Comparative Genomics of Odorant Binding Proteins in Anopheles gambiae, Aedes aegypti, and Culex quinquefasciatus

Malini Manoharan1,2,3, Matthieu Ng Fuk Chong1, Aurore Vaitinadapoulé1,2, Etienne Frumence1, Ramanathan Sowdhamini2,*, and Bernard Offmann1,4,*

1Université de La Reunion, DSIMB, INSERM UMR-S 665, La Reunion, France
2National Centre for Biological Sciences, Tata Institute for Fundamental Research, GKVK Campus, Bangalore, Karnataka, India
3Manipal University, Madhav Nagar, Manipal, Karnataka, India
4Université de Nantes, UFIP CNRS FRE 3478, Nantes, France
*Corresponding authors: E-mail: bernard.offmann@univ-nantes.fr; bernardoffmann@gmail.com; mini@ncbs.res.in.

Accepted: December 21, 2012

Abstract

About 1 million people in the world die each year from diseases spread by mosquitoes, and understanding the mechanism of host identification by the mosquitoes through olfaction is at stake. The role of odorant binding proteins (OBPs) in the primary molecular events of olfaction in mosquitoes is becoming an important focus of biological research in this area. Here, we present a comprehensive comparative genomics study of OBPs in the three disease-transmitting mosquito species Anopheles gambiae, Aedes aegypti, and Culex quinquefasciatus starting with the identification of 110 new OBPs in these three genomes. We have characterized their genomic distribution and orthologous and phylogenetic relationships. The diversity and expansion observed with respect to the Aedes and Culex genomes suggests that the OBP gene family acquired functional diversity concurrently with functional constraints posed on these two species. Sequences with unique features have been characterized such as the "two-domain OBPs" (previously known as Atypical OBPs) and "MinusC OBPs" in mosquito genomes. The extensive comparative genomics featured in this work hence provides useful primary insights into the role of OBPs in the molecular adaptations of mosquito olfactory system and could provide more clues for the identification of potential targets for insect repellants and attractants.

Key words: odorant binding proteins, OBP, mosquito, Culex quinquefasciatus, Aedes aegypti, Anopheles gambiae, olfaction, phylogeny.

Introduction

The spread of infectious diseases among humans is mediated primarily by the world’s most dangerous animal, the mosquitoes among which the anthropophilic mosquitoes such as Anopheles gambiae, Anopheles funestus, Aedes albopictus, Aedes aegypti, and Culex quinquefasciatus are the most effective transmitters of viruses and parasites. They are responsible for the spread of a number of life-threatening diseases such as malaria, dengue, and West Nile encephalitis and recently Chikungunya with a lower mortality rate compared with the other diseases. According to the World Health Organization, global climate change is expanding mosquitoes range, heightening the risk of disease for millions of additional people. Primary prevention is one of the most important aspects to subside the spread of diseases either by controlling the population of these vectors or by preventing the interaction between the vector and the host.

Understanding the molecular mechanism for human host recognition mediated by olfaction would help in identifying new strategies for the prevention of the primary contact. Volatile products secreted by the human host in the process of metabolism are responsible for the attraction of these vectors to the host. The ability of recognizing and discriminating thousands of odorant molecules in insects as in mammals relies on specialized chemosensitive neural cells expressing olfactory receptor proteins (ORs) which reside within segregated compartments called sensilla. Each sensillum is a hair-like structure bathed in the sensillum lymph which contains a
number of secreted proteins (McKenna et al. 1994; Pikielny et al. 1994; Wang et al. 1999). The odorant binding proteins (OBPs) are found to be important water-soluble components of this sensillum lymph. It was first identified in the moth as pheromone binding proteins (PBPs) (Vogt and Riddiford 1981). These globular proteins are believed to bind different odorant molecules (Plettner et al. 2000), owing to their high divergence within the family, and transport them to their respective olfactory receptors triggering the mechanism of olfaction (Pelosi and Maida 1995).

