new genes encoding putative CSPs of insect species from 10 different orders. Using tissue-specific EST libraries, we have examined the relative expression of putative CSP genes in many tissues from 22 insect species suggesting that the genes are expressed widely. One *Drosophila* CSPs is expressed sixfold higher in head than other CSPs. One *Bombyx mori* CSPs was found at a very high level in pheromone gland, and for the first time, six CSPs were identified in *B. mori* compound eyes. The different frequencies of CSP transcripts were observed between solitary and gregarious EST libraries of *Locusta migratoria*.

Key words: CSPs, OBPs, odorant binding, olfaction, semiochemicals

Introduction

Insects can sense a wide range of chemical signals (semiochemicals), such as pheromones and host odors, which they use to detect each other and to locate suitable food supplies. The semiochemicals are volatile hydrophobic molecules that enter the insects' antennae and other sensory organs via pores and travel across the hydrophobic space to the olfactory receptors (ORs), where the response of the insect to the signal is initiated.

It is currently assumed that the passage of semiochemicals to the ORs is facilitated by odorant-binding proteins (OBPs) which were first reported in Lepidoptera (Vogt and Riddiford, 1981; Tsuchihara *et al.*, 2005) and have since been characterized from a wide range of insect orders including Diptera (e.g., McKenna *et al.*, 1994), Hemiptera (e.g., Dickens *et al.*, 1998), Hymenoptera (e.g., Danty *et al.*, 1998), Orthoptera (e.g., Ban *et al.*, 2003a), Coleoptera (e.g., Wojtasek *et al.*, 1998), Blattaria (Rivière *et al.*, 2003), and Isoptera (Ishida *et al.*, 2002). The transport role for OBPs is suggested by the findings that they are highly concentrated (up to 10 mM) in the lymph of chemosensilla, some are expressed specifically in antennae (Laue *et al.*, 1994; Pikielny *et al.*, 1994; Shanbhag *et al.*, 2001; Ishida *et al.*, 2002), and many can bind pheromones and other semiochemicals (Vogt and Riddiford, 1981; Du *et al.*, 1994; Pelosi and Maida, 1995). Whether these OBPs act only as carriers or confer some specificity is still controversial; however, the idea of specificity is supported by the finding that many insect species have many different genes encoding putative OBPs. For example, studies of genome databases have identified 59 putative OBP genes in *Drosophila melanogaster* (Hekmat-Scafe *et al.*, 2002) and 57 in *Anopheles gambiae* (Xu *et al.*, 2003).

Based on sequence analyses, insect OBPs can be divided into four distinct groups: the "Classic" OBPs with six highly conserved cysteines, the "Plus-C" OBPs with more than six conserved cysteines, the "Atypical" OBPs found only in *A. gambiae* (Hekmat-Scafe *et al.*, 2002; Xu *et al.*, 2003; Zhou *et al.*, 2004), and the chemosensory proteins (CSPs) with only four conserved cysteines (McKenna *et al.*, 1994). The latter group of proteins are slightly smaller than

other OBPs (100–120 residues) and have their four conserved cysteines linked in a noninterlocked fashion producing two small loops (Angeli *et al.*, 1999). The first member of this group was reported as the protein induced in leg regeneration in *Periplaneta americana* (Nomura *et al.*, 1992). A similar protein [olfactory segment-D protein (OS-D)] was demonstrated to be specifically expressed in sensilla coeloconica of *D. melanogaster* (McKenna *et al.*, 1994). They were first named as CSPs by Angeli *et al.* in 1999 and have so far only been reported in insects where they show high levels of amino acid identity across species (Wanner *et al.*, 2004).

For CSPs, there is some evidence of binding to semiochemicals, for example, in *Mamestra brassicae*, there is binding to pheromone components (Jacquin-Joly *et al.*, 2001), a recombinant *Apis mellifera* protein binds to brood pheromone components (Briand *et al.*, 2002) and a protein from *Locusta migratoria* binds an endogenous oleoamide (Ban *et al.*, 2003b). The localization of some CSPs in the lymph of chemosensilla and the binding to semiochemicals strongly suggest that they are in some ways involved in insect chemoreception. However, there is no direct evidence for the involvement of CSPs in olfaction *in vivo*. Indeed, although many are expressed in the antennae, others are expressed in other tissues including legs (Mameli *et al.*, 1996; Picimbon *et al.*, 2001), labial palps (Maleszka and Stange, 1997), tarsi (Angeli *et al.*, 1999), brain (Whitfield *et al.*, 2002), proboscis (Nagnan-Le Meillour *et al.*, 2000), pheromone gland (Jacquin-Joly *et al.*, 2001), and wings (Ban *et al.*, 2003b).

To date, genes encoding CSPs have been found mostly by conventional molecular techniques, and no genome-wide annotation or extensive expressed sequence tags (EST) analyses have been done. Also there is no universal naming system for these proteins, and they have been variously called sensory appendage proteins (SAPs) (Robertson, *et al.*, 1999; Biessmann *et al.*, 2002), OS-D (McKenna *et al.*, 1994), CSPs (Angeli *et al.*, 1999), and OS-D–like proteins (Wanner *et al.*, 2004; Jacobs *et al.*, 2005).

Although an increasing number of CSPs have been discovered in insects (Arthropoda; Uniramia), no study has extended the search to other subphyla of Arthropoda. Indeed, the genome-wide analysis of Arthropods is itself limited by the availability of completed, publicly accessible genome sequences which are only so far available for insect species. However, recent EST projects provide another resource for large-scale analyses (NCBI dbEST), and a large number of EST sequences for other subphyla of Arthropoda are available. In addition, ESTs are available for many insect species with some being assigned to specific tissues.

In this study, we report the genome annotation and EST analysis of CSPs. We have collated 58 previously identified genes and used this to build a CSP sequence motif, and to use such motif to search for protein sequences that contain such a motif in the completed genome sequences of 32 species from six kingdoms, and in the EST sequences of 23 species from the phylum Arthropoda. We have then compared gene expression profiles for the CSP genes within and between species using their presence in the EST libraries to quantify the relative tissue abundance for each gene.

Experimental procedures

Structure of the CSP motif

Published CSP sequences were downloaded from the National Center for Biotechnology Information (NCBI) database using the keywords "CSP," "SAP," and "olfactory-specific protein." A total of 58 sequences were obtained in FASTA format and aligned using Clustal X (Thompson *et al.*, 1997) with default gap penalty parameters of gap opening 10 and extension 0.2. The conserved cysteines were identified, and the number of residues between them was counted (in some cases, manual adjustment was necessary to align the cysteines), giving a "CSP Motif" of C1-$X_{6–8}$-C2-$X_{16–21}$-C3-X_2-C4-X_3. For this motif there are 18 ($3 \times 6 \times 1 \times 1$) possible spacing combinations between the conserved cysteines, and the program "MotifSearch" takes "C1-X_{6-8}-C2-X_{16-21}-C3-X_2-C4-X_3" as the input, expands it to the 18 spacing combinations, and uses each combination to search the peptide sequences for the presence of the CSP motifs. The MotifSearch program is a home-developed window program and written in Delphi.

Identification of motif-containing sequences

The whole-genome peptide sequences of 32 species from six kingdoms (Metazoa, Plantae, Fungi, Protista, Archaea, and Bacteria) were downloaded from their genome databases in the FASTA format and searched for CSP motif–containing sequences using the MotifSearch program. For the EST libraries, the sequences were retrieved as FASTA files from the NCBI EST database (http://www.ncbi.nlm.nih.gov/dbEST/) using either the species or order name to search "text" and then translated in six frames into peptide sequences. The CSP motif–containing peptide sequences were used in Blast searches against the sequences in the NCBI database, and the sequences that had significant matches to known CSPs were selected. The EST id was used to retrieve full-length EST sequences that were used for assembly to get the consensus sequence. Briefly, all the sequences of CSP-encoding ESTs from different tissue-specific EST libraries of one species were combined into a FASTA file as the input file and then assembled with Vector NTI Contig Express software (InforMax, Oxford, UK). The gaps and mismatching bases were corrected manually, and the consensus sequence of each contig was retrieved. For simplicity and to distinguish the retrieved sequences from published CSPs, they were not assigned specific names; instead, they were given four letters to represent the species and then EST followed by the contig number. CSP sequences from genome data were named using a four-letter species identifier and the letters CSP.

