5-1-2000

Molecular Evolution of Odorant-Binding Protein Genes OS-E and OS-F in Drosophila

Daria S. Hekmat-Scafe
Yale University

Robert L. Dorit
Yale University, rdorit@smith.edu

John R. Carlson
Yale University

Molecular Evolution of Odorant-Binding Protein Genes *OS-E* and *OS-F* in Drosophila

Daria S. Hekmat-Scafe,*,† Robert L. Dorit‡ and John R. Carlson*

**Department of Molecular, Cellular, and Developmental Biology, ‡Department of Ecology and Evolutionary Biology, Yale University, New Haven, Connecticut 06520-8103 and †Biology Department, Mills College, Oakland, California 94613*

Manuscript received August 2, 1999
Accepted for publication January 6, 2000

ABSTRACT

The Drosophila olfactory genes *OS-E* and *OS-F* are members of a family of genes that encode insect odorant-binding proteins (OBPs). OBPs are believed to transport hydrophobic odorants through the aqueous fluid within olfactory sensilla to the underlying receptor proteins. The recent discovery of a large family of olfactory receptor genes in Drosophila raises new questions about the function, diversity, regulation, and evolution of the OBP family. We have investigated the *OS-E* and *OS-F* genes in a variety of Drosophila species. These studies highlight potential regions of functional significance in the OS-E and OS-F proteins, which may include a region required for interaction with receptor proteins. Our results suggest that the two genes arose by an ancient gene duplication, and that in some lineages, one or the other gene has been lost. In *D. virilis*, the *OS-F* gene shows a different spatial pattern of expression than in *D. melanogaster*. One of the *OS-F* introns shows a striking degree of conservation between the two species, and we identify a putative regulatory sequence within this intron. Finally, a phylogenetic analysis places both OS-E and OS-F within a large family of insect OBPs and OBP-like proteins.

INSECT odorant-binding proteins (OBPs) are small, soluble proteins found in the aqueous lymph that fills the olfactory sensilla on the antenna (Pelosi and Maida 1995). OBPs are believed to shuttle hydrophobic odorants through the sensillar lymph to the underlying odorant receptors (Vogt *et al.* 1991; Pelosi 1994; Prestwich *et al.* 1995). Each insect species contains multiple, distinct OBPs, which are often sequestered within distinct subsets of olfactory sensilla (Steinbrecht *et al.* 1995; Steinbrecht 1996). Since different OBPs may display different odorant specificities (Du and Prestwich 1995; Prestwich *et al.* 1995), OBPs may play a role in olfactory coding; in particular, the OBP(s) present in a particular olfactory sensillum might determine the spectrum of odorants to which the underlying receptors have access.

Recently a large family of candidate odorant receptors has been identified in Drosophila (Clyne *et al.* 1999b; Vosshall *et al.* 1999). The family contains on the order of 100 genes predicted to encode a highly divergent family of seven-transmembrane-domain proteins. A number of these genes has been shown to be expressed in subsets of olfactory neurons, and some of the genes are not expressed in a mutant in which some of the olfactory neurons show abnormal odor specificities (Clyne *et al.* 1999a). The large number of receptor genes, their restricted expression in subsets of olfactory neurons, and their high degree of sequence divergence raise questions about their relationship to OBPs. If individual OBPs interact with specific receptors, has the OBP gene family evolved in concert with the receptor gene family? Given the remarkable sequence divergence among receptors, do individual OBPs contain a highly divergent domain that promotes interactions with individual receptors? Has the spatial regulation of OBPs evolved to allow interactions between specific OBPs and specific receptors or sets of receptors?

The genome of the fruit fly *Drosophila melanogaster* encodes a variety of predicted OBPs, each of which is expressed in a characteristic portion of the antenna (McKenna *et al.* 1994; Pikielny *et al.* 1994; Kim *et al.* 1998). Although most of these OBPs differ markedly from one another, those encoded by the olfactory-specific genes *OS-E* and *OS-F* show substantial sequence similarity: 69% amino acid identity for the mature proteins (McKenna *et al.* 1994). The *OS-E* and *OS-F* genes are located <1 kb apart and were suggested to have arisen by gene duplication (Hekmat-Scafe *et al.* 1997). They are coexpressed within two morphological types of olfactory sensilla that are located in the same region of the antenna, the ventrolateral region.

To address the functional significance and evolution of the OS-E and OS-F proteins in Drosophila, we examined *OS-E* and *OS-F* gene homologues in a variety of Drosophila species, with particular emphasis on *D. virilis*, a species thought to have shared a common ancestor with *D. melanogaster* $\sim$40 million years ago (mya; Powell and DeSalle 1995; Russo *et al.* 1995). Our analysis uncovered an *OS-F* homologue in *D. virilis*, but no *D.*

Corresponding author: John R. Carlson, Department of Molecular, Cellular, and Developmental Biology, Yale University, P.O. Box 208103, New Haven, CT 06520-8103. E-mail:john.carlson@yale.edu

virilis counterpart to the *OS-E* gene. *D. melanogaster* and *D. virilis* OS-F proteins show remarkable conservation but diverge notably in two regions: the N terminus and a C-terminal region that exhibits heterogeneity in other insect OBPs. We observed that *OS-F* transcripts are expressed in a different spatial pattern within the antenna of *D. virilis* than in *D. melanogaster*, possibly reflecting the presence of this OS-F protein in an additional class of olfactory sensilla in *D. virilis*. We find that the *OS-F* intron shows a surprisingly high degree of sequence conservation, and we identify a putative regulatory element within it.

Our examination of *OS-E* and *OS-F* homologues in a variety of Drosophila species suggests that the duplication that gave rise to *OS-E* and *OS-F* is an ancient one. These studies also highlight regions of potential functional importance in the OS-E and OS-F proteins, one of which might mediate binding to odorant receptor proteins. Finally, our phylogenetic analysis illustrates that OS-E and OS-F are members of a diverse and ancient family of OBP-related insect proteins.

MATERIALS AND METHODS

Drosophila stocks and culture: *D. virilis*, *D. simulans*, *D. mauritiana*, *D. teissieri*, and *D. willistoni* flies were kindly provided by Jeffrey Powell (Yale University); *D. lebanonensis* flies were obtained from the National Drosophila Species Resource Center (NDRSC, Bowling Green, OH). *D. virilis* flies were grown at 18° on standard molasses-corn meal medium (Ashburner 1989).

Isolation of *OS-E* and *OS-F* homologues from different Drosophila species: Genomic DNA was isolated from aliquots of ~30 flies essentially as described (Raha *et al.* 1994). PCR amplification was performed on an ~1/2-fly equivalent of DNA. The 5′ primer (EF3), corresponding to the sequence CY(M/I)NC, was 5′ CGGAATTCTG(T/C)TA(T/C)ATIAA(T/C)TG, and the 3′ primer (EF11), corresponding to the sequence CHKAWW, was 5′ GCTCTAGACCACCAIGC(C/T)T T(A/G)TG(A/G)C. (5′ *Eco*RI and *Xba*I sites, respectively, are underlined.) In pilot experiments that led to the isolation of *D. lebanonensis OS-E1*, a longer 3′ primer (EF7), corresponding to the sequence CHKAWWFHQC, was used [5′ CA(C/T)TG(A/G)TG(A/G)AACCACCAIGC(C/T)TT(A/G)TG(A/G)C]. The PCR conditions were the following: 94° for 5 min, followed by 35 cycles of 54° for 1 min, 72° for 2 min, and 94° for 1 min, and then one cycle of 54° for 1 min and 72° for 10 min. In some experiments (those leading to the isolation of *D. lebanonensis OS-E1* and *E2* and *D. teissieri OS-F*) the 35 cycles described above were preceded by 3 cycles of 37° for 1 min, 72° for 2 min, and 94° for 1 min. AmpliTaq DNA polymerase (Perkin-Elmer Biosystems, Norwalk, CT) was used for all PCR reactions.

Purified PCR products were digested with *Eco*RI and *Xba*I and subcloned into pBluescript II (Stratagene, La Jolla, CA). DNA sequencing of both strands of the inserts in these plasmids was performed with the T3 and T7 primers using the Sequenase 2.0 kit (U.S. Biochemicals Corp., Cleveland) according to the manufacturer's instructions.