The arthropod OBPs form a large specific multi-gene family. They are 10–30 kDa globular and water-soluble proteins that are characterized by a specific six a-helical domain comprising of six highly conserved cysteines that have distinct disulphide connectivities. These structural features are now considered the hallmark of this protein family (Calvo et al. 2002; Valenzuela et al. 2002; Calvo et al. 2006). OBPs have been identified in a number of insect species, including four dipterian species Drosophila melanogaster (Galindo and Smith 2001; Graham and Davies 2002; Hekmat-Scafe et al. 2002; Valenzuela et al. 2002; Zhou et al. 2004; Vieira et al. 2007; Vieira and Rozas 2011), A. gambiae (Vogt 2002; Xu et al. 2003; Zhou et al. 2004; Li et al. 2005; Vieira and Rozas 2011), Aed. aegypti (Zhou et al. 2008), and C. quinquefasciatus (Pelletier and Leal 2009, 2011). These proteins are very divergent in terms of the sequences within the family, and sequence identities between the family members from the different species could drop as low as 8% (Vieira and Rozas 2011). In Drosophila, a subgroup of (i) OBPs lacking two of the six conserved cysteines, called MinusC OBPs and (ii) OBPs carrying additional conserved cysteines called PlusC OBPs have been identified (Hekmat-Scafe et al. 2002). The MinusC OBPs typically lack the second and fifth Cys residues. However, this definition appears to be somewhat ambiguous, since there are three Drosophila OBPs among this cluster which contain all the six hallmark cysteines (Pelosi and Maida 1995). MinusC OBPs have never been described to date in mosquito genomes.

In mosquitoes, three subfamilies of OBP genes have been characterized so far: (i) the Classic OBPs that carry the six conserved cysteines characteristic motif of the OBP family; (ii) the PlusC OBPs that have the same conserved cysteines and disulphide connectivity but which contain six additional cysteines with novel disulphide connectivities; (iii) the Atypical OBPs that are among the longest known OBPs and that have initially been described as containing a single Classic OBP domain in its N-terminal extended by a less characterized C-terminal extension. Very recently, it was shown that Atypical OBPs comprises two domains that are in fact homologous to the Classic OBP domain and were hence considered as “dimer OBPs” (Vieira and Rozas 2011).

In A. gambiae and Aed. aegypti, OBPs from the three different subfamilies have been reported to date while in C. quinquefasciatus, only the Classic and PlusC members of this family have been reported so far (Pelletier and Leal 2009, 2011). Atypical OBPs have not yet been reported in this genome.

An additional multi-gene family, known as D7 salivary proteins, is known to be distantly related to the arthropod OBP superfamily (Calvo et al. 2002, 2006, 2009). There are two types of D7 salivary proteins in the mosquito genome, the short and the long forms which contain one and two OBP-like domains, respectively (Valenzuela et al. 2002; Kalume et al. 2005; Choumet et al. 2007). The available structures of the D7 proteins indicate that the domains adopt a similar fold to the OBP domains but decorated with additional structural features and a seventh helix. In the two-domain D7 protein, the C-terminal OBP-like domain has been shown to bind to biogenic amines in A. gambiae and Aed. aegypti (Mans et al. 2008; Calvo et al. 2009), while the N-terminal domain in Aed. aegypti was shown to have a specific bioactive lipid-binding activity (Calvo et al. 2009). These members serve as important representatives for the construction of phylogenetic trees serving as outgroups for the OBP gene family in the current analysis.

This work describes the identification and extension of OBPs in the mosquito genomes of A. gambiae, Aed. aegypti, and C. quinquefasciatus. We provide a significant extension of the OBP gene family to a total of 110 new members in these three genomes and report the presence of all three classes of OBPs in the three mosquito genomes. In particular, we identified Atypical class of OBPs in C. quinquefasciatus. We further confirm that “Atypical OBPs” are composed of two domains that are homologous to Classic OBPs and provide in-depth characterization of their origin and structural features. This work also provides for a comprehensive and robust subclassification of the different OBP classes through structure-based alignments and phylogenetic analysis which could possibly reflect on the functional divergence of these proteins. We also provide a detailed primary structural and phylogenetic characterization of all these novel OBP subtypes. An extensive set of supplementary materials that detail our analyses and results are provided.

