Insect Biochemistry and Molecular Biology 39 (2009) 448–456

Contents lists available at ScienceDirect

Insect Biochemistry and Molecular Biology

journal homepage: www.elsevier.com/locate/ibmb

The insect SNMP gene family

Richard G. Vogt [a,*,1], Natalie E. Miller [a], Rachel Litvack [a], Richard A. Fandino [a], Jackson Sparks [a], Jon Staples [a], Robert Friedman [a], Joseph C. Dickens [b]

[a] Department of Biological Sciences, University of South Carolina, Columbia, SC 29208, USA
[b] USDA, ARS Henry A. Wallace Beltsville Agricultural Research Center Plant Sciences Institute, Invasive Insect Biocontrol and Behavior Laboratory, Bldg. 007, Rm. 030, Beltsville, MD 20705, USA

ARTICLE INFO

Article history:
Received 19 January 2009
Received in revised form
22 March 2009
Accepted 25 March 2009

keywords:
Pheromone
Receptors
Olfactory
Gustatory
Chemosensory
Gustatory
Mosquito
Fly

ABSTRACT

SNMPs are membrane proteins observed to associate with chemosensory neurons in insects; in *Drosophila melanogaster*, SNMP1 has been shown to be essential for the detection of the pheromone *cis*-vaccenyl acetate (CVA). SNMPs are one of three insect gene clades related to the human fatty acid transporter CD36. We previously characterized the CD36 gene family in 4 insect Orders that effectively cover the Holometabola, or some 80% of known insect species and the 300 million years of evolution since this lineage emerged: Lepidoptera (e.g. *Bombyx mori*, *Antheraea polyphemus*, *Manduca sexta*, *Heliothis virescens*, *Helicoverpa assulta*, *Helicoverpa armigera*, *Mamestra brassicae*); Diptera (*D. melanogaster*, *Drosophila pseudoobscura*, *Aedes aegypti*, *Anopheles gambiae*, *Culex pipiens quinquefasciatus*); Hymenoptera (*Apis mellifera*); and Coleoptera (*Tribolium castaneum*). This previous study suggested a complex topography within the SNMP clade including a strongly supported SNMP1 sub-clade plus additional SNMP genes. To further resolve the SNMP clade here, we used cDNA sequences of SNMP1 and SNMP2 from various Lepidoptera species, *D. melanogaster* and *Ae. aegypti*, as well as BAC derived genomic sequences from *Ae. aegypti* as models for proposing corrected sequences of orthologues in the *D. pseudoobscura* and *An. gambiae* genomes, and for identifying orthologues in the *B. mori* and *C. pipiens q.* genomes. We then used these sequences to analyze the SNMP clade of the insect CD36 gene family, supporting the existence of two well supported sub-clades, SNMP1 and SNMP2, throughout the dipteran and lepidopteran lineages, and plausibly throughout the Holometabola and across a broad evolutionary time scale. We present indirect evidence based on evolutionary selection (dN/dS) that the dipteran SNMPs are expressed as functional proteins. We observed expansions of the SNMP1 sub-clade in *C. pipiens q.* and *T. castaneum* suggesting that the SNMP1s may have an expanded functional role in these species.

1. Introduction

SNMPs are insect membrane proteins which associate with pheromone sensitive neurons in Lepidoptera and Diptera (Rogers et al., 1997, 2001a,b; Vogt, 2003; Benton et al., 2007; Forstner et al., 2008; Jin et al., 2008). SNMPs comprise a sub-clade of insect genes related to the human protein fatty-acid transport protein CD36 (Fig. 1A; Nichols and Vogt, 2008). The first SNMPs were identified from Lepidoptera (Rogers et al., 1997, 2001a,b; Krieger et al., 2002; Forstner et al., 2008). SNMPs from several lepidopteran species, here referred to as SNMP1, were shown to be antenna specific, associating with pheromone specific olfactory neurons in a manner suggesting they play a central role in pheromone detection (Rogers et al., 1997, 2001a,b; Krieger et al., 2002). A second lepidopteran SNMP, SNMP2, also associates with pheromone sensitive sensilla but has been shown to express in sensilla support cells rather than neurons (Rogers et al., 2001b; Forstner et al., 2008). Recently, a *Drosophila melanogaster* SNMP1 (SNMP1*Dmel*, cg7000) was found to be essential for the detection of the pheromone *cis*-vaccenyl acetate (CVA) (Benton et al., 2001; Jin et al., 2008); this protein was not only expressed in antennae, but also in other body parts such as legs and wings (Benton et al., 2007; Miller et al., 2007).

We recently surveyed the insect CD36/SNMP gene family from the genomes of 6 insect species: the fruitflies (Diptera) *D. melanogaster* and *Drosophila pseudoobscura*; the mosquitoes (Diptera) *Anopheles gambiae* and *Aedes aegypti*; the honeybee (Hymenoptera) *Apis mellifera*, and the beetle (Coleoptera) *Tribolium castaneum* (Nichols and Vogt, 2008). This study suggested that the SNMPs

* Corresponding author. Tel. :+1 803 777 8101; fax: +1 803 777 4002.
E-mail address: vogt@biol.sc.edu (R.G. Vogt).
[1] http://www.biol.sc.edu/faculty/vogt.html

0965-1748/$ – see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ibmb.2009.03.007

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
2009		**00-00-2009 to 00-00-2009**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
The insect SNMP gene family	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **University of South Carolina,Department of Biological Sciences,Columbia,SC,29208**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

SNMPs are membrane proteins observed to associate with chemosensory neurons in insects; in Drosophila melanogaster, SNMP1 has been shown to be essential for the detection of the pheromone cisvaccenyl acetate (CVA). SNMPs are one of three insect gene clades related to the human fatty acid transporter CD36. We previously characterized the CD36 gene family in 4 insect Orders that effectively cover the Holometabola, or some 80% of known insect species and the 300 million years of evolution since this lineage emerged: Lepidoptera (e.g. Bombyx mori, Antheraea polyphemus, Manduca sexta, Heliothis virescens, Helicoverpa assulta, Helicoverpa armigera, Mamestra brassicae); Diptera (D. melanogaster Drosophila pseudoobscura, Aedes aegypti, Anopheles gambiae, Culex pipiens quinquefasciatus); Hymenoptera (Apis mellifera); and Coleoptera (Tribolium castaneum). This previous study suggested a complex topography within the SNMP clade including a strongly supported SNMP1 sub-clade plus additional SNMP genes. To further resolve the SNMP clade here, we used cDNA sequences of SNMP1 and SNMP2 from various Lepidoptera species, D. melanogaster and Ae. aegypti, as well as BAC derived genomic sequences from Ae. aegypti as models for proposing corrected sequences of orthologues in the D. pseudoobscura and An. gambiae genomes, and for identifying orthologues in the B. mori and C. pipiens q. genomes. We then used these sequences to analyze the SNMP clade of the insect CD36 gene family supporting the existence of two well supported sub-clades, SNMP1 and SNMP2, throughout the dipteran and lepidopteran lineages, and plausibly throughout the Holometabola and across a broad evolutionary time scale. We present indirect evidence based on evolutionary selection (dN/dS) that the dipteran SNMPs are expressed as functional proteins. We observed expansions of the SNMP1 sub-clade in C. pipiens q. and T. castaneum suggesting that the SNMP1s may have an expanded functional role in these species.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**9**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