The procedures described above led to the identification of *D. simulans OS-F*, *D. mauritiana OS-F*, *D. teissieri OS-E*, *D. virilis OS-F*, *D. lebanonensis OS-E1*, and both *D. willistoni OS-E* and *F*. In subsequent experiments designed to determine the number of *OS-E*/*OS-F*-related genes in each of the Drosophila species, the purified ~170-bp PCR products were digested, in separate reactions, with two restriction enzymes that had a six-base-pair recognition site within the sequenced *OS-E* or *F* gene. In *D. simulans*, *D. mauritiana*, and *D. virilis*, no uncut PCR product was discernible in either digest, consistent with the notion that these species possess no additional *OS-E* or *F* genes. In cases where an original-sized PCR product remained after one or both restriction digests (*D. teissieri* and *D. lebanonensis*), we surmised that an additional *OS-E* or *F* gene was likely present.

The uncut *D. teissieri* and *D. lebanonensis* PCR products were purified from a low-melting-point agarose (FMC)/Tris acetate EDTA gel treated with β-agarase (New England Biolabs, Beverly, MA) and subjected to sequence analysis on both strands at the W. M. Keck Foundation Biotechnology Resource Laboratory at Yale University. The sequencing primers were 5′ tes (5′CGGAATTCTGCTACATGAACTG), 3′ tes (5′GCTCTAGACCACCAGGCTTTGTGAC), 5′ leb (5′CGGAATTCTGCTATATGAATTG), and 3′ leb (5′GCTCTAGACCACCAGGCCTTATGAC). In both cases, a single DNA sequence, which was distinct from the one previously identified in that species, was obtained. We note that since the various Drosophila species stocks were not isogenic, polymorphic variants carrying one or a small number of base pair changes may exist for *OS-E* and/or *OS-F* sequences we obtained from some of these species; all of the genes characterized in this study differ substantially from each other.

Isolation and analysis of the *D. virilis OS-F* genomic locus: The PCR products amplified from *D. virilis* DNA using the EF3 and EF11 primers were labeled with ^{32}P using a modified hexamer-labeling procedure (Feinberg and Vogelstein 1983) wherein the EF3 and EF11 primers were substituted for random hexamers. This labeled *D. virilis E*/*F* probe was used to screen a library of *D. virilis* genomic DNA in λEMBL3 (a gift of Ron Blackman, University of Illinois). Hybridization and washes for the screen were as described previously (McKenna *et al.* 1994).

Two overlapping clones (λV1 and λV2) were identified and subjected to restriction analysis (Sambrook *et al.* 1989). A 5.8-kb *Sal*I fragment of λV2 shown by Southern analysis (Sambrook *et al.* 1989) to contain all sequences in λV2 that hybridized with the *D. virilis E*/*F* probe was subcloned into *Bam*HI-cut pGEM7zf(+) (Promega, Madison, WI) as described (Hung and Wensink 1984) to create pDH117. Initial sequence information was obtained from pDH117 using the EF11 primer and the Sequenase 2.0 kit (Stratagene) and used to design additional sequencing primers. Subsequent sequence analysis of both strands was done by the W. M. Keck Foundation Biotechnology Resource Laboratory at Yale University.

An 800-bp *Dra*I/*Sac*I fragment of pDH117 (carrying the *D. virilis OS-F* coding region excluding the last five codons, along with 180 bp of 5′ noncoding DNA) was subcloned into pGEM7zf(+) to create pDH137. The *D. virilis* insert in pDH137 was hexamer-labeled (Feinberg and Vogelstein 1983) and used to probe a Southern blot (Sambrook *et al.* 1989) of *D. virilis* DNA under conditions of both high and low stringency. High-stringency conditions were as described (McKenna *et al.* 1994). For low-stringency conditions, the hybridization buffer contained 30%, rather than 50%, formamide, and the washes were done in 1× SSPE/0.1% SDS. *D. melanogaster OS-E* and *OS-F* are known to cross-hybridize under these same low-stringency conditions (D. Hekmat-Scafe and J. Carlson, unpublished data).

Antisense and sense DIG-RNA probes (Boehringer-Mannheim, Indianapolis) for *D. virilis OS-F* prepared from pDH137 were used to examine *D. virilis OS-F* expression in *D. virilis*

heads by *in situ* hybridization (Tautz and Pfeifle 1989). The hybridization conditions (McKenna *et al.* 1994) were those under which *D. melanogaster OS-E* and *OS-F* do not cross-hybridize.

Computational analysis: The ClustalW package (Higgins and Sharp 1988) was used to align OS-E and OS-F amino acid sequences derived from *OS-E* and *OS-F* PCR products amplified from the various Drosophila species. This alignment was used to construct both a parsimony and distance (neighbor-joining) tree with the phylogenetic analysis using Parsimony (PAUP* 4.0b2 test version) package (Swofford 1999). PBPRP-1 (pheromone-binding protein related protein 1; Pikielny *et al.* 1994) was used as an outgroup to root the trees. A majority rule consensus parsimony tree was obtained for 16 equivalent trees found by a heuristic search using tree-bisection-reconnection with random sequence addition (100 replicates). Bootstrap values were obtained from 1000 replicates using the same algorithm. A majority rule consensus distance tree, derived from 3 equivalent length trees, was also obtained using tree-bisection-reconnection.

A list of 49 insect OBPs and related proteins was obtained by performing four iterations of a Psi-Blast search of the Non-Redundant GenBank CDS starting with the OS-E protein sequence. The 49 corresponding amino acid sequences were then extracted from GenBank. These sequences were subsequently aligned using the ClustalW package (Higgins and Sharp 1988), and the resulting protein alignments inspected to ensure that the landmark cysteine residues were properly aligned. To prevent overweighing of insertion/deletion events, gaps were recorded as single events (regardless of gap length) for subsequent analyses.

This alignment was used to construct a phylogenetic network on the basis of either maximum parsimony or distance (neighbor-joining) with the PAUP* 4.0b2 test version package (Swofford 1999). Unrooted parsimony trees were created using the heuristic search algorithm; 1000 replicate runs were performed with stepwise random sequence addition. All general odorant-binding proteins (GOBP) and pheromone-binding proteins (PBP) sequences were designated as "outgroups" and everything else as "ingroup." Bootstrap values were assigned based on the partition functions obtained from 1000 replicate runs using the tree-bisection-reconnection option for branch-swapping and random stepwise taxon addition. The distance tree was generated using the standard algorithm and employing mean character difference as the standardized measure. Ties were resolved randomly.

RESULTS

The *OS-F* gene of *D. virilis*, its organization, and a highly conserved intron: We have identified a single *D. virilis* gene whose product shows extensive sequence similarity to *D. melanogaster OS-F*. Briefly, we used two PCR primers—one corresponding to an amino acid sequence present in many insect OBPs and the other a more specific primer corresponding to a sequence present in *D. melanogaster OS-E* and *OS-F*—to amplify a product from *D. virilis* genomic DNA. The amplified sequence was then radiolabeled and used to probe a library of *D. virilis* genomic DNA. We thereby isolated genomic sequences that encompass the *D. virilis OS-F* gene, along with ~7 kb of upstream DNA and ~14 kb of downstream DNA.

Restriction maps of the *D. virilis OS-F* genomic locus and the corresponding genomic region in *D. melanogaster* are shown in Figure 1A. In contrast to *D. melanogaster*, in which a related gene, *OS-E*, is present <1 kb upstream of *OS-F*, *D. virilis* has only a single OBP-related gene in the region, *OS-F*. Both low- and high-stringency Southern hybridization of *D. virilis OS-F* sequences to four different restriction digests of *D. virilis* DNA revealed single bands (Figure 1B). Complete sequence analysis of *D. virilis* genomic DNA corresponding to the smallest of these bands, the ~1-kb Dra I fragment, revealed no additional *OS-F*-related sequences. Hence, *D. virilis* apparently has only one gene closely related to *D. melanogaster OS-F*.