Calculation of the relative abundance of each CSP in each tissue type

This was calculated from $N_{\text{CSP-clones}}/N_{\text{total-clones}}$, where $N_{\text{CSP-clones}}$ is the number of EST sequences in that tissue type which are assembled to give the consensus CSP sequence (i.e., the number of ESTs in one contig) and $N_{\text{total-clones}}$ is the total number of EST sequences used in the assembly (i.e., total number of the motif-containing ESTs identified as CSPs from the EST libraries for one species) (see also Supplementary Tables 1–7).

Phylogenetic analyses

The sequences identified from genomic sequences and EST libraries were aligned using Clustal X (8.1) (Thompson *et al.*, 1997) with default gap penalty parameters of gap opening 10 and extension 0.2. The phylogenetic trees were then constructed from these multiple alignments using MEGA3 software (Kumar *et al.*, 2004). The final unrooted consensus tree was generated with 1050 bootstrap trials using the neighbor-joining method (Saitou and Nei, 1987) and presented with a cutoff bootstrapping value of 70.

Results and discussion

Identification of CSPs in genome data

Using the MotifSearch program (see Experimental Procedures) in combination with Blast, we have searched for CSPs in the genome databases of 32 species in the six kingdoms: Archaea, Bacteria, Protista, Fungi, Plantae, and Metazoa. Many CSP motifs were found in representatives of all of the kingdoms, but only in members of the phylum Arthropoda did the motif-containing sequences show similarity to known CSP sequences (Table 1).

We have further analyzed all CSP motif–containing sequences listed in Table 1. Insect sequences were not included in this analysis. Most of these CSP motif–containing proteins are at least twice as long as than insect CSPs. No single CSP hit during the Blast was obtained regardless of search score against insect CSPs using NCBI Blastp with the threshold of EXSPECT = 100 and DISPLAY = 50. However, this does not preclude the possibility that sequence divergence in more distant species prevents CSP detection. These CSP motif–containing sequences were then aligned with *D. melanogaster* CSP OS-D, and the sequences with the highest identity in each species were selected. The identities to OS-D (155 amino acid long) range from 7% of human intestine G-protein signaling regulator (accession no. BAC04934) to a maximum 21% of *Arabidopsis thaliana* transcription factor/zinc ion–binding protein (accession no. NP_200258). Furthermore, the CSP motif–containing sequences with similar size to known CSPs (between 100 to 300 amino acid) include human retinal pigment epithelium (RPE)–spondin (accession no. AAH42877) and keratin-associated proteins (accession no. CAA45283, CAF31639), pig plasmin trypsin inhibitor (accession no. NP_999036), as well as proteins that contain ring finger and CHY zinc finger domain, such as *Danio rerio* zinc finger protein 658 (accession no. XP_688998), human transcription elongation factor A and zinc finger protein (accession no. XP_528356 and AAL09356), pig transcription elongation factor A and zinc finger protein (accession no. XP_528356 and XP_517222), and *A. thaliana* transcription factor and ubiquitin–protein ligase (accession no. NP_200258, NP_194461 and NP_177614), and proteins that have a rabphilin-3A effector domain, such as *Mus musculus* Rab3-interacting molecule 2 (accession no. BAD32696) and *Caenorhabditis elegans* Rim (accession no. BAD32699). All bacterial proteins that contain CSP motifs are involved in energy production and conversion such as polyferredoxin and anaerobic dimethyl sulfoxide reductase chain B of *Escherichia coli* (accession no. BAB36804 and BAB36806, respectively). It was found that to each of all smaller CSP motif–containing proteins, there are Drosophila counterparts annotated with putative functions rather than chemoreception when these proteins were used in Blast against *D. melanogaster* sequences. For example, all plant proteins are similar to Drosophila zinc finger protein (accession no. AAF47094) and all *E. coli* proteins to Drosophila protein containing 4Fe-4S–binding domain (accession no. AAF50342). The mammal proteins (BAC04934, NP_999036, XP_5172222, and BAD32696) match to Drosophila fatty acid desaturase, extracellular matrix protein, zinc finger protein, and Rim (accession no. AAB17283, AAO84908, AAF52385, and AAF55479, respectively) with E values of 0.03, 9×10^{-12}, 3×10^{-76}, and 9×10^{-30}, respectively. All these proteins also have no similarity to any of the known CSPs in both primary sequences and predicted tertiary structures (data not shown), suggesting that CSPs identified in the current study seem to be confined to the phylum Arthropoda.

Within the phylum Arthropoda, only insect (Uniramia) genomes are available, so it does not tell us how widespread these genes are in other subphyla of Arthropoda (see later). Indeed, the only available genomes are from insect species, that is, *D. melanogaster*, *Drosophila pseudoobscura*, *A. gambiae*, *A. mellifera*, and *Bombyx mori*, and from these 18,289, 18,331, 15,802, 24,640, and 21,302 peptide sequences were, respectively, searched giving a total of 102 motif-containing sequences of which 28 also showed similarity to known CSPs (Table 1). An alignment of these along with sequences identified from EST libraries (see later) is supplied as Supplementary Figure 1.

For *D. melanogaster*, four of the 24 motif-containing sequences showed similarity to CSPs (Table 1) and are presented as DmelCSP1-4. DmelCSP1 has not been annotated previously but is identical to CG30172 in the *D. melanogaster* genome annotation (FlyBase) and to *D. melanogaster* head ESTs (see Supplementary Table 1). DmelCSP4 was identified previously as a *D. melanogaster* OS-D (accession no. Q27377) (McKenna *et al.*, 1994) or A10 (Pikielny *et al.*, 1994). The deduced amino acid sequence of DmelCSP2 is

Table 1 Identification of putative CSPs from annotated genome sequences in the six kingdoms

Kingdom	Phylum	Species	No. sequence[a]	No. motif[b]	No. CSP[c]	Life form
Metazoa	Chordata	*Homo sapiens*	34,091	54	0	Human
	Chordata	*Pan troglodytes*	38,822	39	0	Chimpanzee
	Chordata	*Mus musculus*	32,281	44	0	Mouse
	Chordata	*Rattus norvegicus*	28,545	96	0	Rat
	Chordata	*Gallus gallus*	28,416	43	0	Chicken
	Chordata	*Danio rerio*	30,783	59	0	Zebrafish
	Arthropoda	*Anopheles gambiae*	15,802	20	7	Malaria mosquito
	Arthropoda	*Drosophila melanogaster*	18,289	24	4	Fruit fly
	Arthropoda	*Drosophila pseudoobscura*	18,331	20	4	Fly
	Arthropoda	*Bombyx mori*	21,302	21	8	Silk moth
	Arthropoda	*Apis mellifera*	24,640	17	5	Honeybee
	Nematoda	*Caenorhabditis elegans*	22,215	12	0	Nematodes
	Nematoda	*Caenorhabditis briggsae*	14,713	9	0	Nematodes
Plantae	Streptophyta	*Oryza sativa*	59,712	19	0	Rice
	Streptophyta	*Arabidopsis thaliana*	28,581	10	0	Thale cress
Fungi	Microsporidia	*Encephalitozoon cuniculi*	1996	5	0	Parasite
	Ascomycota	*Neurospora crassa*	10,082	5	0	Bread mold
	Ascomycota	*Fusarium graminearum*	11,640	4	0	Pathogen
	Ascomycota	*Eremothecium gossypii*	4718	4	0	Cotton pathogen
	Ascomycota	*Schizosaccharomyces pombe*	4964	3	0	Fission yeast
	Ascomycota	*Saccharomyces cerevisiae*	6700	1	0	Baker's yeast
Protista	Apicomplexa	*Plasmodium falciparum*	5267	1	0	Malaria parasite
Archaea	Euryarchaeota	*Methanococcus maripaludis*	1722	2	0	Methanogen
	Euryarchaeota	*Methanococcus jannaschii*	1715	1	0	Methanogen
	Euryarchaeota	*Methanosarcina acetivorans*	4540	1	0	Methanogen
	Euryarchaeota	*Methanosarcina mazei*	3371	1	0	Methanogen
	Euryarchaeota	*Methanobacterium thermoautotrophicum*	1869	1	0	Methanogen
Bacteria	Proteobacteria	*Escherichia coli O157*	5361	2	0	Bacterium
	Proteobacteria	*Escherichia coli CFT073*	5379	1	0	Bacterium
	Proteobacteria	*Geobacter sulfurreducens PCA*	3446	1	0	Bacterium
	Proteobacteria	*Desulfotalea psychrophila LSv54*	3118	2	0	Bacterium
	Proteobacteria	*Legionella pneumophila str. Lens*	2878	1	0	Pathogen

