

ODORANT BINDING PROTEIN HOMOLOGUES OF THE MALARIA MOSQUITO *Anopheles gambiae*; POSSIBLE ORTHOLOGUES OF THE OS-E AND OS-F OBPS OF *Drosophila melanogaster*

RICHARD G. VOGT^{1,*}

¹*Department of Biological Sciences, University of South Carolina, Columbia SC 29208*

(Received August 12, 2002; accepted August 27, 2002)

Abstract—Twenty-nine *Anopheles gambiae* candidate Odorant Binding Proteins (OBPs) were characterized for similarity to OBPs of *Drosophila melanogaster* and other insects. Twenty-five of these sequences were identified by BLAST searching the *A. gambiae* genome database. Several *A. gambiae* sequences were significantly similar to the *D. melanogaster* OBPs OS-E/OS-F, LUSH and PBPRP2/PBPRP5. Exon boundary comparisons suggests that two *A. gambiae* genes are orthologues of OS-E and OS-F, justifying the names *AgamOS-E* (EAA01090, AF437886) and *AgamOS-F* (EAA14641, AF437884). If these are orthologues, then the gene duplication establishing the OS-E and OS-F lineages predated the divergence of mosquitoes and flies. The identification of orthologous OBPs and other chemosensory genes between *D. melanogaster* and *A. gambiae* should accelerate the transfer of physiological and behavioral information between these two species.

Key Words—Odorant Binding Protein, OBP, evolution, odor detection, pheromone detection, chemosensory proteins, *Anopheles gambiae*, mosquito, olfaction.

*To whom correspondence should be addressed. E-mail: VOGT@BIOL.SC.EDU

INTRODUCTION

Odorant Binding Proteins (OBPs) are small water-soluble proteins thought to transport odors to receptor proteins in olfactory neuron membranes (Vogt et al., 1999). OBPs comprise a multigene family; thirteen have been identified in *Manduca sexta* through direct sequencing, and 38 in *Drosophila melanogaster* through direct sequencing and genome analysis (Robertson et al., 1999; Galindo and Smith, 2001; Graham and Davies, 2002; Vogt et al., 2002). OBPs are differentially expressed among olfactory sensilla, contributing to the unique phenotypes and odor sensitivities of these sensilla (de Bruyne et al., 2001; Shanbhag et al., 2001). In this study, OBP genes of *D. melanogaster* and other insect species were used to identify OBPs from the *A. gambiae* genome, and the likely *A. gambiae* orthologues of the *D. melanogaster* OBPs OS-E and OS-F (also known as PBPRP3) are identified.

METHODS

Amino acid sequences of previously reported OBPs were used in a BLAST-P search of the *A. gambiae* genome through the NCBI web site (8/2002). This database identifies hypothetical gene coding regions and the corresponding amino acid sequences. Sequences with an E-value of 0.05 or lower were accepted. Sequences were aligned using ClustalX, and a neighbor joining tree was constructed using MEGA2. No effort was made to modify sequences, although previously reported versions of these sequences were used if available. Exon boundaries were determined by translating the genomic sequences into multiple reading frames to identify the amino acid sequences bounding exon/intron transitions.

RESULTS AND DISCUSSION

Table 1 lists 29 candidate OBP genes from *A. gambiae*. Twenty-five unique genes were identified through BLAST search analysis of the *A. gambiae* genome and are listed with the protein accession number (EAA#) and the corresponding gene scaffold (AAA#). Four pairs of identified sequences were considered to represent the same genes; small variations in nucleic acid and amino acid sequences and very low variation in third base positions of codons suggested that differences between members of each pair was either due to allelic variation or sequencing error. Six of the 25 sequences were previously reported from direct sequence analysis (Biessmann et al., 2002). Three additional sequences were also previously reported as *A. gambiae*