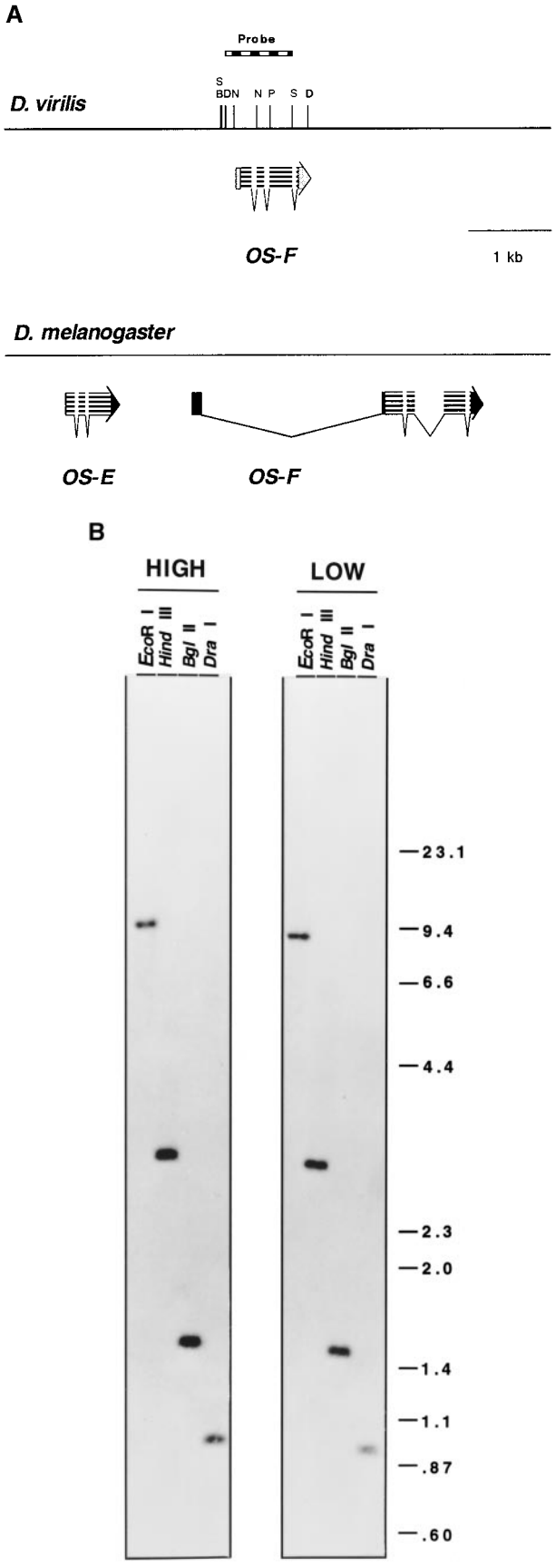
An amino acid sequence alignment of *D. virilis* OS-F with *D. melanogaster* OS-E and OS-F is shown in Figure 2A. All three are small proteins, with a predicted primary translation product of ~16–17 kD. All carry an N-terminal signal sequence, and all have the six aligned cysteine residues that are diagnostic of insect OBPs (Pelosi and Maida 1995).

D. virilis OS-F protein shows 76% sequence identity to *D. melanogaster* OS-F and 57% identity to *D. melanogaster* OS-E (Table 1). The sequence identity between *D. virilis* and *D. melanogaster* OS-F is greater than that seen when comparing *D. melanogaster* OS-E and OS-F (62%).

Two regions of the protein exhibit a high degree of sequence divergence. The majority of the amino acid substitutions between mature *D. virilis* and *D. melanogaster* OS-F occur in a 22-amino-acid stretch, which we call the "heterogeneous region" and which extends from L^{107} to H^{128} in *D. virilis* OS-F (Figure 2A). *D. virilis* and *D. melanogaster* OS-F are only 55% identical within this 22-amino-acid stretch, whereas the remaining portions of the mature proteins are 86% identical (Table 1). *D. melanogaster* OS-E and OS-F show an even greater degree of heterogeneity in this region: 39% identity, as compared to 76% identity in the remaining portions of the mature proteins. A second region of heterogeneity is at the N terminus. Much of the N-terminal heterogeneity resides within the signal sequence, but the N-terminal region of the mature proteins is predicted to exhibit substantial heterogeneity as well.

The *D. virilis* and *D. melanogaster OS-F* genes display a similar intron-exon structure (Figures 1A and 2A). There are three small introns within the *D. virilis OS-F* coding region. They are located between N^{43} and Y^{44} (76 bp), between E^{68} and A^{69} (78 bp), and between K^{154} and H^{155} (67 bp) (Figure 2A). These three introns are present at positions corresponding to those of the three introns within the coding region of *D. melanogaster OS-F*; the first two of these intron insertion sites also correspond to those of the two introns in *D. melanogaster OS-E* (Hekmat-Scafe *et al.* 1997).

The first intron in the coding region of *D. virilis OS-F* shows a surprisingly high degree of nucleotide sequence identity to that of *D. melanogaster OS-F* (Figure 2B). Overall, the two introns are 76% identical. This similarity



suggests that the intron harbors conserved regulatory sequences needed for appropriate gene expression in the olfactory system. One possibility for such a regulatory element is the sequence GCCACGC, which is also present in the first intron within the coding region of the pheromone-binding protein-related protein, *PBPRP-1* (data not shown). *PBPRP-1* encodes a predicted OBP, which, like *OS-F*, is expressed in regions of the *D. melanogaster* antenna rich in trichoid sensilla (Pikielny *et al.* 1994).

The spatial regulation of *D. virilis OS-F* is different from that of *D. melanogaster OS-F*: *In situ* hybridization revealed that *D. virilis OS-F* transcripts are expressed predominantly, perhaps exclusively, in the antenna (Figure 3 and data not shown), as has been observed previously for *D. melanogaster OS-E and OS-F* (McKenna *et al.* 1994). Interestingly, the distribution of *D. virilis OS-F* transcripts within the antenna is different from that observed for *D. melanogaster OS-E* and *OS-F*. *D. melanogaster OS-E* and *OS-F* are coexpressed specifically in the ventrolateral region of the antenna in a pattern that mimics the distribution of one morphological class of sensory hairs, the trichoid sensilla (McKenna *et al.* 1994; Hekmat-Scafe *et al.* 1997). However, an antisense probe for *D. virilis OS-F* mRNAs shows a broader distribution, extending to include a portion of the antenna immediately ventral to the sacculus, a chamber lined with sensory hairs (Figure 3, A and B). Visual inspection by light microscopy revealed no major differences in the distribution of trichoid sensilla between the two species (D. Hekmat-Scafe and K. Störtkuhl, unpublished results), suggesting the possibility that the distribution of OS-F among the different morphological classes of sensilla is different between the two species. No hybridization was observed with the *D. virilis OS-F* sense probe (Figure 3C).

Figure 1.—The *OS-F* gene in *D. virilis.* (A) Genomic organization of the *D. virilis OS-F* gene (top) contrasted with that of *D. melanogaster OS-E* and *OS-F* (bottom). All of the genes are transcribed from left to right, as indicated by the arrows. Exons are indicated by thick boxes and introns by thin lines connecting these boxes. Within exons, coding regions are striped, whereas noncoding sequences are black (*D. melanogaster*) or gray (*D. virilis*). The 5′ and 3′ untranslated regions of *D. virilis OS-F* have not been characterized in detail, and the sizes of the gray boxes are designated arbitrarily. Positions of restriction enzyme recognition sites in genomic DNA encompassing the *D. virilis OS-F* gene are shown. The *Dra*I-*Sac*I fragment used as a probe for the Southern blot of *D. virilis* genomic DNA shown in B is indicated. S, *Sac*I; B, *Bgl*II; D, *Dra*I; N, *Nsi*I; P, *Pst*I. (B) Southern analysis of the *D. virilis OS-F* gene. *D. virilis* genomic DNA was digested with *Eco*RI, *Hin*dIII, *Bgl*II, or *Dra*I as indicated and subjected to 0.6% agarose-Tris borate EDTA electrophoresis and Southern transfer. Filters were probed with a ^{32}P-labeled *Dra*I-*Sac*I fragment (shown in A) carrying the *D. virilis OS-F* gene under either high (left) or low (right) stringency conditions. Positions of molecular weight markers are shown on the right.