[a]The peptide sequences of *C. elegans*, *H. sapiens*, *M. musculus*, *R. norvegicus*, *G. gallus*, *S. cerevisiae*, and *E. coli* were from ftp://ftp.ebi.ac.uk/pub/databases/genomes/. The peptide sequences of *O. sativa* and *A. thaliana* were downloaded from ftp://ftp.tigr.org/pub/data/Eukaryotic_Projects/. The *Drosophila* sequences were downloaded from FlyBase http://flybase.bio.indiana.edu/annot/. *Apis mellifera* and *A. gambiae* peptide sequences were from http://www.ensembl.org/info/data/download.html. *Bombyx mori* peptide sequences were from http://silkworm.genomics.org.cn/. All other sequences were downloaded from NCBI database as FASTA files.

[b]The number of CSP motif–containing sequences was identified by MotifSearch program (see text).

[c]The number of peptide sequences in the motif-containing sequences that have a similarity with known CSPs in NCBI database.

identical to a virus-induced protein Pherokine-2 (Phk-2) (Sabatier *et al.*, 2003) and has a very high identity (63.2%) to the ejaculatory bulb protein III (PEBmelIII) (Accession no. Q23981). The annotation of PEBmelIII as a possible CSP is interesting because this protein belongs to a family that is thought to be responsible for changes in the activity of *D. melanogaster* (in terms of movement, secretion, enzyme production, gene expression, etc.) or in responses to a virus stimulus (Brieger and Butterworth, 1970; Sabatier *et al.*, 2003). A possible involvement of PEBmelIII in storing and delivering chemical messages has also been proposed on the basis of its expression in glands producing pheromones (Dyanov and Dzitoeva, 1995), and the high similarity of PEBmelIII to the OBP MbraOBP2 of *M. brassicae* has led to the suggestion that vaccenyl acetate could be a specific ligand for both proteins (Bohbot *et al.*, 1998). DmelCSP3 resembles *D. melanogaster* Pherokine-3 (Phk-3) that was reported to be induced by virus and bacterial infections (Sabatier *et al.*, 2003). For *D. pseudoobscura*, all four genome sequences identified as CSPs (from a total of 20 sequences with a CSP motif, Table 1) are homologues of the four *D. melanogaster* CSPs, with more than 75% amino acid identity, implying a very high conservation of CSPs in the genus *Drosophila*.

For the *A. gambiae* published genome annotation, the predicted peptide sequences are given as the longest open reading frames, most of them being without identified start codons (R. Holt, personal communication), so we used GenScan full-length coding sequences for each of the 20 CSP motif–containing sequences found by our MotifSearch. This showed that of the 20 motif-containing sequences, seven have significant similarity to known insect CSPs (Table 1), and these were compared with entries in GenBank and the mosquito Ensembl database and with three independent genome-wide annotations of *A. gambiae* OBPs (Vogt, 2002; Xu *et al.*, 2003; Zhou *et al.*, 2004). It was apparent that some of the annotated genes are from the same contig sequence and have almost identical sequences despite being given different names and accession numbers. One of our sequences, AgamCSP4, has been cloned previously and named as the SAP1, AgamSAP-1 (Biessmann *et al.*, 2002), but the other six are newly identified as CSPs.

The 17 motif-containing sequences from the genome of the honeybee *A. mellifera* (Table 1) included five with similarity to published CSPs with one, AmelCSP4, being identical to a previously reported honeybee CSP (ASP3c) (Danty *et al.*, 1998; Briand *et al.*, 2002) and another, AmelCSP2, being identical to the worker antennal protein 1 (Kamikouchi *et al.*, 2004). We also identified eight CSP genes in the *B. mori* genome, including two previously cloned CSPs, BmorCSP1 and BmorCSP2 (Picimbon, *et al.*, 2000).

Identification of CSPs in EST libraries

As stated previously, within the phylum Arthropoda, the only organisms with genome sequences are insects in the subphylum Uniramia. We therefore used EST data to look for a wider distribution of CSPs in other subphyla of Arthropoda (Table 2). There are 96,276 EST sequences available for the subphylum Chelicerata, but although many were found to contain CSP motifs, there were no sequences with a significant similarity to any known CSPs. However, from 63,304 EST sequences of the subphylum Crustacea, 15 had CSP motifs including one (BQ563140) in the brine shrimp *Artemia franciscana* which had significant similarity to CSPs (24% identity to the *Cactoblastis cactorum* CSP CLP-1, 22% identity to the CSP-I of *Schistocerca gregaria*, and 18% identity to Mbra-SAP of *M. brassicae*). Thus, the origin of CSPs appears to predate the divergence of Crustacea and Uniramia, about 130 million years earlier than previously reported from the separation of the class Neoptera (Wanner *et al.*, 2004).

For the subphylum Uniramia, CSP motif–containing sequences were identified in 25 different species, and all of these had one or more sequences with significant similarity to known CSPs (Table 2) (Each CSP sequence was obtained by assembling large numbers of ESTs, and the information is provided in the Supplementary Tables 1–7.). For the species with genome sequences available, we were able to compare the genome sequence with the EST (see next section), but we also found CSPs in other insect orders for which the genome data are not available and some of which have not previously been found to have CSPs. Thus in the order Hemiptera, there are five CSPs in the pea aphid *Acyrthosiphon pisum* and four in the brown citrus aphid *Toxoptera citricida*. In Siphonaptera there is one CSP in the cat flea *Ctenocephalides felis*, in Phthiraptera one CSP in the human head-louse *Pediculus humanus*, in Coleoptera one from *Biphyllus lunatus* and two from *Tribolium castaneum*, and for Orthoptera a new group of CSPs with high similarity to CSPSgre-III-1 in the locust *L. migratoria*. For Lepidoptera where CSPs have been widely reported previously, we found additional six CSPs in the moth *Manduca sexta* and 10 in the silk moth *B. mori* of which five were found in both genome sequences and EST libraries, one only in genome sequences, and four only in EST libraries. We also found CSPs in species where they had not been reported previously, two in the butterfly *Heliconius melpomene*, and four in the diamondback moth *Plutella xylostella*. These data show that CSPs are extremely widespread throughout the insect species.

Comparison of CSP sequences in different insect species

An alignment of the deduced CSP sequences identified from genome data and EST libraries in this study is supplied as Supplementary Figure 1, and the data have been used to construct phylogenetic trees and to compare sequences for CSPs as given in Figure 1.

