Cellular and Molecular Life Sciences

RESEARCH ARTICLE

Characterisations of odorant-binding proteins in the tsetse fly *Glossina morsitans morsitans*

**Renhu Liu · Stella Lehane · Xiaoli He · Mike Lehane ·
Christiane Hertz-Fowler · Matthew Berriman ·
John A. Pickett · Linda M. Field · Jing-Jiang Zhou**

Abstract Odorant-binding proteins (OBPs) play an important role in insect olfaction by mediating interactions between odorants and odorant receptors. We report for the first time 20 OBP genes in the tsetse fly *Glossina morsitans morsitans*. qRT-PCR revealed that 8 of these genes were highly transcribed in the antennae. The transcription of these genes in the antennae was significantly lower in males than in females and there was a clear correlation between OBP gene transcription and feeding status. Starvation over 72 h post-blood meal (PBM) did not significantly affect the transcription. However, the transcription in the antennae of 10-week-old flies was much higher than in 3-day-old flies at 48 h PBM and decreased sharply after 72 h starvation, suggesting that the OBP gene expression is affected by the insect's nutritional status. Sequence comparisons with OBPs of other Dipterans identified several homologs to sex pheromone-binding proteins and OBPs of *Drosophila melanogaster*.

R. Liu · X. He · J. A. Pickett · L. M. Field · J.-J. Zhou (✉)
Department of Biological Chemistry, Rothamsted Research,
Harpenden AL5 2JQ, UK
e-mail: jing-jiang.zhou@bbsrc.ac.uk

S. Lehane · M. Lehane
Liverpool School of Tropical Medicine, Pembroke Place,
Liverpool L3 5QA, UK

C. Hertz-Fowler · M. Berriman
Parasite Genomics, Wellcome Trust Sanger Institute,
Hinxton CB10 1SA, UK

Introduction

Tsetse flies (Diptera: Glossinidae) are blood-sucking insects and vectors of many African trypanosome species including *Trypanosoma brucei rhodesiense* and *T. brucei gambiense* causing human sleeping sickness and *T. brucei brucei* causing nagana in live stock [1]. *Glossina morsitans morsitans* is a major species in the Morsitans group involved mainly in the transmission of nagana, but also, on occasion, in the transmission of sleeping sickness or Human African Trypanosomiasis (HAT). Efficiency in killing the vector insects for the control of HAT and nagana using insecticide-treated baits was greatly increased by the addition of appropriate odour attractants [2, 3].

The female fly is slightly larger than the male and females usually mate only once, before or at about the same time as the taking of the first blood meal. Males can mate several times and older males are more likely to mate successfully. The larvae develop inside the female flies and there are three larval instars up to the time when the fully grown larva is dropped by the female fly and the pupae develop underground. Shortly after larviposition, the female ovulates again, depositing a third-instar larva about every 7–11 days, when there are good food supplies and the correct environmental conditions. The life span of females is about 20–40 days and males apparently survive 14–21 days [4, 5]. Newly emerged flies sometimes travel long distances (60–120 m) to locate hosts for their blood meal, with adults of both sexes feeding exclusively on blood.

To find their hosts, tsetse flies use both visual cues and their antenna to detect odours [6]. The flies do not necessarily feed on hosts that happen to be in the same habitat. For example, common animals, such as zebra and wildebeest, are not hosts [7], and each tsetse species has a specialist range of host animals; for example, the preferred host for *G. m. morsitans* is the warthog but it also feeds on ox, buffalo, kudu and human. The reasons for specialisation on different groups of vertebrates are not fully understood. In tsetse, the antennae are comprised of three segments and the upper side of the third and largest segment has a long thin structure, the arista, with a row of branched hairs on its upper side. The third segment also has two small holes leading to sensory pits, which contain many sensory hairs (sensillae) and are important for the detection of odours (e.g. attractants). The electrophysiological recordings of signal cells in the sensory pits identified 141 cells as olfactory receptors, of which 52% responded to 1-octen-3-ol [8]. The effect of age, gender and feeding status on the electroantennogram (EAG) responses to host odours such as l-octen-3-ol, 4-heptanone, 3-nonanone and acetone has been determined [9]. Although nearly all behavioural and electrophysiological studies use female flies before their first blood meal, it has been shown that EAG activity decreased with age in both sexes of *G. m. morsitans*, with responses in 5-day-old flies being significantly lower than those in 1-day-old flies. The EAG responses of males were higher than those of females, and sensitivity increased with increasing starvation in both sexes [9]. Despite the electrophysiological studies with tsetse flies, there is little known about the molecular mechanisms which underpin the olfactory responses.

Antennae contain a high concentration of odorant-binding proteins (OBPs) [10–12] which are small, water-soluble and present at high concentrations in the sensillum lymph surrounding the olfactory receptor neurons, and are believed to be involved in the first step of olfactory molecular recognition and signal transduction by ferrying airborne host odorants across the sensillum lymph to the odorant receptors (ORs) [13–16]. It has been proposed that OBPs could potentially function in odorant recognition [16–20]. Recent work on one of the *Drosophila* OBP LUSH revealed that it is required for the perception of the sex pheromone, with deletion of the gene encoding LUSH suppressing electrophysiological and behavioural responses to the pheromone 11-vaccenyl acetate [20]. LUSH when bound to the pheromone makes a pheromone-specific conformational change that triggers the firing of pheromone-sensitive neurons [21]. Another two OBP genes (*Obp57d* and *Obp57e*) in *D. melanogaster* were found to determine the differential behaviour of two *Drosophila* species, *D. melanogaster* and *D.*

sechellia, to octanoic and hexanoic acids, which act as oviposition attractants for *D. sechiella* but as repellents for other *Drosophila* species [22]. Coexpression of the pheromone-binding protein BmorPBP1 of the moth *Bombyx mori* with the pheromone-receptor BmOR1 in the OR-lacking neuron of *D. melanogaster* facilitates response to the sex pheromone of *B. mori* [23]. Such functional evidence is lacking for OBPs of other insect species.

There are a large number of OBPs present within a variety of insect species [13] and identification of genes encoding OBPs has been mainly by means of bioinformatic approaches based on the characteristic features of the protein families [24–30]. These features include the six-cysteine OBPs signature, a size of 15–20 kDa, the α-helix pattern, the globular water-soluble nature and the presence of a signal peptide. Insect OBPs are very divergent and classed into subfamilies: classic OBPs (possessing all the features), dimer OBPs (having two-six-cysteine signatures), plus-C OBPs (having two additional conserved cysteines plus one proline), minus-C OBPs (having lost two conserved cysteines) and atypical OBPs (having 9–10 cysteines and a long C-terminus) [25, 26, 29]. The highly conserved six cysteines of classic OBPs are a very important feature and contribute to forming disulphide bridges and the tertiary structure of OBPs.

In this study, we took a genome-wide approach and constructed an antennal cDNA library of *G. m. morsitans* for EST sequencing. We also searched for OBP genes in antennal ESTs, four other EST libraries and whole genome shotgun reads and recent genome assembly. The tissue distribution of the transcripts of these genes was then analysed in heads, bodies and antennae. In order to identify the OBP genes associated with host location or sex pheromone detection, transcription profiles of the OBP genes in the antennae were analysed separately in males and females using different starvation periods. Phylogenetic relationships of the OBPs in *G. m. morsitans* were also analysed with other OBPs in Diptera and the evolution of OBP genes in *G. m. morsitans* is discussed.