**Results**

**Extension of OBPs Family in All Three Mosquito Genomes**

In the already published works, 65 OBPs from A. gambiae (Vogt 2002; Xu et al. 2003; Zhou et al. 2004, 2008), 64 from Aedes aegypti (Zhou et al. 2008), and 53 OBPs from C. quinquefasciatus (Pelletier and Leal 2009) were previously identified. These OBPs have been characterized by these groups into three main subfamilies Classic, PlusC, and Atypical based on sequence features (fig. 1). Only very recently, Vieira and Rozas (2011) added four new putative genes to the A. gambiae OBP gene repertoire and 13 PlusC OBPs to the C. quinquefasciatus genome. These new genes
FIG. 1.—Cysteine conservation patterns across the different subfamilies and subgroups of OBPs from *Anopheles gambiae*, *Aedes aegypti*, and *Culex quinquefasciatus* genomes. The six conserved cysteines in GOBP domain are denoted C1–C6. The six additional cysteines in the C-term of the Atypical OBPs are denoted C1^0–C6^0. The lines connecting cysteines represent the disulphide bonds and dotted lines represent the additional disulphide bonds in the PlusC OBPs.
were also identified by our sequence searches and bioinformatics analysis (see Materials and Methods) (table 1 and supplementary table S1a–e, Supplementary Material online). The fasta sequences of the identified genes are available for download as supplementary material.

In this study, a major expansion is provided in the Atypical OBP subfamily of the mosquitoes where 31 new members (AagOBP6 to AagOBP68, AaeOBP67 to AaeOBP114, CquOBP54 to CquOBP112) in all three mosquito genomes. Detailed results are provided in accompanying supplementary tables S1a–e, Supplementary Material online.

Interestingly, sequences that lack C2 and C5 cysteines were observed in the alignments. OBPs which lack these two particular cysteines, called the MinusC OBPs, have been characterized and expressed in other insect genomes such as Drosophila, Bombyx mori, Tribolium castaneum, and Apis mellifera (Vieira and Rosas 2011), but their presence in the mosquito genome has not been shown previously. AaeOBP78 from Aed. aegypti and 15 proteins from C. quinquefasciatus (CquOBP59–CquOBP62, CquOBP64–CquOBP74) were found to lack these two cysteines. As all these sequences retained the N-terminal signal peptide or the presence of the PBP/GOBP domain, they were retained in our analysis as MinusC OBPs (supplementary tables S3 and S4, Supplementary Material online).

We also observed interesting cysteine conservation patterns among the Atypical OBPs. The Atypical OBPs were previously described as proteins that hold a Classic OBP domain in the N-terminal end with an uncharacterized C-terminal domain. However, the close analysis of the extended C-terminal end of Atypical members highlighted the presence of six additional cysteines conserved within this subfamily, with a cysteine spacing pattern very similar to the conserved cysteines (C1–C6) at their N-terminal end. The observed cysteine conservation pattern in the case of the Atypical OBPs is purely the reflection of the annotation of new members in this subfamily and has never been described before to our knowledge. We hence propose to annotate these cysteines as C1’–C6’.

### Table 1

Identification of OBPs in Anopheles gambiae, Aedes aegypti, and Culex quinquefasciatus Genomes

<table>
<thead>
<tr>
<th>Subfamily</th>
<th>Classic</th>
<th>PlusC</th>
<th>Atypical</th>
<th>Not Determined</th>
<th>New</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. gambiae</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Previously reported</td>
<td>29</td>
<td>16</td>
<td>16</td>
<td></td>
<td>69</td>
<td></td>
</tr>
<tr>
<td>Newly identified</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aed. aegypti</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Previously reported</td>
<td>33</td>
<td>17</td>
<td>14</td>
<td></td>
<td>111</td>
<td></td>
</tr>
<tr>
<td>Newly identified</td>
<td>6</td>
<td>10</td>
<td>31</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. quinquefasciatus</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>109</td>
<td></td>
</tr>
<tr>
<td>Previously reported</td>
<td>48</td>
<td>12</td>
<td>26</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Newly identified</td>
<td>21</td>
<td>12</td>
<td>26</td>
<td></td>
<td>2</td>
<td></td>
</tr>
</tbody>
</table>

Notes.—The table shows statistics of previously and newly identified OBP members (AagOBP6 to AagOBP68, AaeOBP67 to AaeOBP114, CquOBP54 to CquOBP112) in all three mosquito genomes. Detailed results are provided in accompanying supplementary tables S1a–e, Supplementary Material online.