For species in the order Diptera, the sequence analysis of CSPs identified from *D. melanogaster*, *D. pseudoobscura*, *Drosophila yakuba*, *A. gambiae*, and *Glossina morsitans*, as well as reported SAPs of *Aedes aegypti* and *Culicoides sonorensis* (Figure 1A) shows that many CSPs within each species

Table 2 Identification of CSP from ESTs and their tissue expression in the phylum Arthropoda

Subphylum	Order	Species	No. EST	No. Motif[a]	No. CSP[b]	Tissue type
Chelicerata	Ixodida	*mix*	96,276	84	0	Mix
Crustacea	Decapoda	*mix*	63,304	14	0	Mix
Crustacea	Anostraca	*Artemia franciscana*	1746	1	1	Whole organism
Uniramia	Odonata	*mix*	6223[c]	0	NA	Mix
	Blattaria	*mix*	6842[c]	2	2[d]	Leg
	Phasmatodea	*mix*	10,629[c]	3	3[d]	Antennae
	Orthoptera	*Locusta migratoria*	45,481	88	19	See Figure 2
	Hemiptera	*Acyrthosiphon pisum*	3707	15	5	Whole nymphs and female adults
	Hemiptera	*Acyrthosiphon pisum*	13,055	14	3	Female nymphs (L3) heads
	Hemiptera	*Acyrthosiphon pisum*	6380	2	2	Nymphs (L3) antennae
	Hemiptera	*Acyrthosiphon pisum*	5467	0	NA	Embryos
	Hemiptera	*Toxoptera citrida*	4304	11	4	Whole adults
	Coleoptera	*Biphyllus lunatus*	671	2	1	Whole organism
	Coleoptera	*Tribolium castaneum*	2475	41	2	Embryos
	Diptera	*Anopheles gambiae*	21,916	21	3	Head
	Diptera	*Anopheles gambiae*	10,346	55	1	Fat body
	Diptera	*Drosophila melanogaster*	84,414	226	2	Heads
	Diptera	*Drosophila melanogaster*	146	0	NA	Antennae
	Diptera	*Drosophila melanogaster*	28,326	15	1	Larval fat body
	Diptera	*Drosophila melanogaster*	16,215	17	1	Pupae
	Diptera	*Drosophila melanogaster*	102,141	31	1	Embryos
	Diptera	*Drosophila pseudoobscura*	727	30	0	Embryos
	Diptera	*Drosophila yakuba*	10,236	1	1	Testes
	Diptera	*Drosophila yakuba*	441	2	1	Embryos
	Diptera	*Drosophila yakuba*	373	1	1	Whole adults
	Diptera	*Drosophila simulans*	4121	2	0	Testes
	Diptera	*Glossina morsitans*	21,463	8	1	Adult gut
	Siphonaptera	*Ctenocephalides felis*	4846	1	1	Hindgut and Malpighian tubules
	Trichoptera	*Hydropsyche sp. T20*	222	1	0	Larval silk glands
	Lepidoptera	*Mamestra brassicae*	71[c]	29	6[d]	Antennae
	Lepidoptera	*Manduca sexta*	1354	25	11	Male antennae
	Lepidoptera	*Manduca sexta*	400	43	5	Female antennae
	Lepidoptera	*Manduca sexta*	260	0	NA	Larval fat body
	Lepidoptera	*Manduca sexta*	12	0	NA	Larval nerve cord and adult wings
	Lepidoptera	*Bombyx mori*	118,426	121	10	See Figure 2
	Lepidoptera	*Cactoblastis cactorum*	1[c]	1	1[d]	Labial palps
	Lepidoptera	*Plutella xylostella*	1144	5	4	Larvae
	Lepidoptera	*Heliconius melpomene*	568	2	2	Pupae
	Hymenoptera	*Apis mellifera*	202	5	2	Female antennae

Table 2 Continued

Subphylum	Order	Species	No. EST	No. Motif[a]	No. CSP[b]	Tissue type
	Hymenoptera	*Apis mellifera*	227	0	NA	Male antennae
	Hymenoptera	*Apis mellifera*	19,037	7	1	Female worker brain
	Hymenoptera	*Apis mellifera*	5208	6	0	Body (larval, pupal, and female)
	Hymenoptera	*Apis mellifera*	22	0	NA	Ovaries
	Hymenoptera	*Linepithema humile*	66	1	1	Antennae
	Phthiraptera	*Pediculus humanus*	1127	1	1	Whole organism (engorged adult)

[a]The number of CSP motif–containing sequences was identified by MotifSearch program (see text).
[b]The number of peptide sequences in the motif-containing sequences that have a similarity with known CSPs in NCBI database.
[c]There is no EST data available and the protein sequences were obtained from NCBI database.
[d]Published CSP sequences that were also identified with MotifSearch program. The ESTs of *D. yakuba* were from ftp://hgdownload.cse.ucsc.edu/goldenPath/droYak1/database/, and the ESTs of *D. pseudoobscura* were downloaded from ftp://ftp.hgsc.bcm.tmc.edu/pub/data/Dpseudoobscura/EST/.

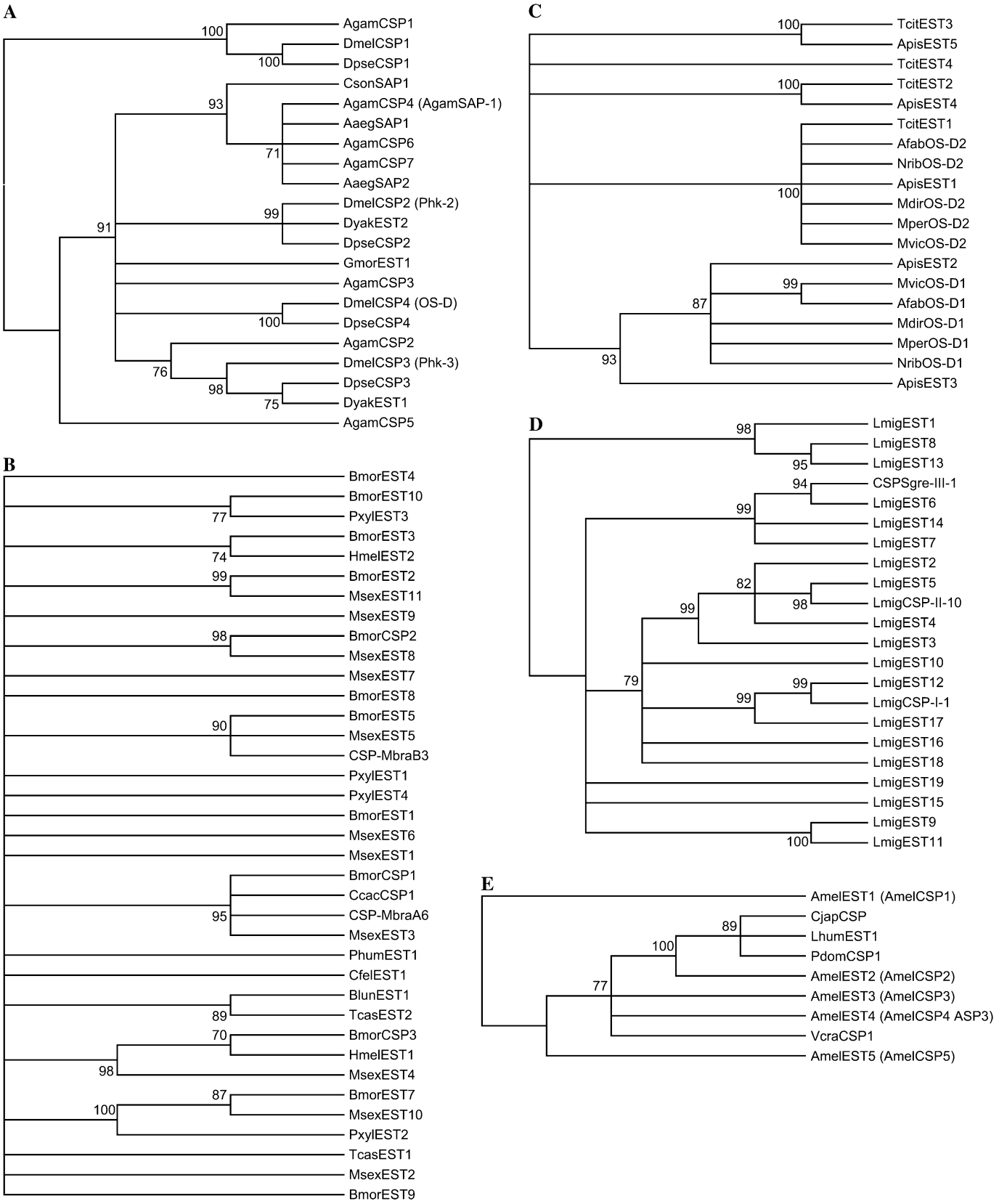
are similar (e.g. AgamCSP4, 6, and 7) and that there are highly conserved genes between the three *Drosophila* species (e.g. DpseCSP2, DyakEST2, and DmelCSP2). Perhaps more surprisingly, some *Drosophila* and *Anopheles* sequences are very similar, for example, a group of three CSPs, DmelCSP1, DpseCSP1, and AgamCSP1, which are well segregated from other CSPs with a bootstrap value of 100. This suggests that they are a group of CSPs with a common ancestral protein and which have stayed the same because they perform similar vital functions in all three species. Despite some closely related groups of CSPs, there is clearly a diversity of sequences within these species, and all of this is consistent with the view that this family of genes arose by duplication events prior to the separation of the species and then diverged in each species. This would also seem to be true in the Lepidoptera where the CSPs from six species (*B. mori*, *M. sexta*, *P. xylostella*, *H. melpomene*, *M. brassicae*, and *C. cactorum*) are compared. For sequence comparison, two previously reported CSPs (CSP-MbraA6, CSP-MbraB3) of *M. brassicae* and CSPs identified in this study from Coleoptera, Siphonaptera, and Phithiraptera (Table 2) were also included. Figure 1B shows that moth CSPs, although the tree does not include all previously reported Lepidoptera CSPs, like those of Diptera species, form groups across species with a CSP from one species often being more similar to a CSP from a different species than to others of its own. In *B. mori*, eight of 12 CSPs have their homologues in other moth species, only four (BmorEST1, BmorEST4, BmorEST8, and BmorEST9) are separated from these clusters. However, their homologues would be found when more genome sequences of other moth species become available. Two *B. mori* CSPs (BmorCSP2 and BmorCSP3) were only found in the genome sequence, but their homologues (MsexEST8 and HmelEST1, respectively) were found in the EST libraries of other moth species (Figure 2B). MsexEST8 was identified from the male antennal ESTs of *M. sexta*, and HmelEST1 was found in the pupal EST library of *Heliconius melpomene* (Table 2).