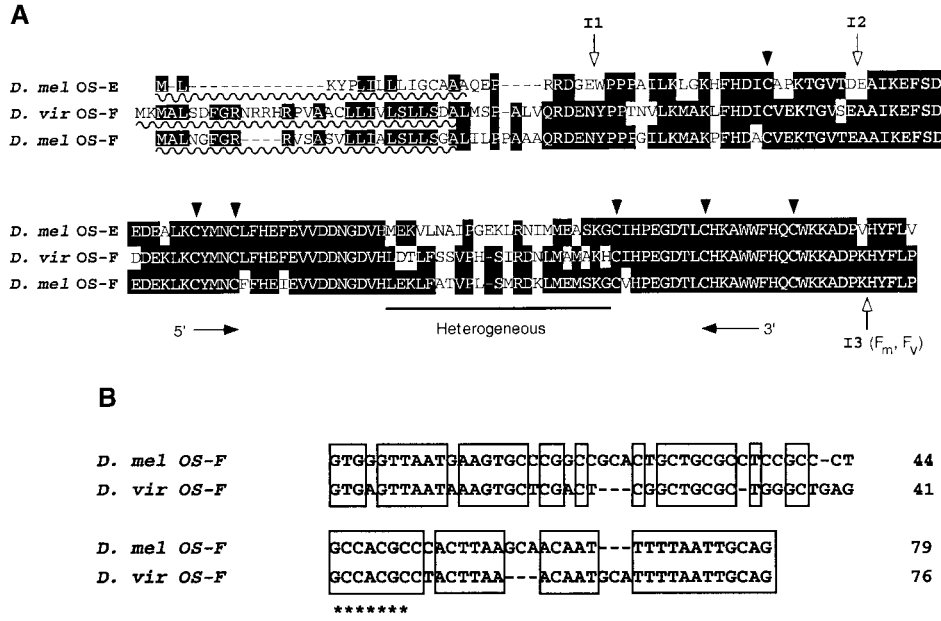


Figure 2.—Comparison of *D. virilis OS-F* with *OS-E* and *OS-F* of *D. melanogaster*. (A) Amino acid sequence alignment between *D. melanogaster* OS-E (top) and OS-F (bottom) and *D. virilis* OS-F (middle). Amino acid identities are indicated by black boxes. Positions of six conserved cysteine residues, which are present in all known insect OBPs, are highlighted by arrowheads. The locations of three introns located within the coding regions of the corresponding genes are indicated by arrows. The first two of these introns (I1 and I2) are located at corresponding positions in all three genes, whereas a third intron (I3) is found only in *D. melanogaster* and *D. virilis OS-F* (F_m, F_v)—at corresponding positions. Positions of likely *D. virilis OS-F* introns were predicted by comparison of the *D. virilis OS-F* genomic sequence to that of *D. melanogaster OS-F* and verification of consensus splice sequences at the predicted splice sites. Predicted N-terminal signal sequences, present in all three proteins, are represented by wavy lines. A stretch of 22–23 amino acids that are particularly heterogeneous in the three proteins is indicated. Positions of the 5′ and 3′ PCR primers used to amplify *OS-E* and *OS-F* genes from a variety of Drosophila species are shown as horizontal arrows. (B) The first introns within the coding regions of *D. melanogaster* and *D. virilis OS-F* demonstrate a high degree of sequence identity. Identical nucleotides are boxed. Overall, the two introns are 76% identical. Asterisks indicate a sequence also found in the first coding region intron of *PBPRP-1*, another predicted OBP gene of *D. melanogaster*.

Analysis of *OS-E* and *OS-F* homologues in different Drosophila species: To expand our analysis of *OS-E* and *OS-F* genes, we isolated *OS-E* and *OS-F* homologues from a wide variety of Drosophila species. Specifically, we amplified a ~170-bp fragment of *OS-E*- and/or *OS-F*-related genes from the genomic DNA of other Drosophila species, using PCR conditions similar to those used to amplify the *D. virilis OS-F* sequences. The amplification products extend between the amino acids corresponding to C^{68} and W^{125} of *D. melanogaster OS-E*. We chose to analyze this region because it includes the heterogeneous stretch of 22–23 amino acids, which represents a region of great sequence divergence both between OS-F in different species (*D. melanogaster* and *D. virilis*) and between OS-E and OS-F in *D. melanogaster*.

Parsimony analysis of OS-E- and OS-F-related protein sequences from the various Drosophila species yielded 16 minimum-length trees (length = 87; confidence interval (CI) = 0.897; retention index (RI) = 0.816). The strict consensus of these is shown in Figure 4A. This tree groups OS-E-related proteins in one cluster and OS-F-related proteins in a sister cluster. Interestingly, in three species, *D. simulans*, *D. mauritiana*, and *D. virilis*, we have found an *OS-F* gene but not an *OS-E* gene, and in one species, *D. lebanonensis*, we have found two genes closely related to *OS-E* (which we named *OS-E1* and *OS-*

TABLE 1

Amino acid sequence identity shared between regions of *D. virilis* OS-F, *D. melanogaster* OS-F, and *D. melanogaster* OS-E

	% amino acid identity			
Comparison	Overall	Mature protein	Mature protein–"Het" region	"Het" region
D. vir OS-F *vs. D. mel* OS-F	76	81	86	55
D. vir OS-F *vs. D. mel* OS-E	57	66	75	17
D. mel OS-E *vs. D. mel* OS-F	62	70	76	39

The heterogeneous ("Het") region corresponds to a 22-amino-acid stretch extending from L^{107} to H^{128} in *D. virilis* OS-F, shown in Figure 2A.

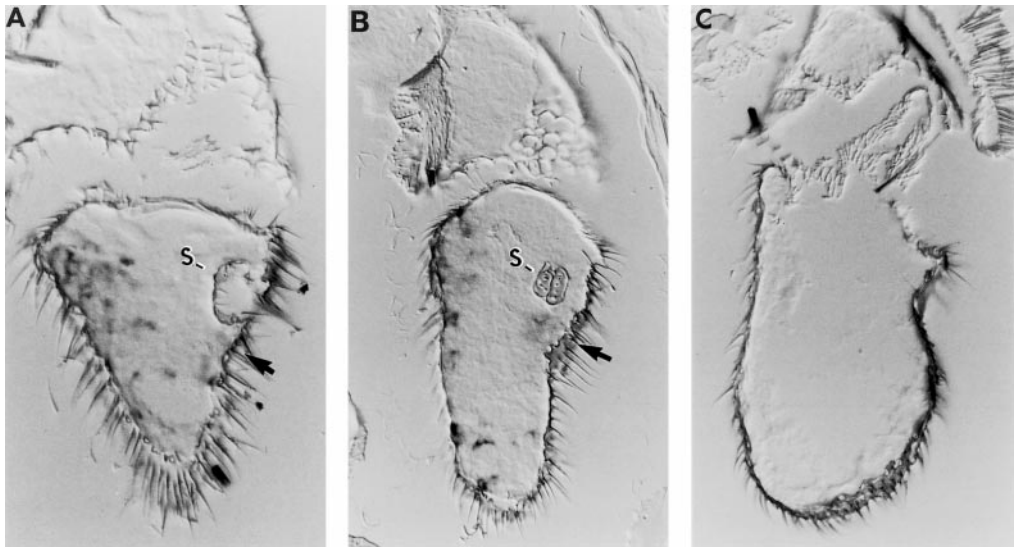


Figure 3.—*In situ* analysis of *D. virilis OS-F* transcripts. *D. virilis* antennal sections were hybridized with a *D. virilis OS-F* antisense (A, B) or sense (C) probe. Ventral is at the bottom and dorsal is at the top. The *OS-F* antisense probe shows punctate labeling in the antenna (A, B), including a region (arrow) beneath the pit-like sacculus (s). No discernible labeling is observed with the *OS-F* sense strand control (C).

E2), but none to *OS-F*. Two other species, *D. teissieri* and *D. willistoni*, are like *D. melanogaster* in that they contain both an *OS-E* gene and an *OS-F* gene. To test further the notion that three species lack an *OS-E* gene, we subjected the PCR products amplified from each of these species to restriction analysis and found no evidence for heterogeneity, as if they represented a unique, *OS-F* species. The simplest interpretation of all these results taken together is that not all Drosophila species carry both an *OS-E* and an *OS-F* gene.

Figure 4B shows the distribution of *OS-E* and/or *OS-F* gene(s) superimposed upon a phylogenetic tree for these species (Russo *et al.* 1995). *D. lebanonensis*, which diverged from *D. melanogaster* ~40 mya (Russo *et al.* 1995), also has two *E*/*F* genes, although both of them are closely related to *OS-E* (*OS-E1* and *OS-E2*). *D. teissieri* and *D. willistoni*, which diverged from *D. melanogaster* more recently (Powell and DeSalle 1995; Russo *et al.* 1995), have both an *OS-E* and an *OS-F* gene. By contrast, we identified an *OS-F* gene but found no evidence for an *OS-E* gene in *D. simulans* and *D. mauritiana*, which diverged from *D. melanogaster* more recently than did *D. teissieri* and *D. willistoni* (Russo *et al.* 1995), as well as in *D. virilis*, which diverged from *D. melanogaster* less recently than did *D. teissieri* and *D. willistoni* (Powell and DeSalle 1995).