TABLE 1. ACCESSION NUMBERS OF *A. Gambiae* OBP CANDIDATES

1	EAA09615 (AAAB01008904) (<i>Agam</i> OS-F)	2	EAA03140 (AAAB01007923)
	EAA14641 (AAAB01008982)		EAA08879 (AAAB01008888)
	AF437884 (OBP-1 Biessmann et al., 2002)		AF437889 (OBP-6 Biessmann et al., 2002)
3	EAA01090 (AAAB01008987) (<i>Agam</i> OS-E)	4	EAA09958 (AAAB01008933)
	AF437886 (OBP-3 Biessmann et al., 2002)		AF437887 (OBP-4 Biessmann et al., 2002)
5	EAA14656 (AAAB01008982)	6	EAA01289 (AAAB01008987)
	AF437885 (OBP-2 Biessmann et al., 2002)		AF437890 (OBP-7 Biessmann et al., 2002)
7	EAA00498 (AAAB01008986)	8	EAA03447 (AAAB01008582)
	EAA03297 (AAAB01008194)		EAA09285 (AAAB01008898)
9	EAA00779 (AAAB01008986)	10	EAA06803 (AAAB01008847)
11	EAA00788 (AAAB01008986)	12	EAA07741 (AAAB01008859)
13	EAA00801 (AAAB01008986)	14	EAA07997 (AAAB01008859)
15	EAA01052 (AAAB01008987)	16	EAA09324 (AAAB01008900)
17	EAA01392 (AAAB01008987)	18	EAA10932 (AAAB01008960)
19	EAA01491 (AAAB01008987)	20	EAA12996 (AAAB01008966)
21	EAA03742 (AAAB01008799)	22	EAA13515 (AAAB01008976)
23	EAA03745 (AAAB01008799)	24	EAA14622 (AAAB01008982)
25	EAA06799 (AAAB01008847)		
26	AF393485	27	AF437888
	(OBP1, Robertson et al., Submission 2002)		(OBP-5 Biessmann et al., 2002)
28	AF393487	29	AF457552 (D7 Salivary Protein,
	(OBP2, Robertson et al., Submission 2002)		Francischetti et al., Submission 2002)

OBP. The *AgamD7* gene, encoding a salivary protein, was also identified as a candidate OBP homologue.

Figure 1A shows a sequence similarity tree assembled from an alignment of the 29 candidate *A. gambiae* OBP sequences, 38 *D. melanogaster* OBP sequences (Galindo and Smith, 2001; Graham and Davies, 2002; Vogt et al., 2002) and OBP related sequences from diverse other species. The long-branch lengths in this tree indicate the considerable sequence divergence between most of these sequences. Most sequences are in branches specific to a given insect group, though several similarity groups do include multiple insect groups. Three similarity groups include both *A. gambiae* and *D. melanogaster* sequences: OS-E/OS-F, LUSH, and PBPRP2/PBPRP5. The OS-E/OS-F group includes 1 sequence from *Culex quinquefasciatus* (Ishida et al., 2002) and 4 sequences from *A. gambiae*; this similarity group was further characterized.

OS-E and OS-F are the most similar of the *D. melanogaster* OBPs (Vogt et al., 2002). The OS-E/OS-F genes are adjacent within a 5000 nucleotide region of chromosome 3, have conserved exon/intron boundaries, and are co-expressed in a