For Hemiptera, we have found CSPs in seven species, all of which are aphids (*A. pisum*, *T. citridae*, *Aphis fabae*, *Nasanovia ribis-nigri*, *Metapolophium dirhodum*, *Myzus persicae*, and *Megoura viciae*). There is a clear conservation of sequences; indeed, one sequence is almost identical in all seven species (Figure 1C and consistent with Jacobs *et al.*, 2005). This is also true for the order Orthoptera where we only have EST libraries from one species *L. migratoria* and where all CSPs are very closely related, presumably because they have arisen from a common gene fairly recently (Figure 1D). Furthermore, three homologues of recently reported CSP CSPSgre-III-1 in *S. gregaria* were identified from *L. migratoria* ESTs (Figure 1D).

For Hymenoptera, EST libraries of two species (*A. mellifera* and *Linepithema humile*) were analyzed (Table 2). However, only three of five *A. mellifera* CSPs identified in the genome were found in the EST libraries. For sequence comparison, previously reported CSPs of other Hymenoptera species (*Polistes dominulus*, *Vespa crabro*, *Camponotus japonicus*) were included in the phylogenetic analysis. As shown in Figure 1E, apart from AmelEST1 and AmelEST5 there are two clusters, one of them contains CSPs from different species (AmelEST2, PdomCSP1, LhumEST1, CjapCSP), suggesting this CSP may perform a similar role in these insects. However, the wasp CSP VcraCSP1 was similar to other two honeybee CSPs AmelEST2 and AmelEST3 in another cluster.

Expression of CSP transcripts

Data from EST libraries not only allow us to identify sequences likely to be CSPs but also, when the origin of the tissues used to construct the libraries is known, we can deduce the relative expression of the CSP transcripts (see Experimental Procedure). This is shown for seven insect species in Table 2 and Figure 2. However, some of the libraries available in NCBI dbEST do not reflect the "natural" expression level of genes, according to the way the EST



libraries have been constructed: some libraries are normalized, some are subtracted, or induced; that could interfere with the number of clones encoding one CSP (enriched or lowered). The data presented in Figure 2A–H were obtained from normalized EST libraries from *A. gambiae* and *D. melanogaster* and natural expressed EST libraries, and no subtracted and induced EST libraries were used to analyze relative expression level of CSPs. The data only represent the relative expression level of one CSP to other CSPs in one tissue-specific EST library of one insect species and is presented individually for each species in Figure 2A–H.

In the 21,916 normalized head ESTs and 10,284 normalized fat body ESTs of *A. gambiae*, there are seven CSPs, and the CSP expressed most frequently is *AgamCSP4* which is expressed equally in both heads with antennae and in female fat body (Figure 2A). In contrast, in *D. melanogaster* (89,576 normalized head ESTs), one CSP is expressed most frequently (*DmelCSP2*) and here this is exclusively in heads with antennae which is more supportive of a possible role in *Drosophila* olfaction (Figure 2B). Phylogenetic analysis shows this CSP to be specific to the *Drosophila* genus (Figure 1A). The *Drosophila* CSP *DmelCSP3* is expressed at low levels in all developmental stages including embryo (102,375 normalized ESTs), pupae (16,299 ESTs), and larvae (10,449 ESTs) (Figure 2B), and closely related genes are found in the embryo EST library of *D. yakuba* and in the genome of *D. pseudoobscura* (Figure 1A). Furthermore, *DmelCSP3* has an identical sequence to Pherokine-3 (*Phk-3*), indicating its expression may be regulated normally at a very low level but that it is induced by viral and bacterial infections (Sabatier *et al.*, 2003). The *Drosophila* CSP *DmelCSP4* has an identical sequence to *OS-D* (*A10*) that has been shown to be an antennal-specific gene; however, we did not find it in the antennal EST libraries although this may be because of the relatively small numbers of ESTs available in the current database. But its homologue DpseCSP4 was identified in the genome sequence of *D. pseudoobscura* (Figure 1A).

In the honeybee, there are EST libraries, 202 from female antennae and 19,051 from female brain. Two CSPs *AmelCSP2* and *AmelCSP4* are present in female antennae with *AmelCSP4* also being present in the female brain at a relatively high level (Figure 2C). The peptide sequence of *AmelCSP4* is identical to a reported antennal-specific protein 3c that binds to brood pheromone components of *A. mellifera* (Briand *et al.*, 2002), and *AmelCSP2* was found to be expressed in the worker antennae (Kamikouchi *et al.*, 2004).