Materials and methods

Animals and tissue

The *G. m. morsitans* colony was maintained at the Liverpool School of Tropical Medicine (colony established in 2002 from the Bristol colony, itself originally derived from flies from Zimbabwe). Flies were kept at 26°C and 70% relative humidity. Male flies were fed defibrinated horse blood every 48 h by artificial membrane [31]. At 24 h after

a fresh blood meal, flies were frozen at –20°C for 5 min, and then the heads with antennae and bodies were separated and placed in 100 μL of ice-cold Trizol regent (Invitrogen) for RNA extractions.

The antennae samples of *G. m. morsitans* were prepared from (1) newly emerged males, (2) young males at 24 h PBM, (3) young males at 48 h PBM, (4) newly emerged females, (5) young females at 24 h PBM, (6) young females at 48 h PBM, (7) young females at 72 h PBM, (8) 10-week-old females at 48 h after last blood meal, and (9) 10-week-old females at 72 h after last blood meal. To prepare the antennal tissue, the living flies were chilled at 4°C then antennae were detached using sterile fine forceps and immediately placed in an Eppendof tube containing 100 μL ice-cold Trizol regent. About 50–60 pairs of antennae were collected from each sample.

RNA and DNA preparation

Twenty heads with antennae weighing about 42 mg and four bodies without head weighing about 92 mg, were ground separately in liquid nitrogen. Then the tissue powder was mixed with Trizol regent (1 mL) and ground again. The homogenised tissue was then transferred into a 1.5-mL RNase-free tube, mixed with 200 μL chloroform and left for 10 min on the bench at room temperature. For antennal RNA isolation, about 60 pairs of antennae were ground directly in a 1.5-mL eppendof tube with 500 μL Trizol regent using a plastic rod. Total RNA was extracted using Trizol reagent according to the manufacturer's instructions (Invitrogen). Two RNA preparations were made from each antennal sample and this gave two biological replicates.

Genomic DNA was isolated to make standard curves for quantitative RT-PCR. The body tissues of a tsetse fly without the head was homogenised in a 1.5-mL eppendof tube with 250 μL DNA extraction buffer (100 mM Tris–HCl pH 9.0, 100 mM EDTA, 1% SDS). The extraction mixture was heated at 70°C for 30 min, mixed with 35 μL of 8 M KAc solution, and incubated on ice for 30 min. The supernatant containing DNA was obtained by centrifugation of the extraction mixture at the highest speed for 10 min, and then extracted further with 280 μL chlorophorm:phenol 1:1. The DNA sample was treated with 2 μL RNase (10 mg/mL) at 37°C for 15 min, extracted again with 250 μL chlorophorm, and finally DNA was precipitated with 2.5 volume of 100% ethanol. The DNA pellet was washed with 75% ethanol and dissolved in 60 μL water (Sigma). Genomic DNA at the concentrations of 20, 2, 1 and 0.2 were used to plot standard curves for calculating transcript abundance of each gene.

Construction of the antennal cDNA library

The antennal cDNA library was constructed using the Creator® Smart cDNA Library Construction Kit (Clontech) according to supplier's instructions. About 700 ng of antennal RNA from flies of mixed ages and sexes was used for the first strand cDNA synthesis in a reaction volume of 10 μL, from which 2 μL of first strand cDNA was used for long-distance PCR with pre-denaturing at 95°C for 2 min, followed by 20 cycles of 95°C for 15 s and 68°C for 6 min and a final elongation cycle of 72°C for 2 min. The purified resultant double-stand cDNA was digested with *Sfi*I, subjected to size fractionation, and ligated into the pDNR-Lib vector (Clontech). The ligation mixture was desalted, electroporated into *E. coli* XL-blue electro-competent cells and plated on the agar plates supplemented with chloramphenicol (34 μg/ml). Clones were picked into wells of 384-well plates for sequencing. Randomly selected bacterial clones in the EST library were subjected to DNA sequencing. Both sides of the inserts were sequenced with M13 and T7 primers using the ABI™ Big Dye Terminator Cycle Sequencing kits. The sequences were clipped for quality using Phred and vector using Cross Match (P. Green, unpublished). Sequence reads were assembled into clusters with Phrap (P. Green, unpublished; http://www.phrap.org).

Identification of putative OBP genes in *G. m. morsitans*

The sequences from the cDNA libraries of the antennae (this study) and of other tissues (http://www.ncbi.nlm.nih.gov/dbEST/) as well as the whole-genome shotgun reads and the genome assembly produced by the Wellcome Trust Sanger Institute (available from ftp://ftp.sanger.ac.uk/pub/pathogens/Glossina/morsitans/) were searched with a combination of methods: reserve-position-specific BLAST (RPS-BLAST) [32], MotifSearch [27, 29] and BLASTx, on the NCBI website (http://blast.ncbi.nlm.nih.gov/Blast.cgi). The sequence hits were collected as putative OBP sequences if their blast scores were less than 10^{-5} to either the PBP-GOBP domain (pfam01395) or to known OBP sequences, and if they had either the six or four conserved cysteine signature, with a low molecular weight and a hydrophobic signal peptide.

Quantitative PCR

Primers were designed for quantitative RT-PCR using Primer3Plus (http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi). The primers were between 19 and 22 base pairs (bp) with a T_m values ranging from 59.5 to 60.5°C. The size of the PCR products was set within the range 120–260 bp. Since genomic DNA was used to

establish standard curves for quantification, the primers were optimized and designed on single exon or exons flanking short introns. Information on the primers is given in the electronic supplementary material, ESM, Table 4.

Invitrogen Platinum® SYBR Green qPCR SuperMix-UDG (Invitrogen) was used for the qPCR reactions. An aliquot (5 μL) of total RNA ranging from 0.2 to 1 μg was first treated with RQ1 DNase in a total volume of 10 μL reaction mixture. Then 6 μL of the reaction mixture was used for reverse transcription using the ImProm-II™ Reverse Transcription System (Promega, Southampton, UK) in 20-μL reaction volumes and then diluted to 500 μL, of which 3 μL was used as template of each qPCR reaction. Each reaction included 1× SuperMix, 200 nM of each of gene-specific primer pairs, 50 nM ROX dye, and 3 μL templates. The qPCR was carried out for each of two RNA preparations from each tissue sample. The copy number of each gene was estimated from the standard curves of genomic DNA based an estimated haploid genome size of 570 Mb for *G. m. morsitans* [33]. On each PCR plate, the β-tublin gene of *G. m. morsitans* was included for initial normalization of the template amount. The PCR reactions were carried out on a Stratagene Mx3000P qPCR system (Agilent Technologies UK, Cheshire, UK) with a thermo-profile of one cycle of 50°C 2 min, 95°C 2 min, then 45 cycles of 95°C 15 s, and 60°C 45 s, followed by a melting curve analysis from 55 to 95°C.