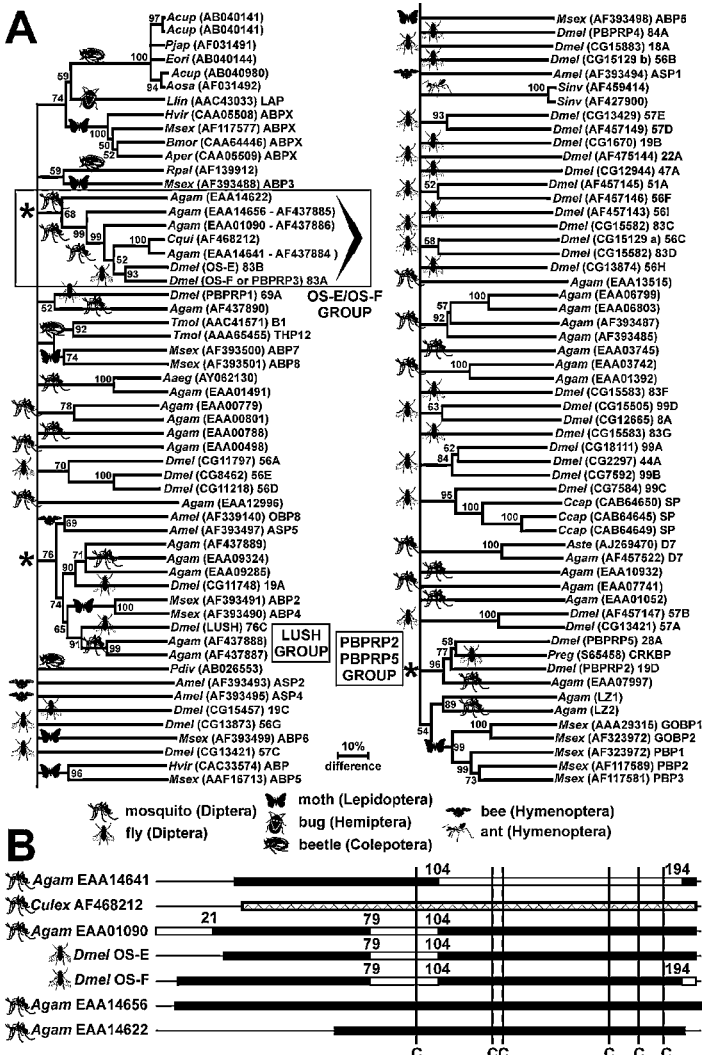


FIG. 1. **A.** A neighbor joining tree is shown; branches are collapsed to 50% bootstrap support based on 1000 replicates. The single tree was separated in two for ease of viewing. *AgamLZ1* and *AgamLZ2* are unpublished from Larry Zwiebel. **B.** A graphical representation of an amino acid alignment of proteins belonging to OS-E/OS-F similarity group (1A). The overall lengths represented include alignment gaps; the numbers refer to the amino acid position at the C-terminal side of a domain in the aligned sequence. Exon domains are indicated by alternating white-filled and black-filled; vertical lines indicate positions of conserved cysteins (C).

subset of olfactory sensilla in the adult antenna, and are, thus, assumed to be co-derived from a gene duplication event (Hekmat-Scafe et al., 2000; Vogt et al., 2002).

Figure 1B represents an amino acid alignment of the 7 proteins of the OS-E/OS-F group. All 7 proteins contain 6 cysteins that co-align. Exon boundaries were determined for each sequence (the *Culex* gene was not available). *DmelOS-E* and *DmelOS-F* have two conserved exon boundaries at positions 79 and 104. *DmelOS-F* has an additional short exon at the extreme C-terminal end (position 194) encoding 5 amino acids (HYFLP), which align with a similar sequence in *DmelOS-E* (HYFLV). *Agam*(EAA01090) also contains exon boundaries at alignment positions 79 and 104 as well as the C-terminal sequence HYFLP, but lacks a C-terminal exon. *Agam* (EAA14641) has an exon boundary at alignment position 104 and a C-terminal exon at position 194 with the sequence HYFLV, but lacks the position 79 boundary. *Agam* (EAA14656) and *Agam*(EAA14622) have only single exons. The genes of *Agam* (EAA14656), *Agam*(EAA14622), and *Agam*(EAA14641) reside in a cluster; *Agam* (EAA14656) and *Agam*(EAA14622) are separated from each other by about 3,000 base pairs (bp) and from *Agam*(EAA14641) by about 200,000 bp. *Agam*(EAA01090) is separated from this cluster by about 40 million bp.

The conservation of exon boundaries suggests the evolutionary history of these genes. The intron-containing *Agam*(EAA14641) and *Agam*(EAA01090) are presumably derived from a gene duplication based on sequence similarity and conservation of exon boundaries; the two genes presumably later became separated by a translocation event. The gene cluster of *Agam*(EAA14641), *Agam*(EAA14656), and *Agam*(EAA14622) likely derived through gene duplications, since the probability of related genes becoming located near each other by translocation is small given the overall size of the genome. If this cluster did derive through gene duplication, then *Agam*(EAA14641) is the likely founder since its relation to *DmelOS-E* and *DmelOS-F* suggests an ancestry that predates the divergence of flies and mosquitoes. It is not obvious how *Agam*(EAA14641) could have produced an intron free copy that then duplicated to *Agam*(EAA14656) and *Agam*(EAA14622); perhaps the *Agam*(EAA14641) transcript was captured by a retrovirus and reinserted into the genome with some homologous recombinatorial bias. This gene cluster is absent in *D. melanogaster* (Vogt et al., 2002) and, thus, may have arisen independently in mosquitos.