An alignment of the various Drosophila OS-E and OS-F protein sequences (Figure 4C) reveals the marked sequence conservation of these proteins. In particular, all contain the motif -HPEGDTL following the fourth conserved cysteine, suggesting that this region is functionally important in both the OS-E and OS-F proteins of Drosophila.

In contrast, certain amino acid residues appear to distinguish OS-E from OS-F unambiguously. The serine at position 27 is present in all OS-F, but no OS-E sequences. Similarly, three residues (G^{25}, L^{28}, and I^{31}) are present in all OS-E, but not in OS-F sequences. An additional residue, N^{21}, is present in all OS-E sequences except for *D. lebanonensis* OS-E2. These diagnostic residues may underlie OS-E- or OS-F-specific functions.

Phylogenetic analysis of the insect OBP family: We have also carried out a broader phylogenetic analysis of the insect OBP family. The maximum-parsimony tree shown in Figure 5A represents a strict consensus tree of 38 trees of length 2605. A distance neighbor-joining network tree is shown in Figure 5B. The two trees represent a hypothesis of relationship among insect OBPs and OBP-like proteins for which complete sequences are available.

Both methods of analysis reveal two major clusters. The first major cluster (Figure 5, top) includes the moth OBPs. It has two major subdivisions, corresponding to the various moth PBPs and GOBPs, respectively. The second major cluster (Figure 5, bottom) corresponds to all other insect OBPs and related proteins. In both trees, OS-E and OS-F are grouped with antennal proteins from a large variety of insect species. These include PBPRP-1 from *D. melanogaster* (Pikielny *et al.* 1994); Rpa12 and Rpa12′, two closely related presumptive OBPs from the beetle *Rhynchophorus palmarum* (GenBank accession nos. AF141865 and AF139912); antennal binding proteins of unknown function (ABPXs) from a variety of species of moth (Krieger *et al.* 1997); closely related PBPs from the beetles *Popillia japonica* and *Anomala osakana* (Wojtasek *et al.* 1998); and LAP, an antennal protein from the hemipteran *Lygus lineolaris* (Vogt *et al.* 1999).

Also included in the second major cluster are a number of proteins expressed in tissues other than the olfactory organs. These include sericotropin, which is present in the brain of the wax moth *Galleria mellonella* (Kodrik *et al.* 1995), the B1 and B2 proteins, which are present in the secretions of the male accessory sex gland of the beetle *Tenebrio molitor* (Paesen and Happ 1995), the *T. molitor* antifreeze protein precursor (GenBank accession no. U24237), the male-specific protein MSSP, which is present in the hemolymph of the medfly *Cerati-*

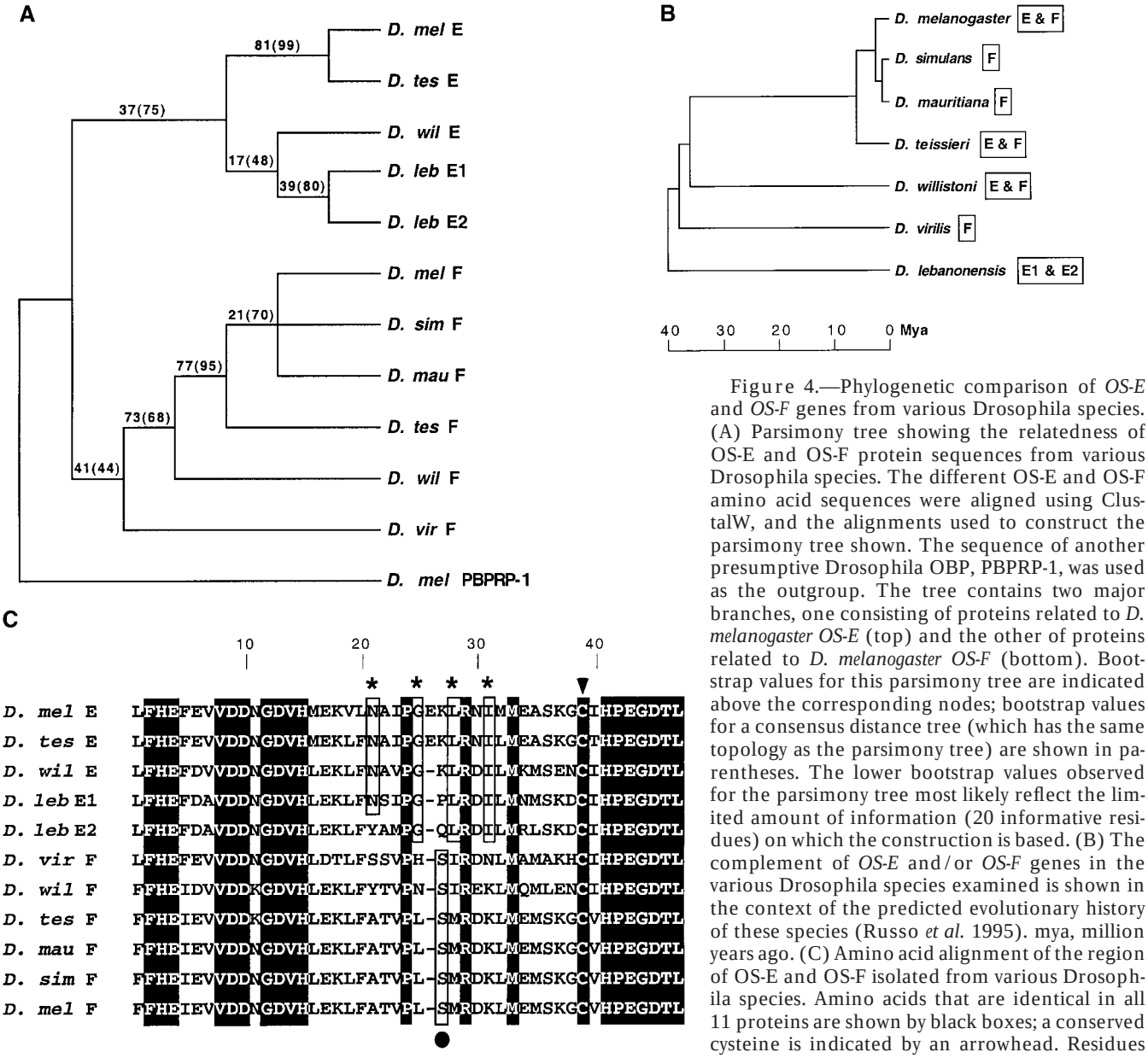


Figure 4.—Phylogenetic comparison of *OS-E* and *OS-F* genes from various Drosophila species. (A) Parsimony tree showing the relatedness of OS-E and OS-F protein sequences from various Drosophila species. The different OS-E and OS-F amino acid sequences were aligned using ClustalW, and the alignments used to construct the parsimony tree shown. The sequence of another presumptive Drosophila OBP, PBPRP-1, was used as the outgroup. The tree contains two major branches, one consisting of proteins related to *D. melanogaster OS-E* (top) and the other of proteins related to *D. melanogaster OS-F* (bottom). Bootstrap values for this parsimony tree are indicated above the corresponding nodes; bootstrap values for a consensus distance tree (which has the same topology as the parsimony tree) are shown in parentheses. The lower bootstrap values observed for the parsimony tree most likely reflect the limited amount of information (20 informative residues) on which the construction is based. (B) The complement of *OS-E* and/or *OS-F* genes in the various Drosophila species examined is shown in the context of the predicted evolutionary history of these species (Russo *et al.* 1995). mya, million years ago. (C) Amino acid alignment of the region of OS-E and OS-F isolated from various Drosophila species. Amino acids that are identical in all 11 proteins are shown by black boxes; a conserved cysteine is indicated by an arrowhead. Residues that are OS-E- or OS-F-specific are boxed. Asterisks indicate those residues present in all of the OS-E sequences (with the exception of N^{21}, which is absent from *D. lebanonensis* OS-E2) but in none of the OS-F sequences, whereas a solid circle indicates S^{27}, which is found in all of the OS-F sequences but in none of the OS-E sequences. Primer-derived sequences (which were identical in all amplified products) are omitted. The positions of the 5′ and 3′ PCR primers used to amplify *OS-E* and *OS-F* gene fragments from the various Drosophila species are shown in the context of the entire *OS-E* and *OS-F* genes in Figure 2A.

tis capitata (Thymianou *et al.* 1998), a variety of D7-related proteins, which are found in saliva of the mosquitoes *Anopheles gambiae* and *Aedes aegypti* (Arca *et al.* 1999), and the SL1 protein, which is present in the saliva of the fly *Lutzomyia longipalpis* (GenBank accession no. 4887114).