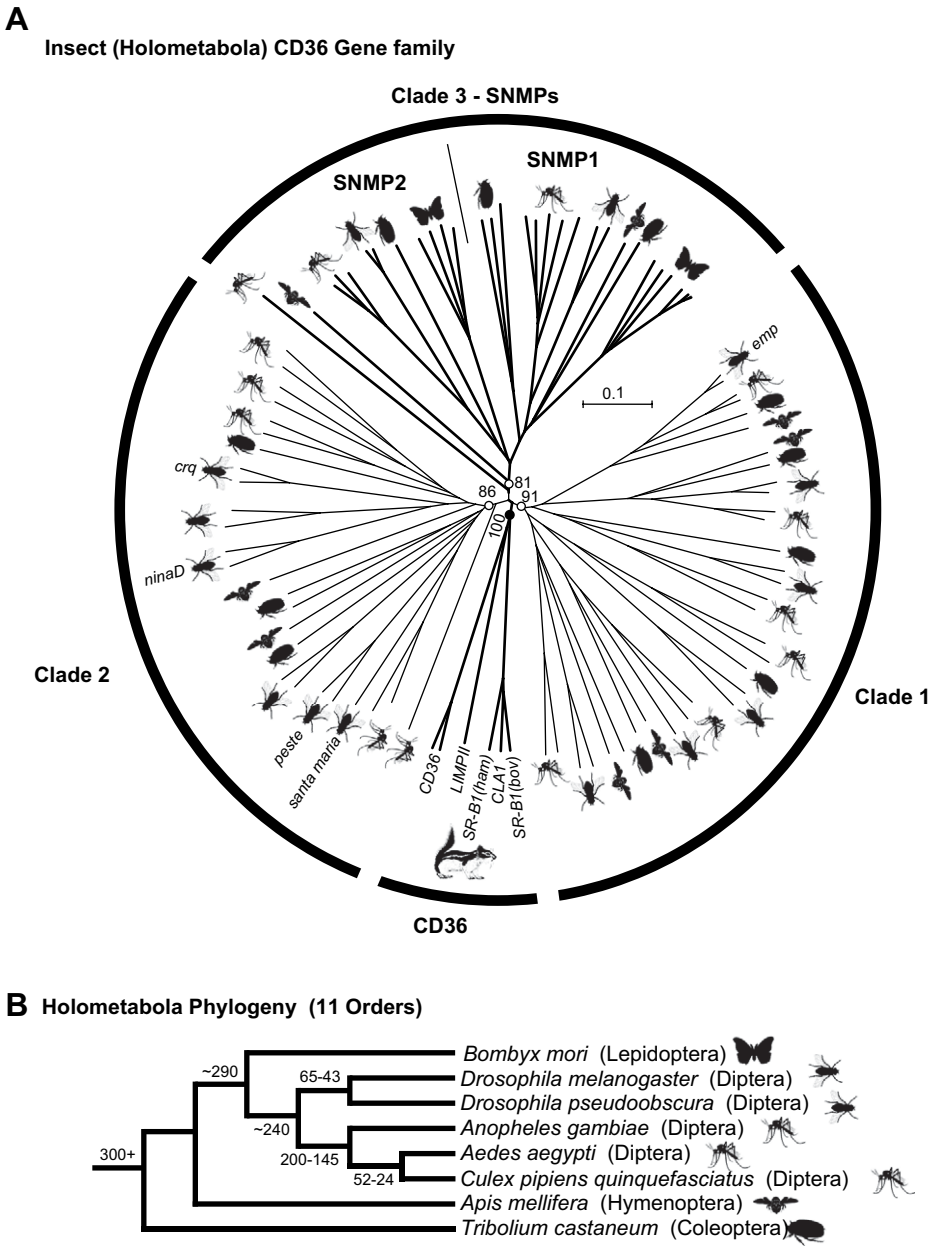

Fig. 1. A. Neighbor joining tree of insect CD36 homologues (MEGA4, pairwise gap deletion due to sequence divergence). All non-SNMP sequences used were identical to those reported in Nichols and Vogt (2008), where they are more fully described; SNMP sequences (Clade 3) are noted in Supplementary Materials (Table 2), and their relationships are shown in more detail in Fig. 4. *emp*, *crq*, *peste*, *ninaD*, *santa maria* refer to characterized *D. melanogaster* genes (see Nichols and Vogt, 2008). Bootstrap support (1000 replicates) is indicated for the major clades. B. Phylogeny of holometabolous lineages and species used in this study. Numbers indicate millions of years (Mya) since indicated lineages diverged (see text).

comprised one of three major clades of the insect CD36 gene family; the SNMP clade showed a complex topography with a well supported SNMP1 sub-clade plus additional SNMP genes. To resolve this topography, our current study focuses on the SNMP clade from the dipteran and lepidopteran genomes, adding the SNMP genes of the mosquito *Culex pipiens quinquefasciatus* and the silk moth *Bombyx mori*. These genes are compared with homologues from Hymenoptera and Coleoptera affording a survey of the majority of the holometabolous insect lineage (Fig. 1B), which includes at least 80% of all known insect species. The Holometabola are thought to have emerged around 300 + million years ago (Mya), and the Lepidoptera/Trichoptera and Diptera to have diverged around 290 Mya (Grimaldi and Engel, 2005). The *Drosophila* and mosquito lineages are thought to have diverged 240 Mya (Grimaldi and Engel, 2005). The *D. melanogaster* and *D. pseudoobscura* lineages diverged 65–43 Mya (O'Grady, 1999; Tamura et al., 2004), and are among the most widely diverged genomes available for this genus. Mosquitoes (Culicidae) are comprised of three suborders, two of which include the blood feeding genera: Culicinae (including *Aedes* sp. and *Culex* sp.) and Anophelinae (including *Anopheles* sp.). The Culicinae and Anophelinae lineages diverged 200–145 Mya (Krzywinski et al., 2006) while the *Aedes* and *Culex* lineages diverged 52–24 Mya (Foley et al., 1998).

Our previous characterization of the insect CD36 gene family relied primarily on the annotated sequences provided in the genome databases (Nichols and Vogt, 2008); many of these annotations were

truncated or otherwise missing elements. For the current study, we used cDNA sequences of SNMP1 and SNMP2 from *D. melanogaster* and *A. aegypti*, as well as BAC (Bacterial Artificial Chromosome) derived genomic sequences from *A. aegypti*, as models for proposing corrected sequences of orthologues in *D. pseudoobscura* and *A. gambiae* genomes and for identifying orthologues in the *Culex pipiens q.* genome. We similarly used published Lepidoptera SNMP1 and SNMP2 cDNA sequences to identify corresponding genes from the *B. mori* genome. We then used these revised sequences to reanalyze the SNMP clade, demonstrating the existence of two well supported sub-clades, SNMP1 and SNMP2, throughout the dipteran and lepidopteran lineages, and plausibly throughout the Holometabola.