Cluster analysis of expression patterns

Transcript data obtained by qRT-PCR was analysed with the hierarchical clustering method using cluster analysis software Cluster 2.11 [34]. Such cluster analysis uses standard statistical algorithms to arrange genes according to similarity in pattern of gene expression. To adjust the data so that each gene has an equal weight, the options of "normalise by genes and arrays", "mean centre by genes and arrays" were selected. The cluster figure was visualised in TreeView 1.6.

Sequence similarity comparison

Protein sequences of OBPs were aligned in ClustalX 2.08 [35] with the option of "iterate each alignment step". The

Table 1 Odorant binding proteins (OBPs) predicted from genes identified from sequence data of *Glossina morsitans morsitans*

Name	Length[a]	GenBank Protein accession number	Closest Dmel OBP (identities, E value)	Sources[b]	Type	Signal Peptide
GmmOBP1	141	FN432779	Obp44a (62%, 1e-48)	H, L, P, T	Classic OBP	1–17aa
GmmOBP2	153	FN432780	Obp99b (51%, 5e-42)	L, P, T	Classic OBP	1–18aa
GmmOBP3	>224	FN432781	Obp83ef (46%, 1e-36)	F, T	Dimer OBP	NA
GmmOBP4	141	FN432782	Obp56e (34%, 5e-15)	F, T	Classic OBP	No SP
GmmOBP5	184	FN432783	Obp19c (32%, 9e-22)	L, T	Classic OBP	1–21aa
GmmOBP6	145	FN432784	Pbprp5 (46%, 1e-28)	L, T	Classic OBP	1–19aa
GmmOBP7	240	FN432785	Obp83cd (42%, 4e-63)	L, F, T	Dimer OBP	1–19aa
GmmOBP8	150	FN432786	Pbprp3 (48%, 6e-40)	A, T	Classic OBP	1–24aa
GmmOBP9	150	FN432787	Pbprp3 (82%, 3e-60)	A, T	Classic OBP	1–26aa
GmmOBP10	>94	FN432788	Pbprp3 (72%, 4e-36)	A, T	Classic OBP	NA
GmmOBP11	>118	FN432789	Obp83g (63%, 3e-47)	T	Classic OBP	NA
GmmOBP12	>95	FN432790	Pbprp3 (69%, 5e-36)	T	Classic OBP	NA
GmmOBP13	>113	FN432791	Obp56h (43%, 3e-24)	T	Classic OBP	NA
GmmOBP14	>121	FN432792	Obp19a (62%, 2e-45)	T	Classic OBP	NA
GmmOBP15	>111	FN432793	Obp56d (32%, 5e-13)	T	Classic OBP	NA
GmmOBP16	>93	FN432794	Obp57c (34%, 6e-12)	T	Classic OBP	NA
GmmOBP17	>102	FN432795	Obp56i (28%, 4e-6)	T	Classic OBP	NA
GmmOBP18	>109	FN432796	Pbprp1 (46%, 8e-11)	T	Classic OBP	NA
GmmOBP19	138	FN432797	No hit	T	Classic OBP	1–19aa
GmmOBP20	>74	FN432798	No hit	T	Classic OBP	1–19aa

NA The signal peptide sequences could not be determined because of partial gene sequences

[a] Total number of amino acids of each OBP; > partial coding sequences

[b] Sequences are identified from EST libraries of: *A* antennae, *H* heads, *L* larvae, *P* pupae, *R* reproductive organs, *F* fat bodies, *G* adult gut, *S* salivary glands, and *T* genomic shotgun reads

alignments were adjusted manually in GenDoc for the conserved cysteine residues and the gaps between codons, and used in the construction of phylogenetic trees with Mega 4 [36].

Results and discussion

Construction of an antennal cDNA library of *G. m. morsitans*

A non-normalised antennal cDNA library was constructed from a pool of total RNAs extracted from the antennae of newly emerged and 10-week-old blood-fed flies of mixed sex. From the library, 11,520 cDNA clones were picked into gridded wells of 384-well plates. From these 2,000 clones were sequenced and gave 1,787 unique cDNAs ranging from 100 to 1,358 bp with an average length of 351 bp. The majority of sequences (91.5%) had no similarity to any protein coding sequences in GenBank. Blast searches against the GenBank non-redundant protein database gave 150 (8.4%) matches with E-values less than 10^{-6}, including 86 (4.8%) highly conserved matches with value E values less than 10^{-20}. These include a polycomb protein (a gene regulator in *D. melanogaster*), spectrin (a cytoskeletal protein), muscle lim protein, ribosomal protein and a tRNA synthase. Other genes showed some identities with known genes of other insects but were very divergent, such as the genes encoding proline-rich proteoglycan (acidic proline-rich protein), G2/mitotic-specific cyclin-B, menin, C-factor (C signal), NF-κ-B inhibitor cactus, and zinc-type alcohol dehydrogenase. The genes found in the antennal library encoding proteins with similarity to known proteins in GenBank are listed in ESM, Table 1. Putative functions of these genes include semiochemical transport and metabolism, neuron or cell signalling, gene regulation, cell constitutive components and basic metabolism.

Identification of genes encoding putative OBPs in *G. m. morsitans*

From the antennal cDNA library, we identified three OBP genes (*GmmOBP8*, *GmmOBP9* and *GmmOBP14*) using the reserve-position-specific BLAST (RPS-BLAST) [32], the MotifSearch [27, 29] and the BLASTx on NCBI website (http://blast.ncbi.nlm.nih.gov/Blast.cgi). In order to identify further OBPs, other available sequences were also searched including tissue-specific cDNA libraries from heads, fat bodies, salivary glands and reproductive organs, and larval and pupal cDNAs (http://www.ncbi.nlm.nih.gov/dbEST/), as well as the whole-genome shotgun (WGS) reads and the genome assembly produced by the *G. m. moristans*
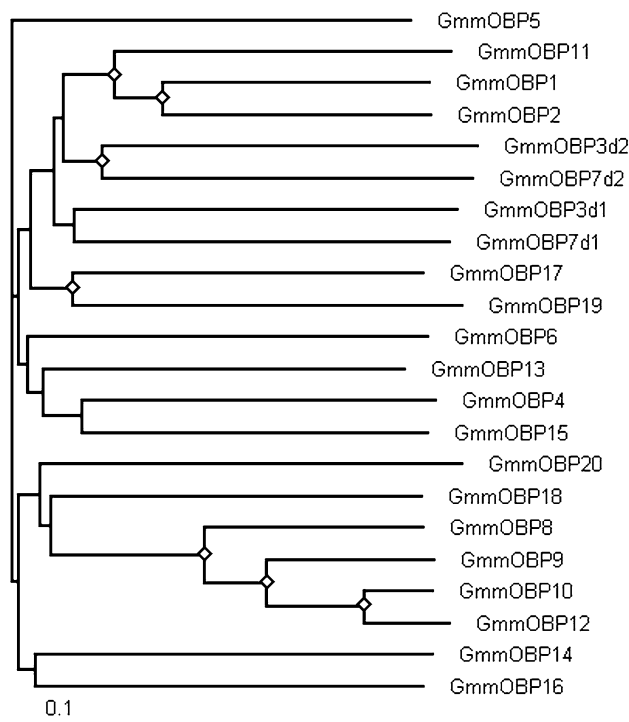