DISCUSSION

In this article we report the results of a phylogenetic analysis of the genes for two presumptive odorant-binding proteins, OS-E and OS-F, in a variety of species of Drosophila. Although the genes are highly conserved, we have identified two regions that are particularly divergent in structure, one at the N terminus and one near the C terminus. Certain amino acid residues within the C-terminal heterogeneous region of the different Drosophila OS-E and OS-F proteins are specific to either OS-E or OS-F. In *D. virilis*, *OS-F* has undergone an alteration in its spatial regulation, such that it is expressed in a region of the antenna in which *OS-E* and *OS-F* are not found in *D. melanogaster* (McKenna *et al.* 1994). We propose that the duplication that gave rise to these two genes occurred at least 40 mya and that one or the other gene has subsequently been lost in certain lineages. Finally, we have produced a phylogenetic com-

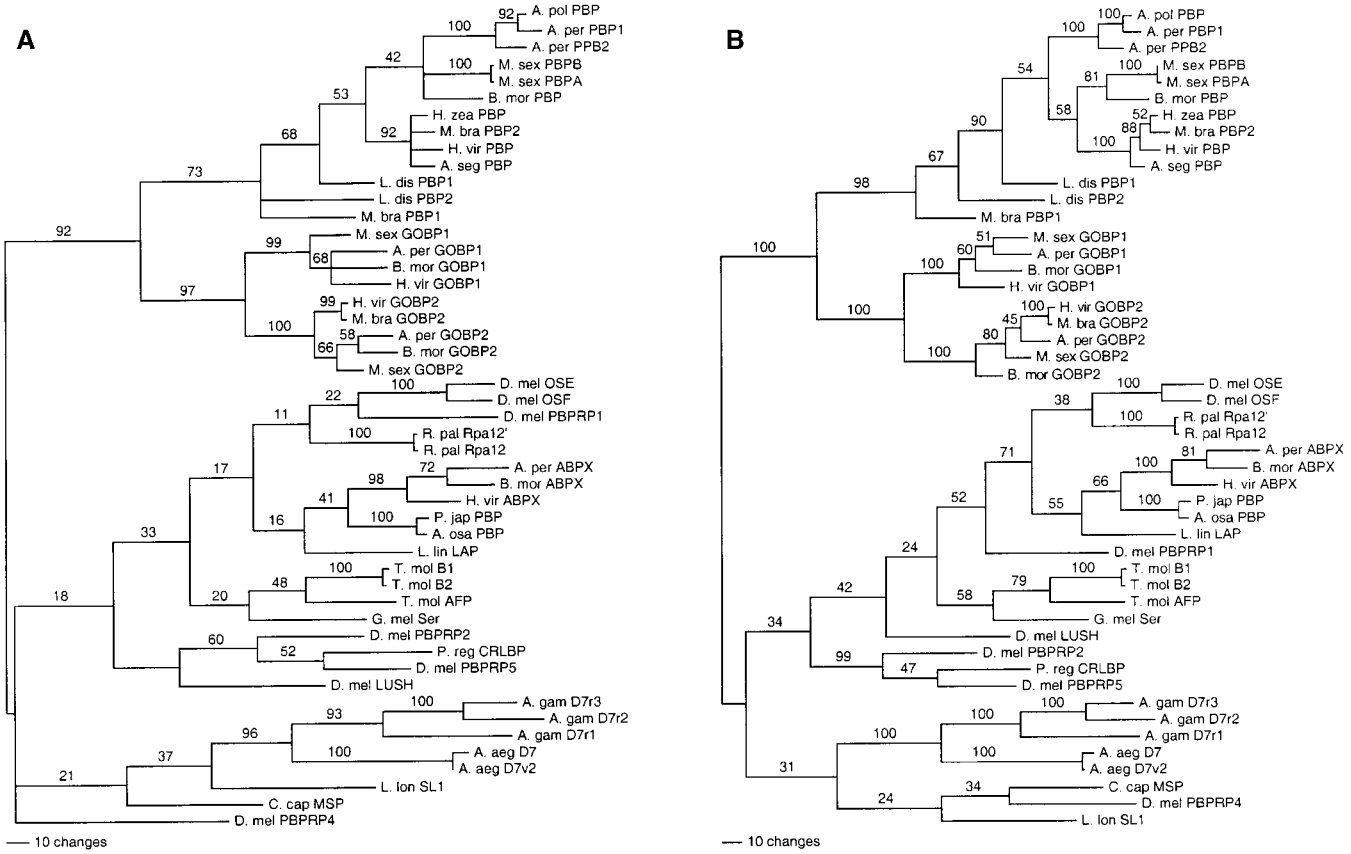


Figure 5.—Phylogenetic trees of insect OBPs and related proteins. Sequences were aligned using the ClustalW package. This alignment was used to construct a maximum parsimony (A) or distance (neighbor-joining) (B) tree using the PAUP package. Of the 340 residues in the global alignment for the parsimony tree (A), 171 positions were parsimony informative; a heuristic search produced 38 trees of length 2605, with a mean consistency index of 0.6. A strict consensus reconstruction with a consensus index of 0.807 is shown. The values above each node indicate bootstrap support for that node. The minimum evolution score for the distance tree (B) is 6.54451. The values above each node indicate bootstrap support for that node. Accession numbers of these sequences are the following: T. mol B1, M97916; T. mol B2, M97917; T. mol AFP3, U24237; G. mel SER, L41640; D. mel PBPRP-4, U05984; D. mel lush, AF001621; D. mel OS-E, U02543; D. mel OS-F, U02542; D. mel PBPRP-1, U05980; D. mel PBPRP-2, U05981; D. mel PBPRP-5, U05985; P. reg CRLBP, S78710; A. per PBP1 (formerly APR1), X96773; A. pol PBP, X17559; A. per PBP2 (formerly APR2), X96860; M. sex PBPA, M21797; M. sex PBPB, M21798; B. mor PBP, X94987; H. vir PBP (formerly HEL-1), X96861; L. dis PBP2, 2444187; A. per GOBP2 (formerly APR10), X96772; M. sex GOBP2, M73798; H. vir GOBP2 (formerly HEL-10), 96863; B. mor GOBP2, X94989; B. mor GOBP1, X94988; M. sex GOBP1, M73797; H. vir GOBP1 (formerly HEL-11), X96862; B. mor ABPX, X94990; H. zea PBP, 3639083; L. lin LAP, 3644030; L. dis PBP1, 2444185; H. vir ABPX, AJ002518; P. jap PBP, 3721994; A. osa PBP, 3721996; A. per ABPX, AJ002519; M. bra GOBP2, 2961244; R. pal Rpa12′, AF141865; R. pal Rpa12, AF139912; M. bra PBP2, 2961240; C. cap MSP, Y08954; A. seg PBP, 176679; M. bra PBP1, 2961242; A. gam D7r1, AJ133852; A. aeg D7, 159559; A. gam D7r3, AJ000035; A. gam D7r2, AJ000036; A. aeg D7v2, 4103771; L. lon SL1, AF132517.

parison of the various members of the diverse family of insect OBP-related proteins, a family that extends to include a number of proteins expressed outside the olfactory system. The potential implications of these findings are discussed below.

Patterns of sequence conservation between *D. virilis* and *D. melanogaster OS-F* genes suggest OBP functional domains: A comparison of OS-F protein sequences in *D. melanogaster* and *D. virilis* suggests that there are varying selective constraints across these sequences. Overall, *D. melanogaster* and *D. virilis* OS-F proteins display 76% amino acid identity. However, this identity masks a marked difference in the level of sequence conservation across the protein. The majority of nonconserved amino acids are found either at the N terminus, of which many, but not all, residues are predicted to lie within the signal sequence or in a 22-amino-acid stretch in the carboxy-terminal half of OS-F. This heterogeneous 22-amino-acid region displays only 55% amino acid identity. In contrast, the remaining portion of mature OS-F is 86% identical in these two species. Furthermore, only conservative amino acid substitutions (D-E, F-I, S-T, L-F, and I-V) are observed in OS-F residues following the first conserved cysteine (excluding the heterogeneous 22-amino-acid region). The 29 most C-terminal amino acids show a remarkable 100% identity.