2. Methods

2.1. Animals and tissue collection

D. melanogaster (W^{1118}) used in this study were reared on a standard diet (20 °C, 16h:8h L:D); 3–5 day old adults were collected, frozen, lyophilized, and processed as described below. *A. aegypti* eggs were graciously provided by Mark Brown (University of Georgia, Department of Entomology) and raised on a larval diet (pond fish food) at 27 °C (16h:8h L:D). Newly ecdysed pupae were transferred with water to small cups in cages (BioQuip Products, #1450B) for adult emergence; adults were provided with sugar water (20% sucrose) via cotton wicks. For adult collection, cages containing 7-day-old adults (and otherwise empty) were lowered into a −70 °C chest freezer and flash frozen; bodies were collected from the cage floors still frozen and subsequently dried by lyophilization. Lyophilized body parts (*D. melanogaster* or *A. aegypti*) were dissected under a stereo microscope at room temperature and collected in vials (1.5 ml) containing 95% ethanol; the ethanol was subsequently removed by pipette, with the final residue removed under vacuum (SpeedVac). Dried tissues were processed as described below for mRNA isolation.

2.2. cDNA and genomic sequencing of D. melanogaster and A. aegypti SNMPs

PCR primers and sequence information including tissue sources are listed in Supplementary Materials (Tables 1 and 2). cDNA and genomic sequences obtained through this project and submitted to GenBank include partial sequences of SNMP1*Dmel* derived from various tissues (EF596938 head, EF560171 wing, EF560170 leg), cDNA sequences of SNMP2*Dmel* (EU189152), SNMP1*Aaeg* (EU246941) and SNMP2*Aaeg* (EU189151), BAC derived genomic sequences of *snmp1Aaeg* (FJ387158) and *snmp2Aaeg* (FJ387159), and sequences of *A. aegypti* BAC clones BL7I-114 (FJ387160, including *snmp1Aaeg*) and BL14K-12 (FJ387161, including *snmp2Aaeg*).

For SNMP expression studies (Fig. 2), mRNA was isolated from tissues collected and pooled from 50 individuals (mixed gender *D. melanogaster* or male or female *A. aegypti*); equivalent aliquots of mRNA were processed for cDNA synthesis. Control target sequences (Fig. 2) were ribosomal protein L32 (RpL32, *D. melanogaster*, href = "genbank:NM_079843">NM_079843) or ribosomal protein S17 (RpS17, *A. aegypti*, AY927787). Non-quantitative PCR reactions were carried out through 30 cycles (see below).

Messenger RNA was isolated using TRIzol® LS Reagent (invitrogen) or by acid-phenol extraction (Chomczynski and Sacchi, 1987). cDNA was synthesized (SuperScript® III RT, invitrogen) and amplified by PCR (Platinum® Taq DNA Polymerase, regular for tissue specificity studies or "high fidelity" for cloning; invitrogen). PCR products were gel purified as needed (Geneclean Turbo®, Q-BIOgene), and inserted into plasmid vectors for subsequent analysis TOPO® TA Cloning Dual Promoter Kit, invitrogen. Cloned PCR products were sequenced by cycle sequencing by the University of South Carolina EnGenCore (Joe Jones, Director). Near full length sequences of SNMP2*Dmel*, SNMP1*Aaeg* and SNMP2*Aaeg* were obtained from cDNAs derived from various tissues, and submitted to GenBank. Partial sequences of SNMP1*Dmel* from head, wing and leg were also obtained and submitted to GenBank.

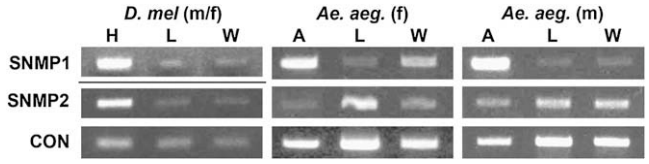

Fig. 2. SNMP1, SNMP2 and control PCR products were amplified (non-quantitatively) from cDNAs of *D. melanogaster* and *A. aegypti* tissues using primers listed in Supplementary Materials (Table 1). For *D. melanogaster*, male and female (m/f) tissues were combined; for each tissue, SNMP2*Dmel* and control (RpL32) products were generated under identical reaction conditions, except for primers; SNMP1*Dmel* products were from a separate experiment and control products are not shown. For *A. aegypti*, male (m) and female (f) tissues were analyzed separately; for each tissue, SNMP1*Aaeg*, SNMP2*Aaeg* and control (RpS17) products were generated under identical reaction conditions, except for primers. H, head with antennae; A, antennae; L, legs; W, wings.

The coding and upstream regions of the *snmp1Aaeg* and *snmp2Aaeg* genes were obtained by sequencing two clones isolated from an *A. aegypti* BAC genomic library (Shizuya et al., 1992; Jiménez et al., 2004; Lobo et al., 2007); the library was screened and clones were generously provided by Dave Severson and Becky deBruyn (Notre Dame University). BAC clones were sequenced using 454 technology by the University of South Carolina EnGenCore. See Supplementary Materials (Methods) for details.

2.3. Annotation of SNMP genes from available genomes

Many of the SNMP sequences characterized in this study were previously identified (see Nichols and Vogt, 2008); new sequences include those from *Culex pipiens q.* and the *snmp2* gene of *B. mori*. For the previous study, sequences were accepted as presented in the published annotations. For the current study, cDNA sequences from *D. melanogaster*, *A. aegypti* and *Manduca sexta* were used as models to modify these annotations for detailed analysis; annotations, modifications and sequences are described in our Supplementary Materials (Methods, Tables 2 and 3). All cDNA sequences used in this study (e.g. Fig. 4) are listed in Supplementary Materials (Sequence Data).

2.4. Alignments, trees, percent identity, dN/dS

ClustalX v1.81 (Larkin et al., 2007) was used for amino acid alignments, using default alignment parameters. Mega 4 (Tamura et al., 2007) was used to construct Neighbor-joining trees (1000 bootstrap replicates, nodes collapsed to 50% bootstrap support) and to derive percent identities and synonymous and non-synonymous (dS, dN) values (Nei-Gojobori model with Jukes-Cantor correction). For dS and dN calculations, nucleotide sequences were aligned to their corresponding codon (amino acid) alignments, non-overlapping ends were trimmed, any remaining start and stop codons were removed and codons split by introns were removed.

3. Results

Complementary DNAs encoding near-full length SNMP2 of *D. melanogaster* (SNMP2*Dmel*) and SNMP1 and SNMP2 of *A. aegypti* (SNMP1*Aaeg* and SNMP2*Aaeg)* were obtained by PCR, using primers