We hence propose to annotate these cysteines as C1’–C6’. This remarkable conservation of cysteines is believed to hold important evolutionary information (Thangudu et al. 2005, 2008). Following this, we characterized the homologues of each of the two domains and identified their closest classic OBP homologue in their corresponding genomes and also the Drosophila genome which confirms that the Atypical OBPs are indeed "two-domain OBPs." It is noteworthy that within the Atypical (two-domain) subfamily, a distinctive subtype called matype2 (see below and fig. 1) showed the presence of only six cysteines (C1, C3, C4, C6, C4’, and C6’), when compared with the other subtypes which
carry the 12 cysteines. The Cys conservation pattern at the N-terminal domain of the OBP is similar to the MinusC OBPs; however, the C-terminal domain is found to have lost more cysteines comparatively.

Analysis of OBP Genes: Orthology across the Three Genomes and Their Corresponding Distribution

We investigated the orthology and gene distribution of OBPs in three genomes. Assembled genome is only available for *A. gambiae* at the date of this work in Ensembl Genomes and VectorBase 3.4 version. The chromosomal mapping for each of the OBP genes in *Anopheles* is hence known with precision (fig. 2). Their chromosomal distribution in the *Anopheles* genome is centrally featured in supplementary fig. S1a–e and further referenced in supplementary table S1a, Supplementary Material online. Though the syntenic relationship between the chromosome arms in *A. gambiae* and their corresponding orthologous chromosome arms in *Culex* and *Aedes* was established by Arensburger et al. (2010) with the help of genetic markers (supplementary table S2, Supplementary Material online), the genomic data of these two *Culicinae* species are only available in the form of supercontigs fragments (Nene et al. 2007; Arensburger et al. 2010) and are yet to be assembled. In these two genomes, a few supercontigs (about 10%) harbor markers that allow their chromosomal localization (Arensburger et al. 2010). Very few of these anchor supercontigs hosted OBP genes. Most supercontigs containing OBP genes did not harbor any genomic markers, hence cannot be assigned to a chromosome in *Aedes* and *Culex*. However, in many cases, direct orthologs in the *Anopheles* genome could be identified (fig. 2, supplementary fig. S1a–e and supplementary table S1a, c, and e, Supplementary Material online). OBP orthologs have been identified using the reciprocal BLAST hit approach (Moreno-Hagelsieb and Latimer 2008) which is widely used in the detection of orthologs. As illustrated in figures 2 and 3 and in supplementary figure S1a–e, Supplementary Material online, three-way orthology (1:1:1) between OBP genes in each of the three genomes were identified in 31 cases while two-way orthology (1:1) between OBP genes from only two genomes were identified in 5 cases between *Anopheles* and *Culex*, 6 between *Anopheles* and *Aedes*, and 19 between *Aedes* and *Culex* (fig. 3), thus confirming the genetic proximity between the *Aed. aegypti* and *C. quinquefasciatus* species. Our proposed analysis was found to be in complete agreement with the microsynteny analysis described very recently in Pelletier and Leal (2011), thus indicating that the orthology detected may serve as the basis of further syntenic analysis.

Interestingly, the overwhelming majority of the OBP genes are organized in gene clusters in the three genomes (supplementary fig. S1a–e, Supplementary Material online). The clusters are mainly composed of gene duplicates. The genes in these genomic clusters hence share high sequence identity (data not shown) and are thereby phylogenetically very close (see below) as it is confirmed by inparalogy data from the inParanoid database (O’Brien et al. 2005). The extension of OBP gene repertoire in *Aed. aegypti* and *C. quinquefasciatus* with respect to *A. gambiae* was mainly driven by these gene duplication events which are more numerous in these two *Culicinae* species. There are a total of 12 OBP gene clusters in *Aed. aegypti* and 13 clusters in *C. quinquefasciatus* genomes when compared with 6 clusters in *A. gambiae*. The largest gene clusters are found in *Aedes* and *Culex*, and a few clusters contain as many as 12 genes. It is observed that 21 out of the 26 newly identified Atypical (two-domain) OBP genes from *C. quinquefasciatus* are in fact distributed into three main gene clusters (fig. 2 and supplementary fig. S1a–e, Supplementary Material online). Similarly, 10 out of the 12 newly identified PlusC proteins are distributed into three gene clusters.