The strong conservation seen in much of OS-F suggests that many of the residues, particularly those in the

extreme C terminus, are functionally important. At the same time, the highly divergent regions may also represent functionally significant domains of OBPs. OBPs are believed to bind odorants and deliver them to receptor molecules (Vogt *et al.* 1991; Pelosi 1994; Prestwich *et al.* 1995). If in fact different OBPs have different odor specificities (Du and Prestwich 1995; Prestwich *et al.* 1995), and if different OBPs deliver their bound odorants to different receptors, one might predict OBPs to contain two variable regions, one for odorant binding and one for receptor binding. It is therefore noteworthy that we have identified two nonconserved regions in OS-F. It will be interesting to determine whether the heterogeneous N-terminal amino acids are in fact on the exterior surface of OS-F, where they might be available to interact either directly or indirectly with receptors, and whether the heterogeneous region near the C terminus binds odors.

Consistent with the possibility that a heterogeneous region of OS-F is responsible for interacting with receptors, we note that the DOR family of candidate odorant receptors is extremely divergent in sequence (Clyne *et al.* 1999b; Vosshall *et al.* 1999). Only one amino acid is conserved among all of the first 17 genes identified, and there are no highly conserved stretches of even a few amino acids. Thus there are no obvious candidates for a highly conserved OBP-binding region; rather, it seems more likely that a structurally heterogeneous region of receptor proteins would bind to a structurally heterogeneous region of OBPs.

We note that sequence conservation between the *D. melanogaster* and *D. virilis OS-F* genes extends beyond the coding sequence. In particular, the high degree of nucleotide identity (76%) within the first coding region intron suggests the possibility that this intron carries conserved regulatory elements, for example, those needed for appropriate *OS-F* expression. One particular sequence, GCCACGC, occurs in at least one additional Drosophila OBP gene, *PBPRP-1*. It will be interesting to determine whether the sequence occurs in other OBP genes and whether it is, in fact, functionally required for an aspect of normal expression in the olfactory system.

We have observed several consistent structural differences between OS-E and OS-F proteins among the various Drosophila species we have examined. Our analysis revealed several amino acid residues that appear specific to either OS-E or OS-F (Figure 4C). Three residues (G^{25}, L^{28}, and I^{31}) are present in all OS-E, but in no OS-F sequences, whereas one residue (S^{27}) is present in all OS-F, but in no OS-E sequences. All of these conserved residues are clustered within the central portion of the region of greatest OS-E / OS-F heterogeneity. The region where the E- and F-specific residues occur is bracketed by residues conserved in all of the E and F proteins (P^{24} and M^{33}) and contains the invariant residue R^{29}. If this short stretch of amino acids is in fact part of a functional domain that differs between OS-E and OS-F, it could underlie critical differences in function between the two proteins.

Spatial expression patterns: The pattern of *OS-F* expression in *D. virilis* differs from that observed in *D. melanogaster* (McKenna *et al.* 1994). Although there is a good deal of overlap between the expression of *OS-F* in *D. virilis* and *OS-F* (as well as *OS-E)* in *D. melanogaster* (McKenna *et al.* 1994), the distribution of *D. virilis OS-F* transcripts includes a region of the antennal surface immediately ventral to the sacculus, a region where *OS-F* is not observed in *D. melanogaster*. The ventrolateral expression pattern of *OS-E* and *OS-F* transcripts in *D. melanogaster* resembles the distribution of the trichoid sensilla in both *D. melanogaster* (Venkatesh and Singh 1984) and *D. virilis* (D. Hekmat-Scafe and K. Störtkuhl, unpublished observations). *D. melanogaster* OS-E and OS-F are coexpressed within most, if not all, trichoid sensilla, as well as in some other interspersed sensilla (Hekmat-Scafe *et al.* 1997). *D. virilis* OS-F is likely expressed within some nontrichoid sensilla since few if any trichoid sensilla have been observed in the region immediately ventral to the sacculus. In *D. melanogaster*, the majority of the sensilla in this portion of the antenna are large basiconic sensilla (Venkatesh and Singh 1984), which respond electrophysiologically to a variety of general odorants (Siddiqi 1987). It will be of interest to determine whether *D. virilis* OS-F is, in fact, found within this class of basiconic sensilla, and, if so, whether the odorant response profiles of such sensilla differ in any way from those of the corresponding basiconic sensilla of *D. melanogaster*.

Duplication and divergence in the *OS-E*/*OS-F* genomic region: We have previously proposed that *OS-E* and *OS-F* arose by a tandem gene duplication, since in *D. melanogaster* the two genes are located in the same orientation ~930 bp apart and share a similar intron-exon organization (Hekmat-Scafe *et al.* 1997). As discussed below, our analysis of *OS-E* and *OS-F* genes in different Drosophila species suggests that this duplication is an ancient one.

Parsimony analysis of the various Drosophila OS-E and OS-F protein sequences (Figure 4A) yielded two major clusters, one containing OS-E-like proteins and the other containing OS-F-like proteins. The OS-E protein sequences are thus more similar to one another than any OS-E protein is to its paralogous OS-F counterpart. The divergence of *OS-E* and *OS-F* gene sequences likely reflects the different selective pressures shaping the OS-E and OS-F proteins.

The extra intron present in the coding region of both *D. melanogaster* and *D. virilis OS-F* (Figure 1A) may reflect either the loss of an intron by the *OS-E* progenitor or, possibly, the insertion of an intron into the *OS-F* progenitor. A 3-bp insertion (resulting in the addition of a single amino acid) observed in the *OS-E* gene of *D. melanogaster* and *D. teissieri*, but not in *D. willistoni*, probably occurred

at some point following the divergence of *D. willistoni* and *D. teissieri* (~35 mya) (Russo *et al.* 1995).

The duplication that gave rise to two *OS-E*/*OS-F* related genes probably occurred before *D. melanogaster* and *D. lebanonensis* diverged (*i.e.*, > ~40 mya; Russo *et al.* 1995). The most parsimonious interpretation of the data shown in Figure 4B is that the *OS-E* gene was lost both in the lineage giving rise to *D. simulans* and *D. mauritiana*, as well as in the one giving rise to *D. virilis* (Figure 4B). The products of the two *D. lebanonensis* genes more closely resemble OS-E than OS-F of *D. melanogaster* (Figure 4C). The high degree of nucleic acid sequence identity between the two *D. lebanonensis OS-E* genes (92%) far exceeds what is required for amino acid conservation, suggesting the presence of gene conversion between the two genes (Sullivan *et al.* 1990).

Phylogenetic analysis of the insect OBP family: Our phylogenetic analysis (Figure 5) illustrates that OS-E and OS-F are members of a large and diverse family of insect OBP-like proteins. OBP family representatives are found in a variety of Endopterygotan orders, including the Lepidoptera (moths), the Diptera (flies and mosquitoes), and the Coleoptera (beetles), as well as one Paraneoptera of the order Hemiptera ("true bug"; Vogt *et al.* 1999). Hence, insect OBP progenitors were likely present in ancient Neoptera (one subgroup of the winged, terrestrial insects, Pterygota).

The presence of two major clusters on both phylogenetic trees shown in Figure 5 suggests an important split between the moth PBPs and GOBPs (comprising the top cluster) and all other insect OBPs and related proteins (which define the bottom cluster). Similar results have recently been reported (Robertson *et al.* 1999; Vogt *et al.* 1999). Only one class of moth OBP-like proteins, the ABPXs, is placed in our bottom cluster, suggesting that it evolved independently from the moth PBPs and GOBPs.

The various Drosophila OBP family members do not define a coherent cluster on either phylogenetic network. Rather, they are found on a number of different branches, each of which includes OBPs from other species of insects. Most likely the divergence of insect OBPs into different subfamilies occurred long ago, before the Endopterygota-Paraneoptera split.

