Available online at www.sciencedirect.com
ScienceDirect
Insect Biochemistry and Molecular Biology 38 (2008) 398–415
Insect Biochemistry and Molecular Biology
www.elsevier.com/locate/ibmb

The SNMP/CD36 gene family in Diptera, Hymenoptera and Coleoptera: *Drosophila melanogaster*, *D. pseudoobscura*, *Anopheles gambiae*, *Aedes aegypti*, *Apis mellifera*, and *Tribolium castaneum*

Zachary Nichols, Richard G. Vogt*

Department of Biological Sciences, University of South Carolina, Columbia SC 29208, USA

Received 10 September 2007; received in revised form 20 October 2007; accepted 7 November 2007

Abstract

Sensory neuron membrane proteins (SNMPs) are membrane bound proteins initially identified in olfactory receptor neurons of Lepidoptera and are thought to play a role in odor detection; SNMPs belong to a larger gene family characterized by the human protein CD36. We have identified 12–14 candidate SNMP/CD36 homologs from each of the genomes of *Drosophila melanogaster*, *D. pseudoobscura*, *Anopheles gambiae* and *Aedes aegypti* (Diptera), eight candidate homologs from *Apis mellifera* (Hymenoptera), and 15 from *Tribolium castaneum* (Coleoptera). Analysis (sequence similarity and intron locations) suggests that the insect SNMP/CD36 genes fall into three major groups. Group 1 includes the previously characterized *D. melanogaster* emp (epithelial membrane protein). Group 2 includes the previously characterized *D. melanogaster croquemort, ninaD, santa maria*, and *peste*. Group 3 genes include the SNMPs, which fall into two subgroups referred to as SNMP1 and SNMP2. *D. melanogaster* SNMP1 (CG7000) shares both significant sequence similarity and five of its six intron insertion sites with the lepidopteran *Bombyx mori* SNMP1. The topological conservation of this gene family within the three major holometabolous lineages indicates that it predates the coleopteran and hymenoptera/dipera/lepidoptera split 300+ million years ago. The current state of knowledge of the characterized insect members of this gene family is discussed.

Keywords: Olfaction; Chemoreception; Evolution; Insect; Pheromone

1. Introduction

Sensory neuron membrane proteins (SNMPs) were identified and characterized as 2-transmembrane domain membrane proteins (519–525 amino acids) of Lepidoptera (moth) olfactory receptor neurons and suggested to play an important but functionally unknown role in odor detection (Rogers et al., 1997, 2001a, b). A Diptera (fly) SNMP homolog, CG7000, has recently been shown to be required for the detection of an aggregation pheromone (Benton et al., 2007). SNMPs belong to a larger family of two-transmembrane domain proteins characterized by the human fatty acid transporter (FAT) CD36 which has a broad range of described roles including 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; Calder and Deckelbaum, 2006; Febbraio and Silverstein, 2007; Fukuwatari et al., 1997; Rac et al., 2007). In addition to the SNMPs, several other CD36 homologs have been described in insects (*Drosophila melanogaster*), including epithelial membrane protein (emp), Croquemort, Peste, NinaD and Santa Maria; similar to CD36, these proteins have functions that include cytoadhesion, caratenoid transport, and chemoreception (see Section 4).

This current report characterizes the SNMP/CD36 family of genes/proteins in the genomes of four species of Diptera (the flies *D. melanogaster* and *D. pseudoobscura*

Abbreviations: CD36, cluster determinant 36; Crq, croquemort; Emp, epithelial membrane protein; NinaD, neither inactivation nor after-potential D, Pes, Peste; SCRB, scavenger receptor type B; SNMP, sensory neuron membrane protein; MYA, million years ago.

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

0965-1748/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ibmb.2007.11.003

and the mosquitoes *Anopheles gambiae* and *Aedes aegypti*), one species of Hymenoptera (the honeybee *Apis mellifera*) and one species of Colepotera (the red flower beetle *Tribolium castaneum*). The *D. melanogaster* genes were previously reported in Rogers et al. (2001b).

The chosen dipteran species represent significant ranges of well established evolutionary distances (Fig. 1): *D. melanogaster* and *D. pseudoobscura* represent the deepest lineage split among the available drosophilid genomes; *An. gambiae* and *Ae. aegypti* represent the deepest lineage split among available mosquito genomes. The coleopteran, hymenopteran and dipteran/lepidopteran lineages represent the three major holometabolous lineages (Grimaldi and Engel, 2005) and thus offer the opportunity to generalize the observed gene patterns across the entire holometabolous group. Collectively, these genomes offer a view of the evolutionary dynamics of the SNMP/CD36 gene family over a variety of time scales ranging from the divergence of the *D. melanogaster* and *D. pseudoobscura* lineages (43–65 MYA, Tamura et al., 2004) to the divergence of the coleopteran and hymenopteran/dipteran/lepidopteran lineages (300+ MYA, Grimaldi and Engel, 2005).