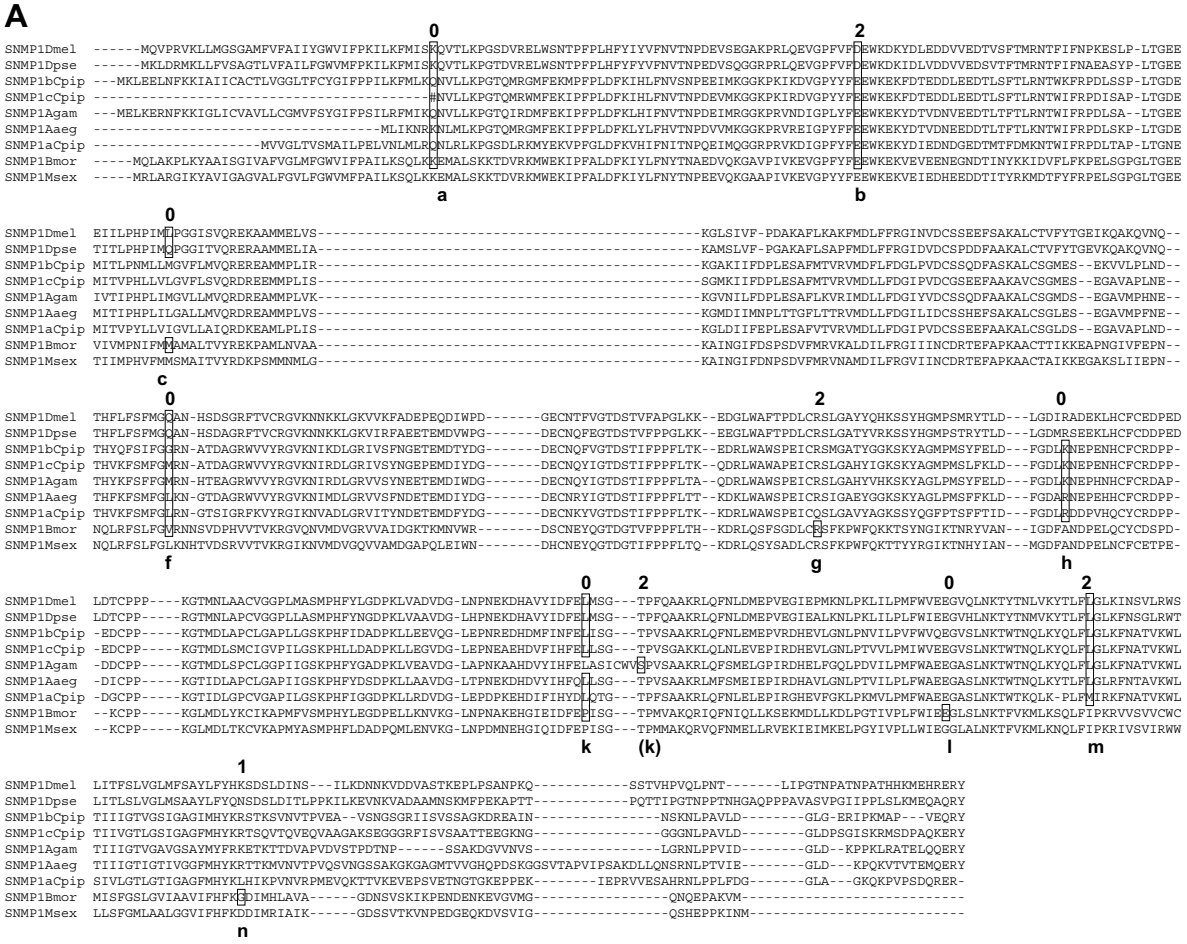

Fig. 3. Alignment (ClustalX) of Diptera and Lepidoptera SNMP amino acid sequences; the sequences included are those for which a valid cDNA model is available. While there is no available genomic sequence for *M. sexta*, this species is included for comparison with *B. mori*. 3A shows the SNMP1 proteins and 3B the SNMP2 proteins; these were aligned as one group and separated for this presentation (six gaps in the SNMP1 alignment are against the SNMP2 alignment; one gap in the SNMP2 alignment is against the SNMP1 alignment). Insertion sites are noted by boxes (3′ amino acid of each exon); phases of boundary codons are noted (0, 1, 2). Homologous intron insertion sites are also noted (a–n) for cross reference with Table 1.

designed to annotated sequences previously identified by Nichols and Vogt (2008). Genomic sequences for SNMP1*Aaeg* and SNMP2*Aaeg* were obtained from BAC clones generously supplied by Dave Severson and Becky deBruyn (Notre Dame University). Expression of *D. melanogaster* and *A. aegypti* SNMP1 and SNMP2 was confirmed by PCR (non-quantitative) of cDNAs synthesized from mRNAs isolated from a variety of tissues including heads/antennae, legs and wings (Fig. 2). The identities of these PCR products were confirmed by sequencing. The near-full length cDNA and genomic (BAC) SNMP sequences were submitted to GenBank, as were partial SNMP1*Dmel* cDNA sequences derived from head, leg and wing tissues (see Supplementary Materials: Table 2).

The *D. melanogaster* and *A. aegypti* SNMP cDNA sequences, and an available SNMP1*Dmel* cDNA were used as models to identify and annotate orthologues from the genomes of *D. pseudoobscura*, *A. gambiae* and *C. pipiens q.* Genes encoding SNMP1 and SNMP2 of *B. mori* (*snmp1Bmor* and *snmp2Bmor*) were identified and annotated from the genome of *B. mori* using a published cDNA of SNMP1*Bmor* and published cDNAs of SNMP2s identified from other Lepidoptera. Please see Supplementary Materials (Methods) for details of these activities, and Supplementary Materials (Tables 2 and 3) for all accession numbers and annotations, and for predicted cDNA and amino acid sequences of all taxa used in this study.

Fig. 3 shows an alignment of SNMP proteins from the Diptera and Lepidoptera species. This figure notes intron insertion sites and their phase (a codon not split by the intron has phase 0, a split codon has phase 1 or 2 depending on whether the split is between nucleotides 1–2 or nucleotides 2–3). Intron insertion sites are noted by letters. All homologous intron insertion sites have the same phase, both within a specific SNMP and between SNMP1 and SNMP2, supporting the evolutionary relatedness between these genes. Many of these sites, including their phase, are conserved between Diptera and Lepidoptera.

Fig. 4 shows a Neighbor Joining tree of all SNMP sequences listed in Supplementary Materials (Tables 2 and 3). The topology of this tree is rooted against the larger CD36 gene family (Fig. 1A) at a position noted by an asterisk. This analysis suggests two SNMP sub-clades: SNMP1 and SNMP2. Among the Lepidoptera and Diptera, only the dipteran mosquito *C. pipiens q.* shows an expansion of the SNMP1 genes. Three SNMP1*Cpip* genes were identified in the genome; these were arrayed in an uninterrupted series in a region spanning 31,381 bp; the three genes are separated by 4704 and 7770 bp respectively, have identical intron insertion site topology (Table 1), and share 67–81% amino acid identity (Table 2), all strongly suggesting they resulted from gene duplication events distinct from the *A. aegypti* and *A. gambiae* lineages.