Fig. 1 Sequence relatedness of the OBP protein families in *Glossina morsitans morsitans*. The unrooted neighbour-joining tree was constructed to show the divergence amongst members of the protein families. Signal peptides were removed before alignment in ClustalX 2.0.8 with the alignment option of "iterate each alignment step". Each of the dimer OBPs, GmmOBP3 and GmmOBP7, is composed of two classic OBP units, which are presented as GmmOBP3d1, GmmOBP3d2, GmmOBP7d1 and GmmOBP7d2, respectively, for alignment. *Diamonds* indicate nodes with 70% bootstrapping support

Sequencing Group at the Sanger Institute and available from the Wellcome Trust Sanger Institute FTP site (ftp://ftp.sanger.ac.uk/pub/pathogens/Glossina/morsitans/). From a combination of these, we identified 20 putative OBP (Table 1) from 75,278 ESTs and 1,536 Mb WGS reads. Of the 20 putative OBP genes, 18 are predicted to encode classic OBPs and 2 (*GmmOBP3* and *GmmOBP7*) to encode dimers with two tandem repeats of the classic OBP domains as found in the *Drosophila* OBP genes *OBP83cd* and *OBP83ef* [29].

The OBP genes in *G. m. moristans* encode a diverse family of proteins (Fig. 1) as found in other Dipterans [27]. The average identity of the peptide sequences is about 25% with most relationships having weak bootstrap support. However, some are more closely related and have high identities, e.g. GmmOBP8, GmmOBP9, GmmOBP10 and GmmOBP12, which are clustered together, and GmmOBP1, GmmOBP2 and GmmOBP11 clustered together with bootstrap support (70%) and high identities of 45% between GmmOBP1 and GmmOBP2 and 84% between GmmOBP10 and GmmOBP12 (Fig. 1).
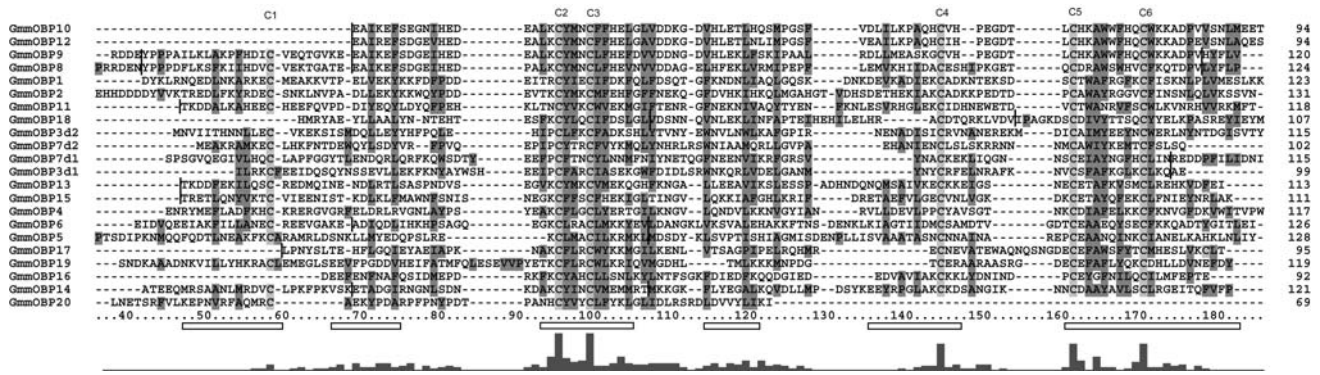
Fig. 2 Alignment of OBPs of *Glossina morsitans morsitans*. The amino acids highlighted in *dark grey* are conserved hydrophobic residues, and the conserved cysteine residues are highlighted in *light grey*. The *rectangular bars* under the alignment represent the positions of the α-helices of the *Drosophila melanogaster* OBP LUSH for GmmOBP8. The splice sites are indicated by *black vertical bars*

Fig. 3 Transcript abundance of genes encoding OBPs in *Glossina morsitans morsitans* as determined by quantitative real-time PCR. The mean values and standard errors of four replicates are presented relative to the internal control gene
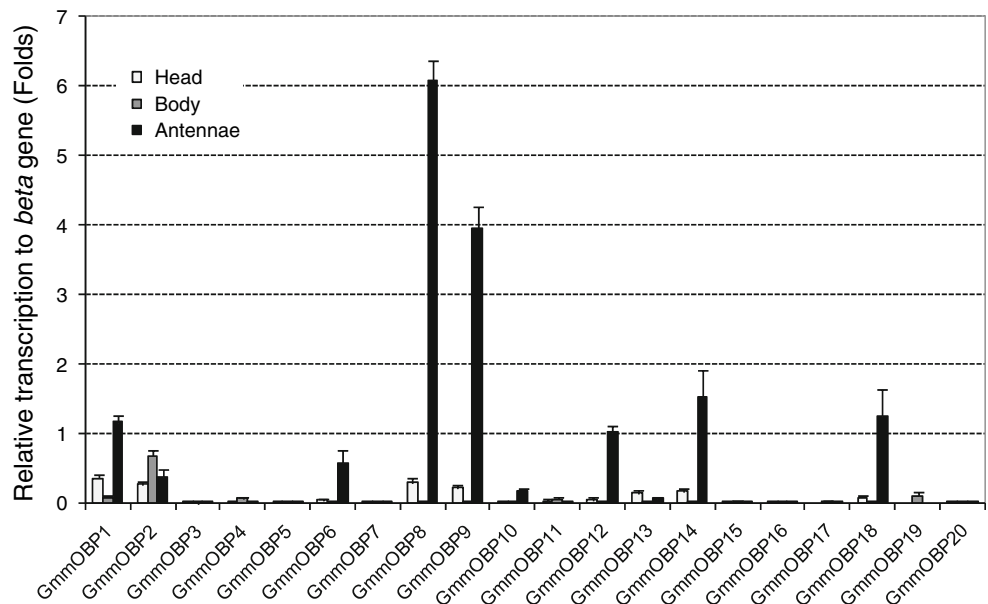
The number of putative OBPs identified in *G. m. morsitans* is small compared to the number in other Dipteran insects. This may be due to a poor representation of the genome in the latest assembly—the total length of the genome assembly is 377 Mb, rather than the previously estimated 570 Mb. We cannot rule out that more OBPs are present in an unclonable, and therefore unsequenced, fraction of the genome. However, given the large number of ESTs and about 66% genome searched, we believe that most of the OBP genes with high levels of transcription have been identified. It is perhaps likely that tsetse flies have a specialised and reduced OBP repertoire. The only free-living stage in the life cycle is the adult (the larvae are nourished entirely within the body of the female insect). So there are none of the complications of having larval stages with a completely different ecology to the adults as found for mosquitoes.

Introns of OBP genes in *G. m. morsitans*

Of the 20 putative OBP genes found in *G. m. morsitans*, two genes have no intron, two have one intron, four have three introns, four have two introns, and the remaining eight genes are partial genes with at least one intron. The positions of introns in the genes are not well conserved between genes (Fig. 2), and the splicing often happens at G↓G, whereas in *A. mellifera* most OBP genes splice at a consensus site GT↓AG [37]. These findings indicate that intron variation has occurred in the evolution of OBP genes in insects.