2. Methods

2.1. Identification of genes and gene structures

Amino acid sequences from 13 previously reported *D. melanogaster* SNMP/CD36 homologs (Rogers et al., 2001b) were used in a BLAST-P search of the *D. pseudoobscura*, *An. gambiae, A. mellifera*, and *T. castaneum* genomes using the NCBI website. The *An. gambiae* sequences were subsequently used in a BLAST-P search of the *Ae. aegypti* genome using the NCBI website. Sequences with an *e*-value of less than 0.005 were selected (Karlin and Altschul, 1990), and these were re-blasted to find any additional presumptive homologs within respective species using the same criterion. All sequences contained at least part of the approximately 400 amino acid CD36 motif, identified by Conserved Domain function of the BLAST-P analysis. Most sequences contained the full motif, shorter sequences not containing the full CD36 motif may be incomplete. The full set of sequences identified are listed in Table 1. Gene positions, orientations and intron insertion sites for *D. melanogaster* and *An. gambiae* were determined using the NCBI Genome Map Viewer (http://www.ncbi.nlm.nih.gov/mapview/) and the sequence view link (SV) on the Map Viewer associated with a specific gene. Gene positions and orientations for *D. pseudoobscura* and *Ae. aegypti* were determined from the annotated scaffolds associating with each gene; intron insertion sites were determined by translating the indicated DNA sequence and noting the exon/intron locations.

Two of the *An. gambiae* sequences were combined out of multiple gene entries. EAA07966 (SCRB11) and EAA07986 (SCRB1) were combined in linear fashion; these computer annotated entries were contiguous in the genome, separated by about 70 bp; alignment with other SNMP homologs suggested SCRB11 comprised the 5′ half and SCRB1 comprised the 3′ half of the CD36 motif. Similarly, EAA11087 (SCRB12), EAA11632 (SCRB2), and EAA11629 (SCRB4) were also combined in linear fashion; these computer annotated entries were contiguous in the genome, separated by 2500 bp (SCRB12-SCRB2) and 6000 bp (SCRB2-SCRB4), and alignment with other SNMP homologs suggested SCRB12 comprised the 5′ third, SCRB2 the middle third and SCRB4 the 3′ third of the CD36 motif. Another gene is present within the proposed intron between SCRB2 and SCRB4, but in the opposite orientation of the SCRB segments (XM_315732, ENSANGP00000015904).

One *Ae. aegypti* sequence, referred to here as Aa-SNMP2, was combined from three gene entries: EAT42493, EAT42492, and EAT42490 (5′ to 3′). All three segments are in the same orientation, the first and second separated by 12,062 bp and the second and third by 12,635 bp. Structural analysis in the Blast data suggest

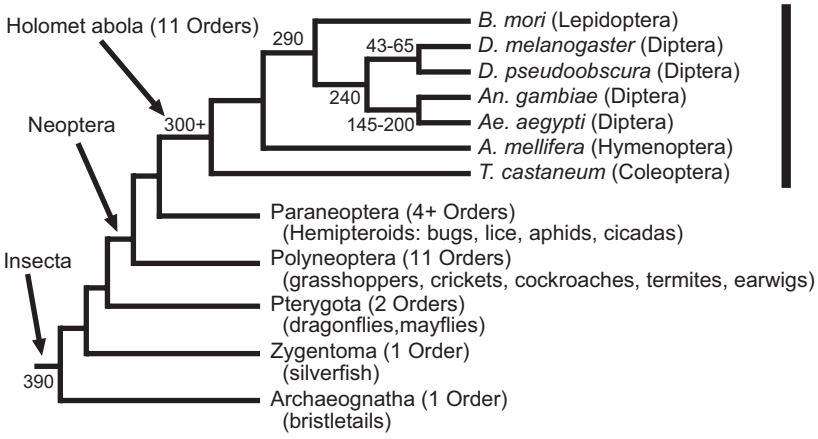

Fig. 1. Phylogenetic relationships of the insect species and groups characterized in this study. Numbers indicate time (MYA) since lineages shared a common ancestor. Organization and times are from Grimaldi and Engel (2005) (see Section 4).