For *B. mori* there are no EST libraries from heads or antennae, so our study for this species is limited to nonolfactory tissues (Figure 2D). Of the 10 CSPs identified, *BmorEST1* is the most frequently expressed, mostly in the pupal and adult pheromone gland (8432 ESTs). It may be involved in the production and/or transportation and release of pheromones. Five out of the 10 *B. mori* CSPs (*BmorEST3*, *5*, *6*, *7*, and *8*) are expressed in the compound eyes, the first time that any CSPs have been reported in this tissue, and the functional role of these genes in eyes could not be olfaction. *BmorEST6* is identical to the reported antennal CSP *BmorCSP1*. Two CSPs identified from the *B. mori* genome (*BmorCSP3* and *5*) and the previously reported *BmorCSP2* were not found in the current EST libraries (a total of 65,389 ESTs but lacking antennae-specific ESTs). In fact, *BmorCSP5* is identical to *BmorCSP2*. For another Lepidopteran species, *M. sexta*, our search found 11 CSP genes (from 1994 ESTs) of which only five have been reported previously from EST databases (*SAP1*, *SAP2*, *SAP3*, *SAP4*, and *SAP5*; Robertson *et al.*, 1999). All 11 are expressed in antennae either in males, females, or both, suggesting a role in olfaction. However, the possibility of the expression in other parts of the insect can not be excluded from the current EST data.

Within the order Hemiptera, ESTs are available for only two species, both of which are aphids. For *T. citricida*, four CSP genes were identified in 2475 ESTs, but the library was constructed from whole insects, so it tells us nothing about expression patterns (Table 2). For *A. pisum*, all five CSP genes found are expressed in whole insects, and for three there is evidence of antennal expression (*ApisEST2* and *4*; Figure 2F) again supportive of a possible role in olfaction. It is interesting that extensive Blast searching and attempts at homology cloning in our laboratory have failed to identify any members of other OBP classes in aphids (Jacobs *et al.*, 2005). The similar expression patterns were reported in the social wasps *P. dominulus* and *V. crabro* where CSPs are specifically expressed in the antennae, while OBPs are also expressed in legs and wings (Calvello *et al.*, 2003).

So far the species with the most CSPs identified is the locust *L. migratoria*, where from 45,481 ESTs, 75 have CSP motifs

Figure 1 Phylogenetic tree of the deduced protein sequences of the CSPs identified from genome data and EST libraries in this study as well as some reported CSPs in insects belonging to the orders **(A)** Diptera (*Drosophila melanogaster*, *Drosophila yakuba*, *Drosophila pseudoobscura*, *Anopheles gambiae*, *Glossina morsitans*, *Aedes aegypti*, and *Culicoides sonorensis*) and **(B)** Lepidoptera (*Bombyx mori*, *Mamestra brassicae*, *Manduca sexta*, *Plutella xylostella*, and *Cactoblastis cactorum*). CSPs identified in the current study from Coleoptera (*Biphyllus lunatus*, *Tribolium castaneum*), Siphonaptera (*Ctenocephalides felis*), and Phthiraptera (*Pediculus humanus*) are also included. **(C)** Hemiptera (*Acyrthosiphon pisum*, *Toxoptera citrida*, *Aphis fabae*, *Nasanovia ribis-nigri*, *Metapolophium dirhodum*, *Myzus persicae*, and *Megoura viciae*). **(D)** Orthoptera (*Locusta migratoria*). **(E)** Hymenoptera (*Apis mellifera*, *Linephithema humile*, *Polistes dominulus*, *Vespa crabro*, and *Camponotus japonicus*). Consensus unrooted trees were generated with 1050 bootstrap trials using the neighbor-joining method (Saitou and Nei, 1987; Kumar *et al.*, 2004) and presented with a cutoff value of 70. Sequences named as "CSP" have either been reported previously as CSPs or were identified in this study from genome sequences. Sequences named as "ESTs" were identified from EST libraries. The names of any identical sequences in the NCBI database are included in parenthesis. The original names of the published CSPs are used for the species if they are not identified in the current study.

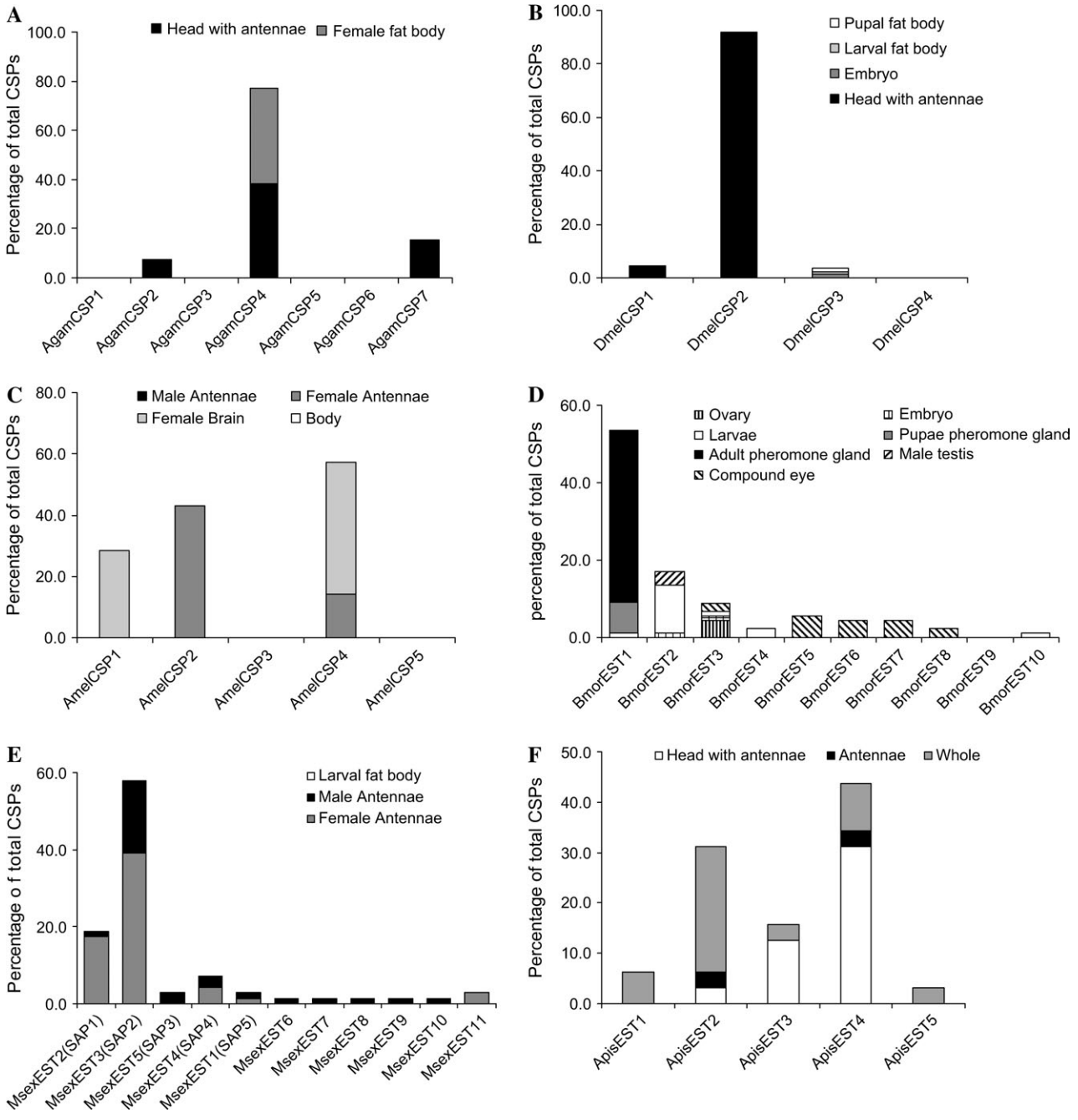


Figure 2 Relative abundance of CSP transcripts in EST libraries. For each CSP transcript in a species, the number of ESTs assembled for each CSP is given as a percentage of all CSP ESTs in that species (see Experimental Procedures and Supplementary Tables). The colors indicate the origin of the tissues used to construct the EST libraries in which the CSPs were identified. **(A)** *Anopheles gambiae*, **(B)** *Drosophila melanogaster*, **(C)** *Apis melifera*, **(D)** *Bombyx mori*, **(E)** *Manduca sexta*, **(F)** *Acyrthosiphon pisum*, and **(G, H)** *Locusta migratoria*.