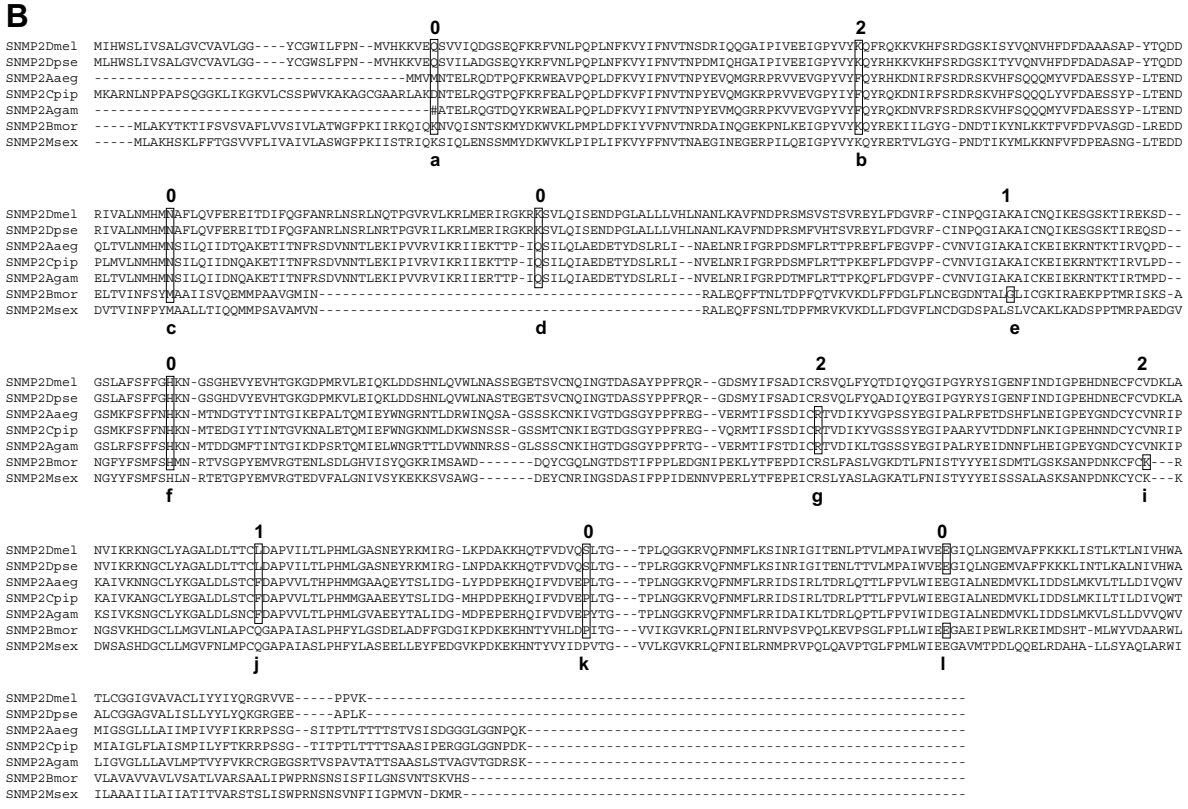

Fig. 3. (*continued*).

The Neighbor Joining tree (Fig. 4) also includes sequences from Hymenoptera (*A. mellifera*) and Coleoptera (*T. castaneum*). SNMP1 and SNMP2 genes are represented in both taxa. *T. castaneum* shows an apparent Coleoptera specific expansion of the SNMP1 genes. Four SNMP1*Tcas* genes are arrayed in an uninterrupted series in a region spanning 12,727 bp; annotated gene sizes range from 1885–31,381 bp, intervals between genes range from 260 to 1820 bp, each annotation indicates 7 or 8 exons. The proximity of these 4 SNMP1*Tcas* genes strongly suggests they derived from a gene duplication event. We have not further analyzed the *A. mellifera* and *T. castaneum* SNMP genes due to an absence of appropriate cDNA models.

Table 1 compares the presence of homologous intron insertion sites and the sizes of homologous introns and exons of those sequences shown in Fig. 3. Some introns, especially those in *A. aegypti*, are quite large, around 10–15 kb. The large size of introns is somewhat species specific, correlating with the concentration of repetitive elements within those species' genomes (e.g. The International Silkworm Genome Consortium, 2008; Nene et al., 2007). In most cases, large introns did not obscure the identification of coding exons. However, initial exons were often difficult to predict if their model genes contained a short first exon followed by a large first intron.

Table 2 compares the amino acid sequence identities of the dipteran and lepidopteran SNMPs included in Fig. 3. Sequence identities are quite high when comparing the same gene within an insect Order, but predictably decrease with phylogenetic distance. SNMP1 vs. SNMP2 sequence identities are quite low, consistent with an early divergence for these two gene sub-clades.

Table 3 compares the synonymous (dS) and non-synonymous (dN) changes that have occurred in the coding nucleotide sequences of the dipteran SNMP1 and SNMP2 genes. In general, dN/dS values are low: 0.08–0.30 for SNMP1, 0.03–0.30 for SNMP2. These values suggest that negative or purifying selection and not positive selection acting on these genes (Nei and Gojobori, 1986), and therefore that each of these SNMP genes is expressed as a functional protein (Torrents et al., 2003).

4. Discussion

This study focuses on the SNMP genes of holometabolous insects, a lineage which emerged within the Neoptera lineage around 300 Mya and comprises >80% of named insect species and the most successful (by number of species) insect Orders (Fig. 1B, see Nichols and Vogt, 2008). SNMPs are related to a larger gene family characterized by the human fatty acid transporter CD36, a membrane protein with a broad range of described roles that include cholesterol transport by macrophage cells, cell–cell recognition or cytoadhesion between a variety of cells, and fatty acid recognition in taste receptor cells (e.g. Rasmussen et al., 1998; Gilbertson et al., 2005; Calder and Deckelbaum, 2006; Febbraio and Silverstein, 2007; Fukuwatari et al., 1997; Rac et al., 2007; Gaillard et al., 2008a,b). In our previous study (Nichols and Vogt, 2008), we characterized and reviewed the CD36 gene family in insects using available genome sequences, all of which were within the holometabolous lineage. We suggested that the insect CD36 gene family is comprised of 3 major clades, one of which includes the SNMP genes. Several of the *D. melanogaster* CD36 family members outside the SNMP clade have been characterized and shown to possess similar functions as CD36: NinaD and Santa Maria function as fatty-acid transporters (carotene) (e.g. Giovannucci and Stephenson, 1999; Kiefer et al., 2002; Yang and O'Tousa, 2007; Wang et al., 2007); Croquemort and Peste function through cell–cell interactions to mediate attacks on apoptotic cells or bacteria (e.g. Franc et al., 1996; Stuart et al., 2005; Philips et al., 2005). Within the

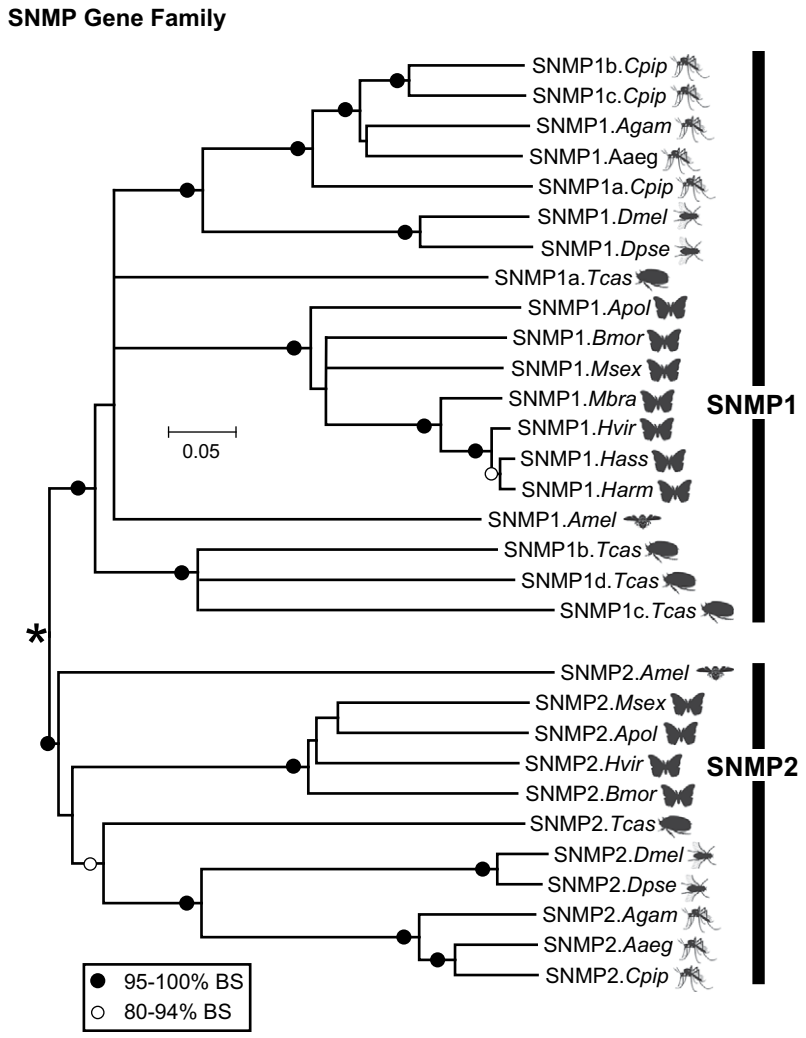

Fig. 4. Neighbor joining tree of SNMP sequences noted in Supplementary Materials (Table 2) (MEGA4, complete gap deletion); bootstrap support is indicated by symbol with branches collapsed at 50% (unmarked nodes have 50–79% bootstrap support). This tree is unrooted; however, the asterisk notes the position the Clade 3 node in Fig. 1A.