Amino acids and sequence motifs of putative OBPs in *G. m. morsitans*

As for OBPs in other insect species, the six-cysteine signature (C_1–C_6 in Fig. 2) is the most striking

Fig. 4 Transcription of OBP genes in the antennae of *Glossina morsitans morsitans*. The transcript levels are presented as mean fold changes relative to the internal reference with four replicates and are divided into three groups: high (**a**), medium (**b**) and low (**c**) expression levels. *PBM* Post-blood meal
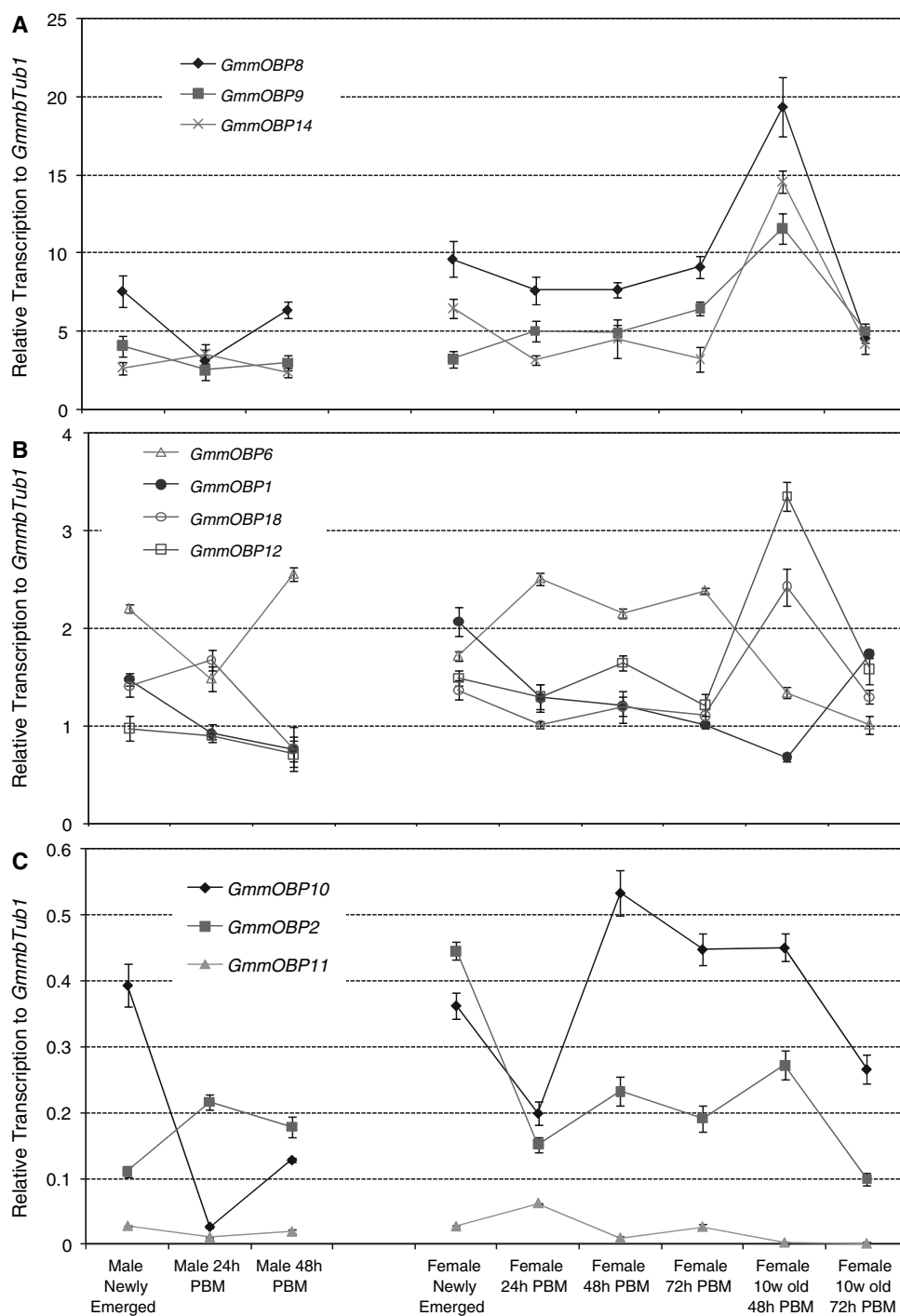
conservation of amino acids in all the mature predicted OBPs of *G. m. morsitans*. However, we found that the physical properties of many of the amino acids are also conserved; for example, the hydrophobic residues (highlighted in grey in Fig. 2), indicating the importance of these residues in forming the cavity which binds to hydrophobic odorants.

Overall, the sequence motif of the *G. m. morsitans* OBPs is $C_1–X_{20-35}–C_2–X_3–C_3–X_{20-30}–C_4–X_{8-12}–C_5–X_8–C_6$ with the variation in spacing between $C_1–C_2$ and $C_3–C_4$ being less than that found in *D. melanogaste*. The conservation of the six cysteines suggests similar bridging occurs to give the tertiary structures similar to those found previously in *D. melanogaster* [38], *Anopheles gambiae* [39] and *Bombyx mori* [19, 40].
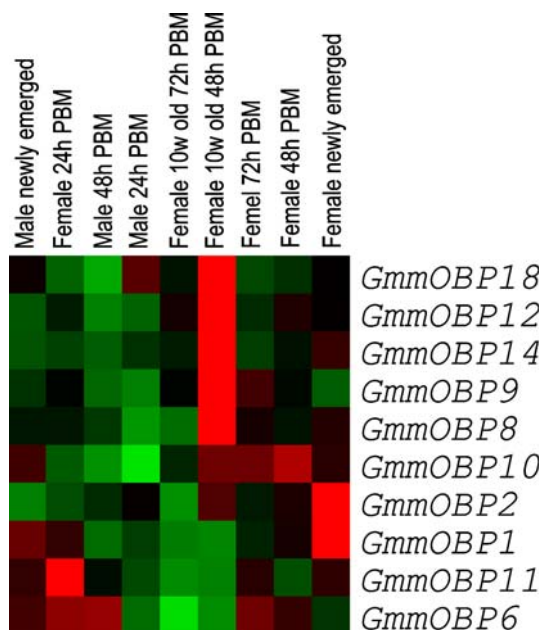
Fig. 5 Cluster analysis of transcription profiles of OBP genes in antennae of *Glossina morsitans morsitans* based on the results in Fig. 4. The transcript abundance is represented with *coloured squares*; *red squares* indicate transcription higher than average and *green squares* indicate transcription lower than average. *PBM* Post-blood meal

Tissue distribution of OBP genes

To characterise further the *G. m. morsitans* OBPs, we carried out quantitative analyses on the tissue distribution of gene transcripts by real time RT-PCR. Firstly, we measured transcriptional levels of two reference genes, *actin* (*GmmAct1*, DV612275) and *β-tublin* (*GmmbTub1*, PUP-002P06.g) in heads, bodies and antennae. According to the standard curves made from genomic DNA, the transcript abundance of *GmmbTub1* was $(3.65 \pm 0.15) \times 10^5$, $(3.65 \pm 0.12) \times 10^5$, and $(4.86 \pm 0.09) \times 10^5$ copies/ng total RNA in heads, bodies and antennae, respectively. However, the transcript levels of *GmmAct1* varied from $(1.58 \pm 0.08) \times 10^3$ in heads to $(7.33 \pm 0.43) \times 10^3$ in bodies and $(1.50 \pm 0.15) \times 10^3$ in antennae. We therefore chose *GmmbTub1* as the reference gene for quantitative RT-PCR analyses to quantify the transcript abundance of the OBP genes in heads, bodies and antennae, respectively, and the results are shown in Fig. 3.