and 19 have significant similarity with other CSPs (Table 2). For this species there are two recognized forms, the so-called "solitary" phase, where individuals live alone, and the "gregarious" phase, where they congregate and form swarms, and it is interesting that expression of many of the CSP transcripts is very different in the solitary (a total of 15,889 ESTs) and gregarious (a total of 17,175 ESTs) phases (Figure 2G and H). For example, the expression of *LmigEST8* in heads is much higher in the solitary phase, whereas *LmigEST12* is not expressed in the solitary phase but is at a high level in the gregarious phase. Most of the *L. migratoria* CSPs are expressed in heads (0.65% of 33,064 total ESTs), and the changes in expression during the developmental switch from solitary to gregarious is an

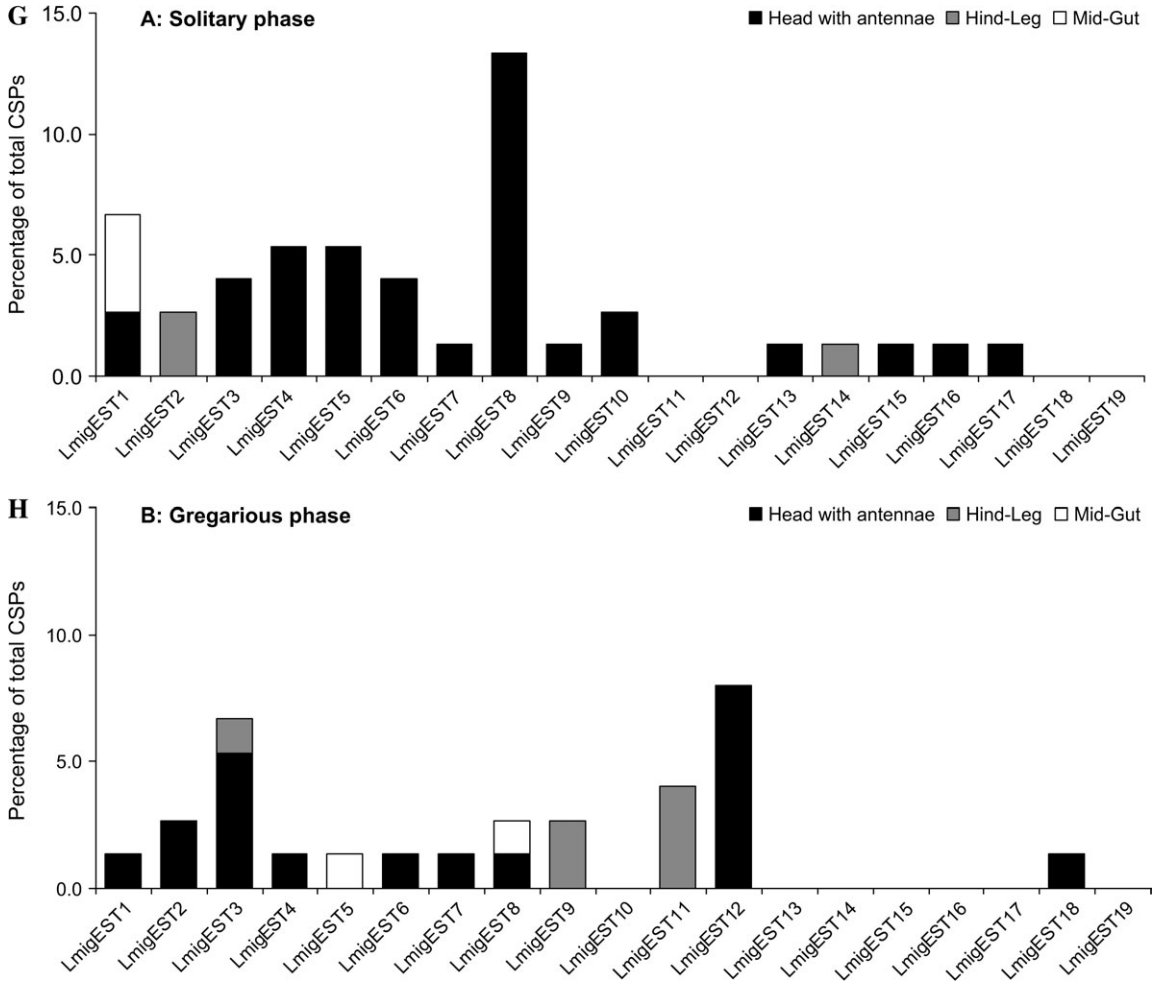


Figure 2 Continued.

interesting finding. Further experiments are needed to confirm the possible role of these CSPs in this phenomenon.

Overall, we have found that CSPs in *D. melanogaster*, *A. gambiae*, *A. mellifera*, *M. sexta*, *A. pisum*, and *L. migratoria* are expressed in antennae, or head material containing antennae, and could therefore possibly have a role in olfaction. It is also likely that in *B. mori*, one CSP may play roles in pheromone release and some in eye function, and in *L. migratoria*, some CSPs may be involved in the developmental transition between the solitary and gregarious life stages.

Conclusions

In this study, we have for the first time reported a CSP in a Crustacean species, suggesting that the origin of CSPs predates the divergence of Uniramia and Crustacea. We have found CSPs in many insect orders including for the first time in Siphonaptera and Phthiraptera. This demonstrates a wide occurrence of these proteins, and their high level of conservation suggests an important role in insect function. Furthermore, the presence of CSPs in different tissue-specific EST libraries has demonstrated a wide distribution for CSPs within each insect species, and they may involve in olfaction as well as in pheromone release and eye function. They are clearly an important family of insect proteins and many questions on their functions remain to be answered.

Supplementary material

Supplementary Figure 1 and Tables 1–7 can be found at http://www.chemse.oxfordjournals.org.

Acknowledgements

Rothamsted Research receives grant-aided support from the Biotechnology and Biological Sciences Research Council of the United Kingdom.

References

Angeli, S., Ceron, F., Scaloni, A., Monti, M., Monteforti, G., Minnocci, A., Petacchi, R. and **Pelosi, P.** (1999) *Purification, structural characterization, cloning and immunocytochemical localization of chemoreception proteins from* Schistocerca gregaria. Eur. J. Biochem., 262, 745–754.

Ban, L., Scaloni, A., D'Ambrosio, C., Zhang, L., Yahn, Y. and **Pelosi, P.** (2003a) *Biochemical characterization and bacterial expression of an odorant-binding protein from* Locusta migratoria*.* Cell. Mol. Life Sci. 60, 390–400.

Ban, L., Scaloni, A., Brandazza, A., Angeli, S., Zhang, L., Yan, Y. and **Pelosi, P.** (2003b) *Chemosensory proteins of* Locusta migratoria*.* Insect Mol. Biol., 12, 125–134.

Biessmann, H., Walter, M.F., Dimitratos, S. and **Woods, D.** (2002) *Isolation of cDNA clones encoding putative odourant binding proteins from the antennae of the malaria-transmitting mosquito,* An. gambiae*.* Insect Mol. Biol., 11, 123–132.

Bohbot, J., Sobrio, F., Lucas, P. and **Nagnan-Le Meillour, P.** (1998) *Functional characterization of a new class of odourant-binding proteins in the moth* Mamestra brassicae*.* Biochem. Biophys. Res. Commun., 253, 489–494.

Briand, L., Swasdipan, N., Nespoulous, C., Bezirard, V., Blon, F., Huet, J.C., Ebert, P. and **Pernollet, J.C.** (2002) *Characterization of a chemosensory protein (ASP3c) from honeybee (*Apis mellifera *L.) as a brood pheromone carrier.* Eur. J. Biochem., 269, 4586–4596.

Brieger, G. and **Butterworth, F.M.** (1970) Drosophila melanogaster*: identity of male lipid in reproductive system.* Science, 167, 1262.