SNMP clade, SNMP1 has been shown to be required for the chemosensory detection of the fatty acid pheromone CVA (Benton et al., 2007; Jin et al., 2008), perhaps similar to a reported association of CD36 with mammalian taste cells and its possible function in fat detection (Fukuwatari et al., 1997; Gilbertson et al., 2005; Gaillard et al., 2008a,b). Our study demonstrated structural similarities of the insect CD36-related genes that argued for a common origin, and further suggested that homologues of these genes are represented in species throughout the holometabolous lineage (Nichols and Vogt, 2008).

Our previous study (Nichols and Vogt, 2008) suggested there may be multiple subgroups within the SNMP clade. For the current study, we cloned and sequenced SNMP genes from *D. melanogaster* and *A. aegypti* and used these to remodel available SNMPs and to identify SNMPs from additional species; we also used available Lepidoptera SNMP sequences to identify SNMP genes from *B. mori*. Analysis of these remodeled and additional sequences suggests the insect SNMPs are organized into two sub-clades, SNMP1 and SNMP2, presumably deriving from a common ancestor.

Both SNMP1 and SNMP2 are expressed in a variety of tissues. In their study of SNMP1*Dmel* function in antenna olfactory neurons, Benton et al. (2007) noted expression in both non-antennal and antennal tissue; within the antennae it was reported in neurons and non-neurons (presumably sensilla support cells). For the current study, we cloned and sequenced SNMP1*Dmel* mRNAs from head (including antennae), leg and wing tissues (see Supplemental Materials Table 2). PCR analysis showed SNMP1*Dmel*, SNMP2*Dmel*, SNMP1*Aaeg* and SNMP2*Aaeg* expression in heads/antennae, legs and wings (Fig. 2). We identified SNMP2*Bmor* cDNAs in larval EST libraries derived from maxillary galea (chemosensory antenna), silk glands and midgut (see Supplemental Materials Table 2). Two recent studies have shown that SNMP1*Dmel* is required for the detection of the pheromone CVA and proposed specific molecular models underlying this requirement through direct interaction with the CVA receptor protein, either mediating the transfer of CVA from odorant binding protein to receptor (Benton et al., 2007), or acting as an inhibitory subunit of the receptor (Jin et al., 2008). However, the broad expression pattern of the SNMPs suggests that the function of these genes may be more general than those proposed, or that the SNMPs have diverse functions specific to the different tissues. Mammalian CD36, the defining member of the overall gene family, also expresses in a wide range of tissues and displays a range of phenotypes that might broadly be described as fatty acid transport and cell-cell recognition, similar to the *D. melanogaster* CD36 homologues NinaD, Santa Maria, Croquemort and Peste (see Nichols and Vogt, 2008). The question remains whether SNMPs show similar functions.

Table 1
Comparison of Intron/Exon Sizes for homologous introns.

In	Dm S1	Dp S1	Aa S1	Ag S1	Cp S1a	Cp S1b	Cp S1c	Bm S1	Dm S2	Dp S2	Aa S2	Ag S2	Cp S2	Bm S2	In
	111	110	20	119	65	119	1	113	110	10	12	1	128	114	
a	61	56	423	337	4707	7770	1	8173	3272	3072	10063	1	240	?	a
	157	157	157	157	157	157	157	157	157	157	157	157	157	157	
b	67	64	67	6384	61	59	57	8745	57	71	74	90	60	5511	b
	147	147						150	147	147	147	147	147	144	
c	446	461						1092	91	118	394	658	198	6102	c
									137	137	134	134	134		
d									1238	1321	10029	1967	3150	168	d
	254	254	399	396	399	399	399	257	260	260	254	254	254		
e														5805	e
														91	
f	69	60	661	69	53	59	51	2118	69	66	65	76	64	799	f
								211			229	229	229	218	
g			296	296	296	296	296	811			67	143	65	585	g
h			57	68	59	52	52							126	h
l	509	506	203	223	203	203	203	285	432	432	200	199	199	1985	l
														187	
j									2109	3366	15490	6146	4830	?	j
									118	118	118	118	118	118	
k	61	62	64	62	66	56	54	534	66	68	75	110	60	809	k
								125	125	125				125	
l	178	178	178	167	175	178	178	1068	63	69				1933	l
m	61	70	16257	400	56	57	61	141	173	173	368	374	368	225	m
n	294	324	327	264	306	270	288	598							n
								118							

Intron (In) positions (a–n) are noted in Fig. 3; intron sizes are noted in yellow cells and exon sizes in white cells. An entry of "1" indicates a non-annotated 5' exon, but an identifiable homologous intron boundary at the 5' end of the next exon. "?" under Bm-S2 indicates a region of missing sequence data (see Supplementary Materials: Methods and Sequence Data). Dm, *D. melanogaster*; Dp, *D. pseudoobscura*; Aa, Ae. aegypti; Ag, An. gambiaea; Cp, *C. pipiens* qu.; Bm, *B. mori*; S1, SNMP1; S2, SNMP2. (For interpretation of the references to colour in this table legend, the reader is referred to the web version of this article.).

A challenge of mining a genome database is confirming the genes are expressed as functional proteins. Here, we used analysis of non-synonymous and synonymous nucleotide changes (dN/dS) to suggest that the dipteran SNMPs analyzed in this study are expressed and functional. dN/dS analysis is a test for evolutionary selection acting on homologous DNA sequences. Nucleotides of individual homologous codons are compared, noting nucleotide changes that change the amino acid (non-synonymous change) and nucleotide changes that do not change the amino acid (synonymous change). Assuming that all nucleotides have an equal probability of changing over evolutionary time, observing a disproportionate number of changes in synonymous or non-synonymous changes suggests positive ($dN/dS > 1$) or negative selection ($dN/dS < 1$) (e.g. Hughes and Nei, 1988). For example, analysis for positive selection has been used to attempt identification of ligand binding sites in chemosensory receptors (e.g. Tunstall et al., 2007). More importantly, selection can arguably only act on genes that are in fact expressed and functional, and therefore dN/dS analysis should be useful to indirectly suggest that a gene is expressed and functional. Torrents et al. (2003) compared 1659 functional and 1703 pseudo- (presumed non-functional) genes from the human genome database: dN/dS values were broadly distributed for both categories, but about 90% of expressed genes had values at or below 0.2, while about 80% of pseudogenes had values at or above 0.2. Our analysis (Table 3) yielded values ranging from 0.03 to 0.30, with most below 0.20; these values suggest that most if not all of the SNMPs are expressed as functional proteins.