Of the 20 OBP genes, 10 were found to be transcribed in antennae at a level of at least 0.1 times higher than *GmmbTub1* (Fig. 3). Of these, three (*GmmOBP8*, *GmmOBP9*, *GmmOBP14*) had the highest transcription level being about 6.1, 4.0 and 1.5 times higher than *GmmbTub1*, equivalent to 2.92×10^6, 1.94×10^6 and 7.43×10^6 copies/ng total RNA, respectively. Thus, these three OBPs are good candidates for a role in *G. m. morsitans*

olfaction. Four OBP genes (*GmmOBP4*, *GmmOBP15*, *GmmOBP17*, *GmmOBP19*) are transcribed only in adult bodies (Fig. 3). One of these (*GmmOBP19*) seems to be a tsetse fly-specific gene, since it has no similarity to any other gene in GenBank (Table 1). Five other OBP genes (*GmmOBP3*, *GmmOBP5*, *GmmOBP7*, *GmmOBP16*, *GmmOBP20*) have low, or no, transcription in adult tissues and three (*GmmOBP3*, *GmmOBP5*, *GmmOBP7*) are found only in larval tissues (Table 1).

Transcription of OBPs in male and female *G. m. morsitans* at different life stages

The transcription of the putative OBP genes in antennae was further analysed in male and female flies at different points in the feeding cycle; newly emerged and 24 or 48 h post-blood meal (PBM). The data are presented in three groupings; those with transcription levels 2.5 times higher than *GmmbTub1* (Fig. 4a), those with levels between 0.5 and 2.5 times that of *GmmbTub1* (Fig. 4b) and those with levels below 0.5 times that of *GmmbTub1* (Fig. 4c). The results show that in all three groups the transcription of OBP genes is significantly higher in female antennae than in male antennae for newly-emerged and 24 or 48 h PBM (Fig. 4). This is most noticeable for the genes that are highly expressed in the antennae (Fig. 4a) with 1.21, 1.64, and 1.89 fold higher expressions in females at 48 h PBM for *GmmOBP8*, *GmmOBP9* and *GmmOBP14*, respectively. Thus, all the genes likely to be involved in olfaction are expressed most highly in the females. The higher levels of expression in female antennae may be due to higher demand for blood meal in females. Both sexes of tsetse fly are obligate haematophages. Males in the Morsitans group of tsetse flies form 'following swarms' in which non-feeding males gather around moving objects (normally host animals) in order to intercept females coming to feed [1]. This behaviour on the part of the male, in which they maintain closer contact with hosts for longer periods of time than do female flies, may mean that the task of host location is simpler for male flies.

For males, there are no significant changes in the transcription of the OBP genes in the antennae over the 48 h PBM compared to that of newly emerged male flies (Fig. 4). In females for seven (*GmmOBP1*, *GmmOBP2*, *GmmOBP8*, *GmmOBP10*, *GmmOBP12*, *GmmOBP14*, *GmmOBP18*) of the ten genes tested, transcription decreased during the first 24 h PBM, suggesting that it is the blood feeding which causes the decrease. For other genes (*GmmOBP9*, *GmmOBP11*, *GmmOBP12*), there are no further significant changes in transcription up to 72 h PBM (Fig. 4b, c), with only the one exception of *GmmOBP6* where transcription is much higher 72 h PBM than that before the blood meal. In all cases, further
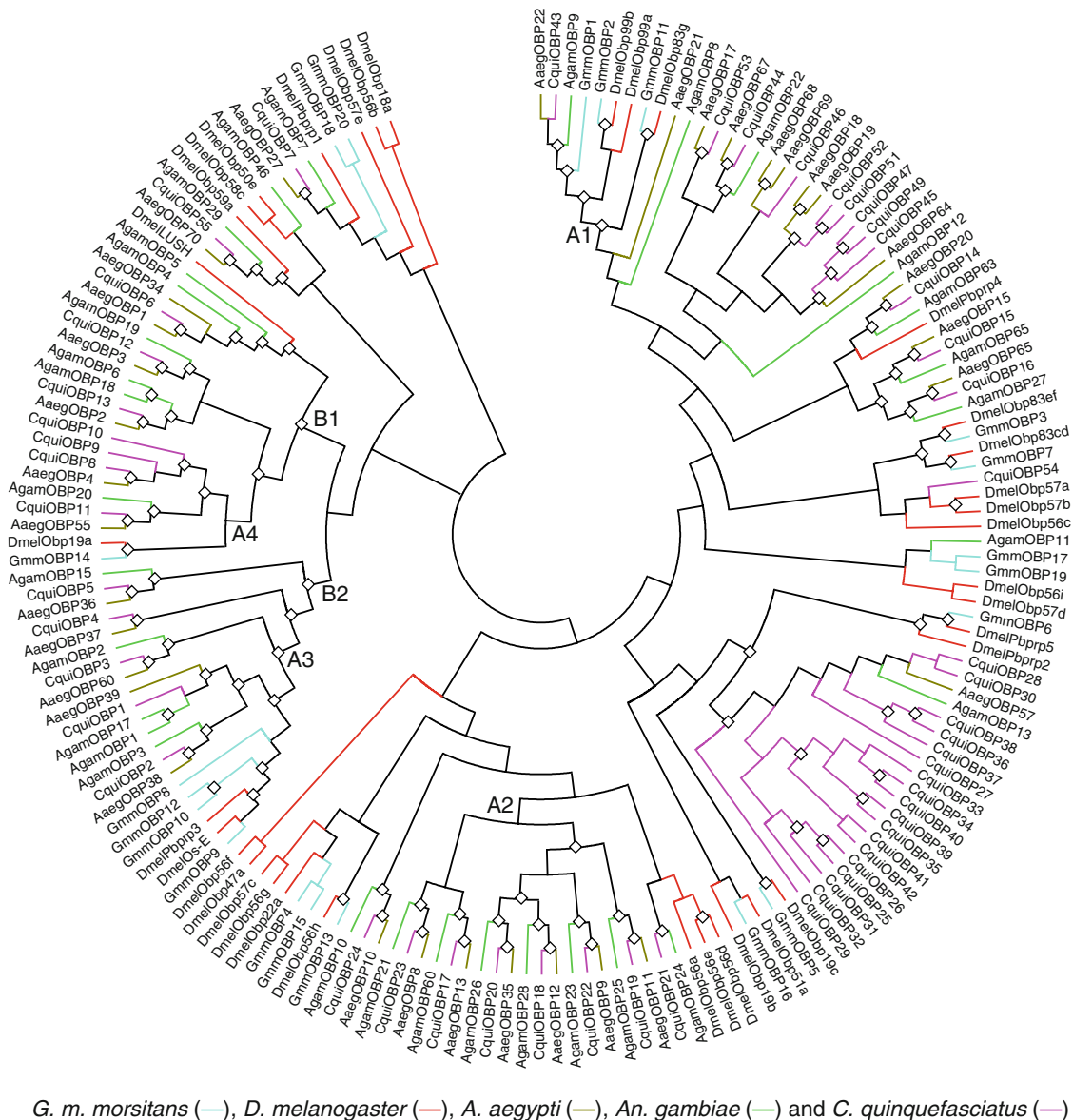
G. m. morsitans (—), *D. melanogaster* (—), *A. aegypti* (—), *An. gambiae* (—) and *C. quinquefasciatus* (—)