Agam(EAA01090) and *Agam*(EAA14641) are probable orthologues of *DmelOS-E* and *DmelOS-F*. Hekmat-Scafe et al. (2000) characterized these genes from

multiple *Drosophila* species and suggested OS-E/OS-F diverged early in the evolutionary history of flies. *Agam*(EAA01090) contains both the 79 and 104 exon boundaries, but lacks the 194 boundary that characterizes *DmelOS-F*. *Agam* (EAA14641) contains the position 194 exon boundary of *DmelOS-F* but lacks the position 79 boundary characteristic of both *DmelOS-E* and *DmelOS-F*. Assuming that loss of an existing intron is more probable than the random gain of an intron at a specific position, then the presence of the position 194 boundary suggests that *Agam* (EAA01090) and *DmelOS-E* are orthologues, and *Agam*(EAA01090) might be named *AgamOS-E*. Similarly, *Agam*(EAA14641) and *DmelOS-F* are apparent orthologues, but *Agam*(EAA14641) has independently lost one of its introns (position 79); *Agam*(EAA14641) might be named *AgamOS-F*. The duplication that established these two gene lineages apparently occurred before the divergence of mosquitoes and flies.

Not all members of the OBP gene family are involved in olfaction (Vogt et al., 2002), and no dipteran OBPs including OS-E and OS-F have actually been shown to bind odor molecules. However, the demonstrated expression of OS-E and OS-F in olfactory sensilla of *Drosophila* (Hekmat-Scafe et al., 1997) and the structural similarities of these two proteins to OBPs that have been shown to bind odors supports their role in odor detection. The conservation of these genes between mosquito and fly suggests OS-E and OS-F have a singularly important function in both species groups.

Acknowledgments—Support to RGV by National Science Foundation IBN0212510 to RGV.

REFERENCES

- 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. *Anopheles gambiae*. *Insect Mol. Biol.* 11: 123–132.
- DE BRUYNE, M., FOSTER, K., AND CARLSON, J. R. 2001. Odor coding in the *Drosophila* antenna. *Neuron* 30: 537–552.
- GALINDO, K. AND SMITH, D. P. 2001. A large family of divergent *Drosophila* odorant-binding proteins expressed in gustatory and olfactory sensilla. *Genetics* 159: 1059–1072.

- GRAHAM, L. A. AND DAVIES, P. L. 2002. The odorant-binding proteins of *Drosophila melanogaster*: annotation and characterization of a divergent gene family. *Gene* 292: 43–55.
- HEKMAT-SCAFE, D. S., STEINBRECHT, R. A., AND CARLSON, J. R. 1997. Coexpression of two odorant-binding protein homologs in *Drosophila*: implications for olfactory coding. *J. Neurosci.* 17: 1616–1624.
- HEKMAT-SCAFE, D. S., DORIT, R. L., AND CARLSON, J. R. 2000. Molecular evolution of odorant-binding protein genes OS-E and OS-F in *Drosophila*. *Genetics* 155: 117–127.
- ISHIDA, Y., CORNEL, A. J., AND LEAL, W. S. 2002. Identification and cloning of a female antenna-specific odorant-binding protein in the mosquito *Culex quinquefasciatus*. *J. Chem. Ecol.* 28: 867–871.
- 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.
- 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 odorant-binding proteins in *Drosophila* olfactory organs. *Microsc. Res. Tech.* 55: 297–306.
- VOGT, R. G., CALLAHAN, F. E., ROGERS, M. E., AND DICKENS, J. C. 1999. Odorant Binding Protein Diversity and Distribution among the Insect Orders, as indicated by LAP, an OBP-related protein of the True Bug *Lygus lineolaris* (Hemiptera, Heteroptera). *Chemical Senses* 24: 481–495.
- VOGT, R. G., ROGERS, M. E., FRANCO, M. D., AND SUN, M. 2002. A comparative study of odorant binding protein genes: differential expression of the PBP1-GOBP2 gene cluster in *Manduca sexta* (Lepidoptera) and the organization of OBP genes in *Drosophila melanogaster* (Diptera). *J. Exp. Biol.* 205: 719–744.