Using dN/dS analysis to indirectly suggest functional expression should be useful for studies involving large gene families such as chemosensory genes (including odor receptors, odorant binding proteins, gustatory receptors), and especially where genome sequences are available from closely related species permitting comparison of orothologous sequences. Direct methods such as PCR (e.g. Robertson and Wanner, 2006) or microarray surveys (e.g. Zhang et al., 2004) require considerable effort and only confirm

Table 2
Percent identities.

%ID	SNMP1									SNMP2					
	Dm	Dp	CP 1b	CP 1c	Ag	Aa	CP 1a	Ms	Bm	Dm	Dp	As	Cp	Ag	Ms
SNMP1*Dpse*	83														
SNMP1b*Cpip*	51	53													
SNMP1c*Cpip*	50	49	81												
SNMP1*Agam*	51	52	74	75											
SNMP1*Aaeg*	51	51	73	75	75										
SNMP1a*Cpip*	49	48	67	70	67	68									
SNMP1*Msex*	42	40	43	42	40	42	41								
SNMP1*Bmor*	40	41	43	41	39	41	43	74							
SNMP2*Dmel*	27	26	27	27	27	27	27	29	29						
SNMP2*Dpse*	27	26	27	27	27	28	27	30	29	93					
SNMP2*Aaeg*	26	27	29	29	29	27	31	31	29	48	49				
SNMP2*Cpip*	26	28	29	29	29	27	30	30	30	49	48	88			
SNMP2*Agam*	27	28	29	28	28	27	30	31	31	48	48	84	81		
SNMP2*Msex*	24	26	29	29	30	29	29	28	28	31	31	30	31	30	
SNMP2*Bmor*	29	30	30	29	31	30	30	30	29	29	29	28	29	30	68

Percent Identity: p-distance (fractional absolute identity differences) values calculated using Mega4, following gap deletion of alignment shown in Fig. 3. Values were converted to percent identity. For species abreviations, see Table 1 legend.

Table 3
dN/dS and Percent Identities.

dN/dS	SNMP1*Dmel*	SNMP1*Dpse*				
SNMP1*Dpse*	0.08(0.12/1.57)	X				
SNMP1*Aaeg*	#(0.44/#)	#(0.42/#)	X			
SNMP1*Agam*	#(0.46/#)	#(0.44/#)	0.09(0.18/1.91)	X		
SNMP1a*Cpip*	#(0.48/#)	0.26(0.47/1.84)	0.15(0.23/1.50)	0.18(0.26/1.42)	X	
SNMP1b*Cpip*	#(0.46/#)	0.18(0.40/2.22)	0.12(0.18/1.49)	0.12(0.19/1.62)	0.30(0.23/0.77)	X
SNMP1c*Cpip*	#(0.46/#)	0.17(0.44/2.64)	0.10(0.17/1.75)	0.13(0.20/1.54)	0.27(0.21/0.79)	0.15(0.12/0.81)

dN/dS	SNMP2*Dmel*	SNMP2*Dpse*	SNMP2*Aaeg*	SNMP2*Agam*
SNMP2*Dpse*	0.03(0.05/1.85)	X		
SNMP2*Aaeg*	0.20(0.51/2.58)	0.25(0.50/2.04)	X	
SNMP2*Agam*	#(0.51/#)	0.30(0.51/1.75)	0.08(0.12/1.57)	X
SNMP2*Cpip*	#(0.49/#	0.25(0.50/1.98)	0.04(0.07/1.74)	0.09(0.13/1.46)

dN/dS: ratios are shown, followed by actual dN and dS values calculated using Mega 4 (Tamura et al., 2007). "#" indicates uncomputed values due to synonymous values approaching saturation. Gaps, start and stop codons, and split codons at exon boundaries were removed for this analysis.

mRNA expression, not the expression of a functional protein; selection-based analysis, while indirect, assumes functional protein expression. Our study suggests a limit for the applicability of this approach in evolutionary distance (time) between compared species (Fig. 1B): comparisons between fly species (65–43 Mya) and between mosquito species (200–245 Mya) generated values, but many comparisons between flies and mosquitoes (~240 Mya) failed due to saturation of changes within sites (the formulae for dS and dN approaches infinity as the number of changes approaches saturation). Thus the analysis may be limited to comparing species which diverged within the past 200–250 million years.

Members of the SNMP1 and SNMP2 sub-clades of the insect CD36 gene family appear to be represented at least throughout the holometabolous lineage. We hope this study will provide the community with information that will encourage further study of these genes in a broad range of species. SNMPs appear to have important functions in chemoreception; comparative analysis should significantly contribute to clarifying those functions.

Acknowledgements

This work was supported by the National Science Foundation (IBN0212510) (to R.G.V.), and USDA Cooperative agreement 58-1275-7-360 (to R.G.V.) funded in part by a grant from the Deployed War-Fighter Protection (DWFP) Research Program funded by the US Department of Defense through the Armed Forces Pest Management Board (AFPMB) (to J.D.D).

Appendix. Supplementary materials

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ibmb.2009.03.007

References

Benton, R., Vannice, K.S., Vosshall, L.B., 2007. An essential role for a CD36-related receptor in pheromone detection in *Drosophila*. Nature 450, 289–293.
Calder, P.C., Deckelbaum, R.J., 2006. CD36: taste the difference? Curr. Opin. Clin. Nutr. Metab. Care 9, 77–78.
Chomczynski, P., Sacchi, N., 1987. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal. Biochem. 162, 156–159.
Febbraio, M., Silverstein, R.L., 2007. CD36: implications in cardiovascular disease. Int. J. Biochem. Cell. Biol. 39, 2012–2030.
Foley, D.H., Bryan, J.H., Yeates, D., Saul, A., 1998. Evolution and systematics of *Anopheles*: insights from a molecular phylogeny of Australasian mosquitoes. Mol. Phylog. Evol. 9, 262–275.
Forstner, M., Gohl, T., Gondesen, I., Raming, K., Breer, H., Krieger, J., 2008. Differential expression of SNMP-1 and SNMP-2 proteins in pheromone-sensitive hairs of moths. Chem. Senses 33, 291–299.
Franc, N.C., Dimarcq, J.L., Lagueux, M., Hoffmann, J., Ezekowitz, R.A., 1996. Croquemort, a novel *Drosophila* hemocyte/macrophage receptor that recognizes apoptotic cells. Immunity 4, 431–443.
Fukuwatari, T., Kawada, T., Tsuruta, M., Hiraoka, T., Iwanaga, T., Sugimoto, E., Fushiki, T., 1997. Expression of the putative membrane fatty acid transporter (FAT) in taste buds of the circumvallate papillae in rats. FEBS Lett. 414, 461–464.
Gaillard, D., Laugerette, F., Darcel, N., El-Yassimi, A., Passilly-Degrace, P., Hichami, A., Khan, N.A., Montmayeur, J.P., Besnard, P., 2008a. The gustatory pathway is involved in CD36-mediated orosensory perception of long-chain fatty acids in the mouse. FASEB J. 22, 1458–1468.