Calvello, M., Guerra, N., Brandazza, A., D'Ambrosio, C.D., Scaloni, A., Dani, F.R., Turillazzi, S. and **Pelosi, P.** (2003) *Soluble proteins of chemical communication in the social wasp* Polistes dominulus*.* Cell. Mol. Life Sci., 60, 1933–1943.

Danty, E., Arnold, G., Huet, J.C., Huet, D., Masson, C. and **Pernollet, J.C.** (1998) *Separation, characterization and sexual heterogeneity of multiple putative odourant-binding proteins in the honeybee* Apis mellifera *L. (Hymenoptera: Apidea).* Chem. Senses, 23, 83–91.

Dickens, J.C., Callahan, F.E., Wergin, W.P., Murphy, C.A. and **Vogt, R.G.** (1998) *Intergeneric distribution and immunolocalization of a putative odourant-binding protein in true bugs (Hemiptera, Heteroptera).* J. Exp. Biol., 201, 33–41.

Du, G.H., Ng, C.S. and **Prestwich, G.D.** (1994) *Odourant binding by a pheromone binding protein: active site mapping by photoaffinity labeling.* Biochemistry, 33, 4812–4819.

Dyanov, H.M. and **Dzitoeva, S.G.** (1995) *Method for attachment of microscopic preparations on glass for in situ hybridisation, PRINS and in situ PCR techniques.* Biotechniques, 18, 822–824.

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

Ishida, Y., Chiang, V.P., Haverty, M.I. and **Leal, W.S.** (2002) *Odorant-binding proteins from a primitive termite.* J. Chem. Ecol., 28, 1887–1893.

Jacobs, S.P., Liggins, A.P., Zhou, J.J., Pickett, J.A., Jin, X. and **Field, L.M.** (2005) *OS-D-like genes and their expression in aphids (Hemiptera: Aphididae).* Insect Mol. Biol., 14, 423–432.

Jacquin-Joly, E., Vogt, R.G., Francois, M.C. and **Nagnan-Le Meillour, P.** (2001) *Functional and expression pattern analysis of chemosensory proteins expressed in antennae and pheromonal gland of* Mamestra brassicae*.* Chem. Senses, 26, 833–844.

Kamikouchi, A., Morioka, M. and **Kubo, T.** (2004) *Identification of honeybee antennal proteins/genes expressed in a sex- and/or caste selective manner.* Zool. Sci., 21, 53–62.

Kumar, S., Tamura, K., and **Nei, M.** (2004) *MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment.* Brief. Bioinform., 5, 150–163.

Laue, M., Steinbrecht, R.A. and **Ziegelberger, G.** (1994) *Immunocytochemical localization of general odourant-binding protein in olfactory sensilla of the silkmoth* Antheraea polyphemus*.* Naturwissenschaften, 81, 178–180.

Maleszka, R. and **Stange, G.** (1997) *Molecular cloning, by a novel approach, of a cDNA encoding a putative olfactory protein in the labial palps of the moth* Cactoblastis cactorum*.* Gene, 202, 39–43.

Mameli, M., Tuccini, A., Mazza, M., Petacchi, R. and **Pelosi, P.** (1996) *Soluble proteins in chemosensory organs of Phasmids.* Insect Biochem. Mol. Biol., 26, 875–882.

McKenna, M.P., Hekmat-Scafe, D.S., Gaines, P. and **Carlson, J.R.** (1994) *Putative* D. melanogaster *pheromone-binding proteins expressed in a subregion of the olfactory system.* J. Biol. Chem., 269, 16340–16347.

Nagnan-Le Meillour, P., Cain, A.H., Jacquin-Joly, E., Francois, M.C., Ramachandran, S., Maida, R. and **Steinbrecht, R.A.** (2000) *Chemosensory proteins from the proboscis of* Mamestra brassicae*.* Chem. Senses, 25, 541–553.

Nomura, A., Kawasaki, K., Kubo, T. and **Natori, S.** (1992) *Purification and localization of p10, a novel protein that increases in nymphal regenerating legs of* Periplaneta americana *(American cockroach).* Int. J. Dev. Biol., 36, 391–398.

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

Picimbon, J.F., Dietrich, K., Angeli, S., Scaloni, A., Krieger, J., Breer, H. and **Pelosi, P.** (2000) *Purification and molecular cloning of chemosensory proteins from* Bombyx mori*.* Arch. Insect Biochem. Physiol., 44, 120–129.

Picimbon, J.-F.J., Dietrich, K., Krieger, J. and **Breer, H.** (2001) *Identity and expression pattern of chemosensory proteins in* Heliothis virescens *(Lepidoptera, Noctuidae).* Insect Biochem. Mol. Biol., 31, 1173–1181.

Pikielny, C.W., Hasan, G., Rouyer, F. and **Rosbash, M.** (1994) *Members of a family of* Drosophila *putative odourant-binding proteins are expressed in different subsets of olfactory hairs.* Neuron, 12, 35–49.

Rivière, S., Lartigue, A., Quennedey, B., Campanacci, V., Farine, J.P., Tegoni, M., Cambillau, C. and **Brossut, R.** (2003) *A pheromone-binding protein from the cockroach* Leucophaea maderae*: cloning, expression and pheromone binding.* Biochem. J., 371, 573–579.

Robertson, H.M., Martos, R., Sears, C.R., Todres, E.Z., Walden, K.K. and **Nardi, J.B.** (1999) *Diversity of odourant binding proteins revealed by an expressed sequence tag project on male* Manduca sexta *moth antennae.* Insect Mol. Biol., 8, 501–518.

Sabatier, L., Jouanguy, E., Dostert, C., Zachary, D., Dimarcq, J.L., Bulet, P. and **Imler, J.L.** (2003) *Pherokine-2 and -3. Two* Drosophila *molecules related to pheromone/odor-binding proteins induced by viral and bacterial infection.* Eur. J. Biochem., 270, 3398–3407.

Saitou, N. and **Nei, M.** (1987) *The neighbour-joining method: a new method for reconstructing phylogenetic trees.* Mol. Biol. Evol., 4, 406–425.

Shanbhag, S.R., Hekmat-Scafe, D., Kim, M.S., Park, S.K., Carlson, J.R., Pikielny, C., Smith, D.P. and **Steinbrecht, R.A.** (2001) *Expression mosaic of odourant-binding proteins in* Drosophila *olfactory organs.* Microsc. Res. Tech., 55, 297–306.

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

Tsuchihara, K., Fujikawa, K., Ishiguro, M., Yamada, T., Tada, C., Ozaki, K. and **Ozaki, M.** (2005) *An odorant-binding protein facilitates odorant*

transfer from air to hydrophilic surroundings in the blowfly. Chem. Senses, 30, 559–564.

Vogt, R.G. (2002) *Odourant binding protein homologues of the malaria mosquito* An. gambiae*; possible orthologues of the OS-E and OS-F OBPs of* D. melanogaster. J. Chem. Ecol., 28, 2371–2376.

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

Wanner, K.W., Willis, L.G., Theilmann, D.A., Isman, M.B., Feng, Q. and **Plettner, E.** (2004) *Analysis of the insect OS-D-like gene family.* J. Chem. Ecol., 30, 889–911.

Whitfield, C.W., Band, M.R., Bonaldo, M.F., Kumar, C.G., Liu, L., Pardinas, J., Robertson, H.M., Soares, B. and **Robinson, G.E.** (2002) *Annotated expressed sequence tags and cDNA microarrays for studies of brain and behaviours in the honey bee.* Genome Res., 12, 555–566.

Wojtasek, H., Hansson, B.S. and **Leal, W.S.** (1998) *Attracted or repelled?—a matter of two neurons, one pheromone binding protein, and a chiral center.* Biochem. Biophys. Res. Commun., 250, 217–222.

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

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

Accepted March 6, 2006