Fig. 6 Phylogenetic tree (topology only) of OBPs in Diptera including *Glossina morsitans morsitans* (Gmm in *turquoise*), *Drosophila melanogaster* (Dmel in *red*), *Aedes aegypti* (Aaeg in *dark yellow*), *Anopheles gambiae* (Agam in *green*) and *Culex quinquefasciatus* (Cqui in *pink*). *Diamonds* indicate the nodes with 70% bootstrap support. Proteins labelled with squares are dimer OBPs. The clades labelled with *A1*, *A2*, *A3* and *A4* reflect the correlation between the phylogenetic relationships of the OBPs and the evolution of the species of Diptera based on classic taxonomy (Fig. 7). There are two main orthologous groups (*B1* and *B2*) with strong boostrapping support and showing a clear phylogenetic relationship amongst the OBPs from the different species

starvation over 72 h after the first blood meal did not significantly affect the OBP gene expression.

Transcription was found to be much higher in the antennae of 10-week-old female flies at 48 h PBM than in younger flies for most of the highly expressed OBP genes (Fig. 4a, b). Thus, the transcription levels are 2.5- (*GmmOBP8*), 2.4- (*GmmOBP9*), 3.2- (*GmmOBP14*), 2.0- (*GmmOBP18*) and 2.0- (*GmmOBP12*) fold higher in the antennae of 10-week-old female flies at 48 h PBM than in younger female flies. Unlike younger female flies, however, the gene transcription decreases dramatically

after 48 h PBM expected for *GmmOBP1* (Fig. 4). This decrease was also observed for the genes (*GmmOBP2* and *GmmOBP10*) that are expressed at the lowest level in the antennae (Fig. 4c). These results clearly demonstrate a correlation between the expression level of some of the OBP genes and the fly's physiological desire for a blood meal. These observations fit well with the behaviour of the fly where the typical feeding interval is believed to be 48–72 h, although considerable variation occurs under field conditions where host availability is a key factor [33].

Fig. 7 Evolutionary tree of Dipteran species based on classic taxonomy obtained from NCBI; the time scale was estimated from a range of published data
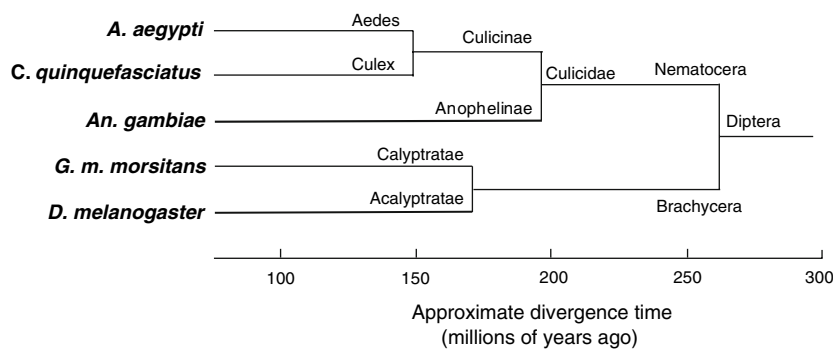
Cluster analysis of transcript expression in antennae

To examine the relative transcription of each of the genes in antennae, we carried out transcriptome cluster analysis on the transcript expression data of nine antennal samples [34]. The analysis first normalises the qRT-PCR values as shown in Fig. 4 to an overall centralised mean value and then arranges the genes according to similarity in their expression patterns over the nine antennal samples. The results (Fig. 5) show that five of the OBP genes (*GmmOBP18*, *GmmOBP12*, *GmmOBP14*, *GmmOBP9*, *GmmOBP8*) have their highest expression in the antennae of 10-week-old female flies at 48 h PBM. This suggests some OBP genes, although they are expressed in antennae at a lower level, also increased their expression in the antennae of the older flies (Figs. 3 and 4). Another two OBP genes (*GmmOBP1*, *GmmOBP2*) have their highest expression in the antennae of newly-emerged female flies, but these genes are also found to have a high transcription in both heads and bodies (Fig. 3). Moreover, they have a similar transcription profile to the *actin* gene (data not shown). It is possible that these genes may have functions other than odour detection.

Comparison of *G. m. morsitans* OBPs with OBPs of other Dipteran insects

To compare the OBP families of *G. m. morsitans* with those of other Dipterans, we compared sequence similarities using protein sequences of reported classic OBPs (Fig. 6), including *Culex quinquefasciatus*, *D. melanogaster*, *Anopheles gambiae* and *Aedes aegypti* whose OBP genes have been fully annotated from their genomes [25, 26, 29, 30]. A total of 184 classic OBPs (including 13 dimers) were analysed. The GenBank accession numbers of all the sequences used are listed in ESM, Tables 2 and 3.

The sequence similarities were used to construct the neighbour-joining tree of the OBPs shown in Fig. 6. The Diptera split into two lineages about 260 million years ago: Brachycera, to which *G. m. morsitans* and *D. melanogaster* belong, and Nematocera to which the mosquito species belong (Fig. 7), and there are four branches (A1–A4) in

which the close similarity between OBP sequences appears to reflect the close relationship between these two Dipteran species (Fig. 7). The OBPs of *G. m. morsitans* are often clustered with the OBPs of *D. melanogaster* whilst all the mosquitoes OBPs are clustered together with at least 70% bootstrapping support (Fig. 6). GmmOBP1 and GmmOBP2 share a common ancestor with mosquito OBPs and fruit fly OBPs, respectively, in branch A1. GmmOBP8 shares a common ancestor with both mosquito and fruit fly OBPs in branch A2. Ten out of the 20 *G. m. morsitans* OBPs have a close homolog in *D. melanogaster* with a minimum identity of 31% between GmmOBP15 and DmelOBP56c and a maximum identity of 82% between GmmOBP9 and OS-E (Fig. 6). There is no homolog of LUSH, the pheromone-binding protein of *D. melanogaster* [20, 21]. It is interesting that all OBP genes highly expressed in the antennae and in 10-week-old flies (Figs. 3, 4, 5) are clustered with OS-E and Pbprp3 (OS-F) of *D. melanogaster* (Fig. 6) which are co-expressed with the *D. melanogaster* pheromone-binding protein LUSH in one neuron [11]. It is very interesting that only GmmOBP1 is clustered with the OBPs from the three mosquito species (branch A1 in Fig. 6) suggesting that it has been positively selected for in blood-feeding insects.