Gaillard, D., Passilly-Degrace, P., Besnard, P., 2008b. Molecular mechanisms of fat preference and overeating. Ann. N Y Acad. Sci. 1141, 163–175.
Gilbertson, T.A., Liu, L., Kim, I., Burks, C.A., Hansen, D.R., 2005. Fatty acid responses in taste cells from obesity-prone and -resistant rats. Physiol. Behav. 86, 681–690.
Giovannucci, D.R., Stephenson, R.S., 1999. Identification and distribution of dietary precursors of the *Drosophila* visual pigment chromophore: analysis of carotenoids in wild type and ninaD mutants by HPLC. Vis. Res. 39, 219–229.
Grimaldi, D., Engel, M.S., 2005. Evolution of the Insects. Cambridge University Press, Cambridge, UK.
Hughes, A.L., Nei, M., 1988. Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. Nature 335, 167–170.
International Silkworm Genome Consortium, 2008. The genome of a lepidopteran model insect, the silkworm *Bombyx mori*. Insect Biochem. Mol. Biol. 38, 1036–1045.
Jiménez, L.V., Kang, B.K., deBruyn, B., Lovin, D.D., Severson, D.W., 2004. Characterization of an *Aedes aegypti* bacterial artificial chromosome (BAC) library and chromosomal assignment of BAC clones for physical mapping quantitative trait loci that influence *Plasmodium* susceptibility. Insect Mol. Biol. 13, 37–44.
Jin, X., Ha, T.S., Smith, D.P., 2008. SNMP is a signaling component required for pheromone sensitivity in *Drosophila*. Proc. Natl. Acad. Sci. U S A 105, 10996–11001.
Kiefer, C., Sumser, E., Wernet, M.F., Von Lintig, J., 2002. A class B scavenger receptor mediates the cellular uptake of carotenoids in *Drosophila*. Proc. Nat. Acad. Sci. USA 99, 10581–10586.
Krieger, J., Raming, K., Dewer, Y.M.E., Bete, S., Conzelmann, S., Breer, H., 2002. A divergent gene family encoding candidate olfactory receptors of the moth *Heliothis virescens*. Eur. J. Neurosci. 16, 619–628.
Krzywinski, J., Grushko, O.G., Besansky, N.J., 2006. Analysis of the complete mitochondrial DNA from *Anopheles funestus*: an improved dipteran mitochondrial genome annotation and a temporal dimension of mosquito evolution. Mol. Phylo. Evol. 39, 417–423.
Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., Higgins, D.G., 2007. Clustal W and Clustal X version 2.0. Bioinformatics 23, 2947–2948.
Lobo, N.F., Campbell, K.S., Thaner, D., Debruyn, B., Koo, H., Gelbart, W.M., Loftus, B.J., Severson, D.W., Collins, F.H., 2007. Analysis of 14 BAC sequences from the *Aedes aegypti* genome: a benchmark for genome annotation and assembly. Genome Biol. 8, R88.
Miller, N.E., Litvack, R.A., Vogt, R.G., 2007. Tissue-specific survey of *Drosophila melanogaster* CG7000 (SNMP1) GenBank accession numbers: EF596938 (head), EF560171 (wing), EF560170 (leg).
Nei, M., Gojobori, T., 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. 3, 418–426.
Nene, V., the Aedes Genome Sequencing Consortium, 2007. Genome sequence of *Aedes aegypti*, a major arbovirus vector. Science 316, 1718–1723.
Nichols, Z., Vogt, R.G., 2008. The SNMP/CD36 gene family in Diptera, Hymenoptera and Coleoptera: *Drosophila melanogaster, D. pseudoobscura*, *Anopheles gambiae*, *Aedes aegypti*, *Apis mellifera*, and *Tribolium castaneum*. Insect Biochem. Mol. Biol. 38, 398–415.
O'Grady, P.M., 1999. Reevaluation of phylogeny in the *Drosophila obscura* species group based on combined analysis of nucleotide sequences. Mol. Phylogen. Evo. 12, 124–139.
Philips, J.A., Rubin, E.J., Perrimon, N., 2005. *Drosophila* RNAi screen reveals CD36 family member required for mycobacterial infection. Science 309, 1251–1253.
Rac, M.E., Safranow, K., Poncyljusz, W., 2007. Molecular basis of human CD36 gene mutations. Mol. Med. 13, 288–296.
Rasmussen, J.T., Berglund, L., Rasmussen, M.S., Petersen, T.E., 1998. Assignment of disulfide bridges in bovine CD36. Eur. J. Biochem. 257, 488–494.
Robertson, H.M., Wanner, K.W., 2006. The chemoreceptor superfamily in the honey bee, *Apis mellifera*: expansion of the odorant, but not gustatory, receptor family. Genome Res. 2006 (16), 1395–1403.
Rogers, M.E., Sun, M., Lerner, M.R., Vogt, R.G., 1997. Snmp-1, a novel membrane protein of olfactory neurons of the silk moth *Antheraea polyphemus* with homology to the CD36 family of membrane proteins. J. Biol. Chem. 272, 14792–14804.
Rogers, M.E., Steinbrecht, R.A., Vogt, R.G., 2001a. Expression of SNMP-1 in olfactory neurons and sensilla of male and female antennae of the silkmoth *Antheraea polyphemus*. Cell Tiss. Res. 303, 433–446.
Rogers, M.E., Krieger, J., Vogt, R.G., 2001b. Antennal SNMPs (Sensory Neuron Membrane Proteins) of Lepidoptera define a unique family of invertebrate CD36-like proteins. J. Neurobiol. 49, 47–61.
Shizuya, H., Birren, B., Kim, U.J., Mancino, V., Slepak, T., Tachiiri, Y., Simon, M., 1992. Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. Proc. Natl. Acad. Sci. USA. 89, 8794–8797.
Stuart, L.M., Deng, J., Silver, J.M., Takahashi, K., Tseng, A.A., Hennessy, E.J., Ezekowitz, R.A., Moore, K.J., 2005. Response to *Staphylococcus aureus* requires CD36-mediated phagocytosis triggered by the COOH-terminal cytoplasmic domain. J. Cell. Biol. 170, 477–485.
Tamura, K., Subramanian, S., Kumar, S., 2004. Temporal patterns of fruit fly (*Drosophila*) evolution revealed by mutation clocks. Mol. Biol. Evol. 21, 36–44.
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.
Torrents, D., Suyama, M., Zdobnov, E., Bork, P., 2003. Genome- wide survey of human pseudogenes. Genome Res. 13, 2559–2567.
Tunstall, N.E., Sirey, T., Newcomb, R.D., Warr, C.G., 2007. Selective pressures on *Drosophila* chemosensory receptor genes. J. Mol. Evol. 64, 628–636.
Vogt, R.G., 2003. Biochemical diversity of odor detection: OBPs, ODEs and SNMPs. In: Blomquist, G.J., Vogt, R.G. (Eds.), Insect Pheromone Biochemistry and Molecular Biology. Elsevier Academic Press, London, pp. 391–446.
Wang, T., Jiao, Y., Montell, C., 2007. Dissection of the pathway required for generation of vitamin A and for *Drosophila* phototransduction. J. Cell Biol. 177, 305–316.
Yang, J., O'Tousa, J.E., 2007. Cellular sites of *Drosophila* NinaB and NinaD activity in vitamin A metabolism. Mol. Cell Neurosci. 35, 49–56.
Zhang, X., Rogers, M., Tian, H., Zhang, X., Zou, D.-J., Liu, J., Ma, M., Shepherd, G.M., Firestein, S.J., 2004. High-throughput microarray detection of olfactory receptor gene expression in the mouse. Proc. Natl. Acad. Sci. USA 101, 14168–14173.