The cluster of proteins most closely related to OS-E and OS-F represent antennal proteins from a large variety of insect species. These include Drosophila PBPRP-1 (Pikielny *et al.* 1994), two presumptive beetle OBPs (Rpa12 and Rpa12′; GenBank accession nos. AF141865 and AF139912), the various moth ABPXs (Krieger *et al.* 1997), two closely related beetle PBPs (Wojtasek *et al.* 1998), and the Hemiptera antennal protein LAP (Vogt *et al.* 1999). It will be of interest to determine whether the various members of the OS-E and OS-F OBP subgroup display functional, as well as structural, homology.

A number of members of the insect OBP family represent small, secreted proteins found outside the olfactory system. All of these nonolfactory OBP-related proteins (except for SL1) carry only four of the six cysteines found in the majority of insect OBPs; they lack the second and fifth cysteine. Since all six cysteines are believed to be disulfide-bonded (Prestwich 1993), most likely the second and fifth cysteines normally form a disulfide bond in the olfactory insect OBPs. The loss of such a disulfide bond in the OBP-related proteins present outside of the olfactory system is likely to have a significant impact on their overall structures.

A knowledge of the tertiary structure of an insect OBP, such as OS-E or OS-F, should provide significant functional insight. It will also be of interest to determine the positions of particularly conserved and divergent stretches of amino acids on the OS-E and OS-F tertiary structures. Such information may suggest whether, for example, the C-terminal heterogeneous region is within a potential binding pocket that might bind odors or whether the N terminus is part of a solvent-exposed outer loop that might bind receptor molecules. In the long term, it will be interesting to correlate the structure of OS-E- and OS-F-related homologues from different species with their functional differences.

We gratefully acknowledge M. Astrid Moise, an undergraduate participant in the NECUSE program at Yale, and Dr. Kenneth Nelson and the members of his Fall '96 Laboratory in Nucleic Acids course at Yale University for their assistance in isolating *OS-E* and *OS-F* PCR products from various species of Drosophila. We also thank Charles Scafe (Stanford University) for assistance with the Psi-Blast search for insect OBP family members. This work was supported by a National Institutes of Health grant DC-02174 to J.C., a Yale Science Development Fund Award to R.D., and an AAUW Summer Faculty Fellowship and a Barrett Faculty Research Fellowship from Mills College to D.H.

LITERATURE CITED

Arca, B., F. Lombardo, M. de Lara Capurro, A. Della Torre, G. Dimopoulos *et al.*, 1999 Trapping cDNAs encoding secreted proteins from the salivary glands of the malaria vector *Anopheles gambiae*. Proc. Natl. Acad. Sci. USA **96:** 1516–1521.

Ashburner, M., 1989 *Drosophila: A Laboratory Handbook*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Clyne, P. J., S. J. Certel, M. de Bruyne, L. Zaslavsky, W. A. Johnson *et al.*, 1999a The odor specificities of a subset of olfactory receptor neurons are governed by Acj6, a POU-domain transcription factor. Neuron **22:** 339–347.

Clyne, P. J., C. G. Warr, M. R. Freeman, D. Lessing, J. Kim *et al.*, 1999b A novel family of divergent seven-transmembrane proteins: candidate odorant receptors in *Drosophila*. Neuron **22:** 327–338.

Du, G., and G. D. Prestwich, 1995 Protein structure encodes the ligand binding specificity in pheromone binding proteins. Biochemistry **34:** 8726–8732.

Feinberg, A., and B. Vogelstein, 1983 A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. **132:** 6–13.

Hekmat-Scafe, D. S., R. A. Steinbrecht and J. R. Carlson, 1997 Coexpression of two odorant-binding protein homologs in *Drosophila*: implications for olfactory coding. J. Neurosci. **17:** 1616–1624.

Higgins, D. G., and P. M. Sharp, 1988 CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. Gene **73:** 237–244.

Hung, M.-C., and P. C. Wensink, 1984 Different restriction enzyme-generated sticky DNA ends joined *in vitro*. Nucleic Acids Res. **12:** 1863–1874.

Kim, M. S., A. Repp and D. P. Smith, 1998 LUSH odorant-binding protein mediates chemosensory responses to alcohols in *Drosophila melanogaster*. Genetics **150:** 711–721.

Kodrik, D., V. A. Filippov, M. A. Filippova and F. Sehnal, 1995 Sericotropin: an insect neurohormonal factor affecting RNA transcription. Netherlands J. of Zool. **45:** 68–70.

Krieger, J., M. Mameli and H. Breer, 1997 Elements of the olfactory signaling pathways in insect antennae. Invert. Neurosci. **3:** 137–144.

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

Paesen, G. C., and G. M. Happ, 1995 The B proteins secreted by the tubular accessory sex gland of the male mealworm beetle, *Tenebrio molitor*, have sequence similarity to moth pheromone-binding proteins. Insect Biochem. Mol. Biol. **25:** 401–408.

Pelosi, P., 1994 Odorant-binding proteins. Crit. Rev. Biochem. Mol. Biol. **29:** 199–228.

Pelosi, P., and R. Maida, 1995 Odorant-binding proteins in insects. Comp. Biochem. Physiol. B **111:** 503–514.

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

Powell, J. R., and R. DeSalle, 1995 *Drosophila* molecular phylogenies and their uses. Evol. Biol. **28:** 87–138.

Prestwich, G. D., 1993 Bacterial expression and photoaffinity labelling of a pheromone binding protein. Protein Sci. **2:** 420–428.

Prestwich, G. D., G. Du and S. LaForest, 1995 How is pheromone specificity encoded in proteins? Chem. Senses **20:** 461–469.

Raha, D., and J. Carlson, 1994 OS9: a novel gene of *Drosophila* expressed in two olfactory organs. J. Neurobiol. **25:** 169–184.

Robertson, H. M., R. Martos, C. R. Sears, E. Z. Todres, K. K. O. Walden *et al.*, 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.

Russo, C. A., N. Takezaki and M. Nei, 1995 Molecular phylogeny and divergence of *Drosopholid* species. Mol. Biol. Evol. **12:** 391–404.

Sambrook, J., E. F. Fritsch and T. Maniatis, 1989 *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Siddiqi, O., 1987 Neurogenetics of olfaction in *Drosophila melanogaster*. Trends Genet. **3:** 137–142.

Steinbrecht, R. A., 1996 Are odorant-binding proteins involved in odorant discrimination? Chem. Senses **21:** 719–727.

Steinbrecht, R. A., M. Laue and G. Ziegelberger, 1995 Immunolocalization of pheromone-binding and general odorant-binding protein in olfactory sensilla of the silkmoths *Antheraea* and *Bombyx*. Cell Tissue Res. **282:** 203–217.

Sullivan, D. T., P. W. Atkinson and W. T. Starmer, 1990 Molecular evolution of the alcohol dehydrogenase genes in the genus *Drosophila*. Evol. Biol. **24:** 107–147.

Swofford, D. L., 1999 PAUP*. Phylogenetic analysis using parsimony (*and other methods), Version 4. Sinauer Associates, Sunderland, MA.

Tautz, D., and C. Pfeifle, 1989 A non-radioactive *in situ* hybridization method for the localization of specific RNAs in Drosophila embryos reveals translational control of the segmentation gene hunchback. Chromosoma **98:** 81–85.

Thymianou, S., M. Mavroidis, G. Kokolakis, K. Komitopoulou, A. Zacharopoulou *et al.*, 1998 Cloning and characterization of a cDNA encoding a male-specific serum protein of the Mediterranean fruit fly, *Ceratitis capitata* with sequence similarity to odourant-binding proteins. Insect Mol. Biol. **7:** 345–353.

Venkatesh, S., and R. Singh, 1984 Sensilla on the third antennal segment of *Drosophila melanogaster* meigen. Int. J. Insect Morphol. Embryol. **13:** 51–63.

Vogt, R. G., G. D. Prestwich and M. R. Lerner, 1991 Odorant-binding-protein subfamilies associate with distinct classes of olfactory receptor neurons in insects. J. Neurobiol. **22:** 74–84.

Vogt, R. G., F. E. Callahan, M. E. Rogers and J. C. Dickens, 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)*. Chem. Senses **24:** 481–495.

Vosshall, L. B., H. Amrein, P. S. Morozov, A. Rzhetsky and R. Axel, 1999 A spatial map of olfactory receptor expression in the *Drosophila* antenna. Cell **96:** 725–736.

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

Communicating editor: T. F. C. Mackay