Conflict of interest statement The authors have no commercial conflicts of interest.

References

1. Leak SG (1999) Tsetse biology and ecology; their role in the epidemiology and control of trypanosomiasis. CABI, Oxon

2. Omolo MO, Hassanali A, Mpiana S, Esterhuizen J, Lindh J, Lehane MJ, Solano P, Rayaisse JB, Vale GA, Torr SJ, Tirados I (2009) Prospects for developing odour baits to control *Glossina fuscipes* spp., the major vector of human african trypanosomiasis. PLoS Negl Trop Dis 3:e435

3. Torr SJ, Hargrove JW, Vale GA (2005) Towards a rational policy for dealing with tsetse. Trends Parasitol 21:537–541

4. Glasgow JP (1963) The distribution and abundance of tsetse. Macmillan, New York

5. Potts WH (1973) Glossinidae (tsetse flies). In: Smith KGV (ed) Insects and other arthropods of medical importance. British Museum (Natural History), London, pp 209–249

6. Vale GA (1977) The xight of tsetse xies (Diptera, Glossinidae) to and from a stationary ox. Bull Entomol Res 67:297–303

7. Clausen PH, Adeyemi I, Bauer B, Breloeer M, Salchow F, Staak C (1998) Host preferences of tsetse (Diptera, Glossinidae) based on bloodmeal identifications. Med Vet Ent 12:169–180

8. den Otter CJ, Van der Goes van Naters WM (1992) Single cell recordings from tsetse (*Glossina m. morsitans*) antennae reveal olfactory, mechano and cold receptors. Physiol Entomol 17:33–42

9. den Otter CJ, Tchicaya T, Schutte AM (1991) Effects of age, sex and hunger on the antennal olfactory sensitivity of tsetse flies. Physiol Entomol 16:173–183

10. Vogt RG, Riddiford LM (1981) Pheromone binding and inactivation by moth antennae. Nature 293:161–163

11. Shanbhag SR, Hekmat-Scafe D, Kim MS, Park SK, Carlson JR, Pikielny C, Smith DP, Steinbrecht RA (2001) Expression mosaic of odorant-binding proteins in Drosophila olfactory organs. Microsc Res Tech 55:297–306

12. Calvello M, Brandazza A, Navarrini A, Dani FR, Turillazzi S, Felicioli A, Pelosi P (2005) Expression of odorant-binding proteins and chemosensory proteins in some Hymenoptera. Insect Biochem Mol Biol 35:297–307

13. Pelosi P, Zhou J-J, Ban LP, Calvello M (2006) Soluble proteins in insect chemical communication. Cell Mol Life Sci 63:1658–1676

14. Smith DP (2007) Odor and pheromone detection in *Drosophila melanogaster*. Pflugers Arch 454:749–758

15. Vogt RG, Prestwich GD, Lerner MR (1991) Odorant-binding-protein subfamilies associate with distinct classes of olfactory receptor neurons in insects. J Neurobiol 22:74–84

16. Leal WS, Chen AM, Erickson ML (2005) Selective and pH-dependent binding of a moth pheromone to a pheromone-binding protein. J Chem Ecol 31:2493–2499

17. Du G, Prestwich GD (1995) Protein structure encodes the ligand binding specificity in pheromone binding proteins. Biochemistry 34:8726–8732

18. Pelosi P (1994) Odorant-binding proteins. Crit Rev Biochem Mol Biol 29:199–228

19. Zhou J-J, Robertson G, He XL, Dufour S, Hooper AM, Pickett JA, Keep NH, Field LM (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

20. Xu P, Atkinson R, Jones DN, Smith DP (2005) Drosophila OBP LUSH is required for activity of pheromone-sensitive neurons. Neuron 45:193–200

21. Laughlin JD, Ha TS, Jones DN, Smith DP (2008) Activation of pheromone-sensitive neurons is mediated by conformational activation of pheromone-binding protein. Cell 133:1255–1265

22. Matsuo T, Sugaya S, Yasukawa J, Aigaki T, Fuyama Y (2007) Odorant-binding proteins OBP57d and OBP57e affect taste perception and host–plant preference in *Drosophila sechellia*. PLoS Biol 5:e118

23. Syed Z, Ishida Y, Taylor K, Kimbrell DA, Leal WS (2006) Pheromone reception in fruit flies expressing a moth's odorant receptor. Proc Natl Acad Sci USA 103:16538–16543

24. Li Z, Pickett JA, Field LM, Zhou J-J (2005) Identification and expression profiling of putative odorant-binding proteins in the malaria mosquitoes, *Anopheles gambiae* and *Anophele arabiensis*. Arch Insect Biochem Physiol 58:175–189

25. 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 12:1357–1369

26. Xu PX, Zwiebel LJ, Smith DP (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

27. Zhou J-J, He XL, Pickett JA, Field LM (2008) Identification of odorant-binding proteins of the yellow fever mosquito *Aedes aegypti*, genome annotation and comparative analyses. Insect Mol Biol 17:147–163

28. Zhou J-J, Kan Y, Antoniw J, Pickett JA, Field LM (2006) Genome and EST analyses and expression of a gene family with putative functions in insect chemoreception. Chem Senses 31:453–465

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

30. Pelletier J, Leal WS (2009) Genome analysis and expression patterns of odorant-binding proteins from the southern house mosquito *Culex pipiens quinquefasciatus*. PLoS One 4:e6237

31. Moloo SK (1971) An artificial feeding technique for *Glossina*. Parasitology 63:507–512

32. Marchler-Bauer A, Panchenko AR, Shoemaker BA, Thiessen PA, Geer LY, Bryant SH (2002) CDD: a database of conserved domain alignments with links to domain three-dimensional structure. Nucleic Acids Res 30:281–283

33. Aksoy S, Berriman M, Hall N, Hattori M, Hide W, Lehane MJ (2005) A case for a *Glossina* genome project. Trends Parasitol 21:107–111

34. Eisen MB, Spellman PT, Brown PO, Botstein D (1998) Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci USA 95:14863–14868

35. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007) Clustal W and Clustal X version 2. 0. Bioinformatics 23:2947–2948

36. Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4. 0. Mol Biol Evol 24:1596–1599

37. Foret S, 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

38. Kruse SW, Zhao R, Smith DP, Jones DN (2003) Structure of a specific alcohol-binding site defined by the odorant binding protein LUSH from *Drosophila melanogaster*. Nat Struct Biol 10:694–700

39. Wogulis M, Morgan T, Ishida Y, Leal WS, Wilson DK (2006) The crystal structure of an odorant binding protein from *Anopheles gambiae*, evidence for a common ligand release mechanism. Biochem Biophys Res Commun 339:157–164

40. Sandler BH, Nikonova L, Leal WS, Clardy J (2000) Sexual attraction in the silkworm moth, structure of the pheromone-binding-protein-bombykol complex. Chem Biol 7:143–151