Phylogeny-Based OBP Clusters

As expected and as already reported previously, OBP family members showed high divergence. The average sequence identity between OBP genes in *A. gambiae*, *Aed. aegypti*, and *C. quinquefasciatus* are 12.5%, 12.8%, and 13.1%, respectively, and their phylogenetic tree (see Material and Methods) also indicated a high sequence divergence (supplementary fig. S2a–c, Supplementary Material online). However, the comparative analysis of the different subfamilies of the OBPs in the mosquito genome provided more meaningful clustering patterns within each subfamily of the OBP members. The analysis was done based on the sequence alignment and phylogenetic trees constructed using sequences from individual subfamilies from all the three mosquito genomes used in this analysis and the *Drosophila* OBPs (Hekmat-Scafe et al. 2002) in the case of the Classic members. A bootstrap consensus tree was constructed using the neighbor joining method (Saitou and Nei 1987) with all the Classic OBPs from the three mosquito genomes and the *D. melanogaster* with 1000 bootstrap replicates (fig. 4). The clustering of the various Classic OBPs into clusters based on significant bootstrap values (50% cutoff) revealed the possibility of 18 different subtypes. These clusters carried orthologous and paralogous sequences from the three genomes. Few members of the mosquito genomes clustered with *Drosophila* OBPs (Hekmat-Scafe et al. 2002), and these clusters were named after their closest *Drosophila* OBPs. Among these OS/EOs-F, *Pbprp1*, LUSH, OBPI9a, and *Pbprp4* have already been described previously (Xu et al. 2003; Zhou et al. 2008; Pelletier and Leal 2009). However, one member from *C. quinquefasciatus* in each of the two subtypes OS/EOs-F (*CquiOBP58*) and OBPI9a (*CquiOBP57*) have been annotated. The huge expansion of sequences (*CquiOBP25–CquiOBP42*) observed by Pelletier and Leal (2009) were found to be homologous to *AaegOBP57* and *AgamOBP13* and were indeed
FIG. 2.—Chromosomal localization of odorant binding proteins from Anopheles gambiae. Details about the gene names used here are shown in supplementary table S1a, Supplementary Material online. Shown to the right of the gene names are their direct three-way orthologs detected in the Aedes aegypti and Culex quinquefasciatus genomes by the reverse BLAST hit methodology. Classic OBPs are featured in blue, PlusC OBPs in red, and two-domain OBPs (or Atypical) in green. The chromosomes were drawn to scale in MapDraw software. The positions of and the distances between gene loci are indicated in megabases. When three-way orthology (1:1:1) was not detected but only two-way orthology (1:1), the two CquiOBP and AaegOBP orthologs are separately connected to the corresponding AgamOBP gene.
closely related to the Pbprp2/Pbprp5 of *Drosophila*. CquiOBP55 and AaegOBP83 identified in this analysis are orthologs of AgamOBP29 and homologous to OBP59a of *Drosophila* and have an unusually long sequence as recently mentioned by Vieira and Rozas (2011). Clustering of three orthologous OBP sequences AgamOBP9, AaegOBP22, and CquiOBP43 with the *Drosophila* MinusC members OBP99a, OBP44a, and OBP99b was observed with a considerable bootstrap support, among which OBP99a alone retains all the six cysteines, while the two others lack the C2 and C5 cysteines (see below). Among the *Drosophila* MinusC OBPs, three members of the MinusC subfamily (Obp83f, Obp99a, and Obp99d) retain all six conserved cysteines, whereas four members of the subfamily (Obp8a, Obp44a, Obp99b, and Obp99c) have C2 and C5 cysteines lacking. Therefore, the mosquito OBPs, which cluster with these *Drosophila* OBPs, do not represent true MinusC OBPs. The other clusters which do not have a close *Drosophila* homologue are named as mclassic1–9 (fig. 4 and supplementary fig. S3, Supplementary Material online). In addition to these subtypes, one group, displaying outstanding sequence features (supplementary fig. S3a, Supplementary Material online) with 16 members lacking C2 and C5 cysteines, has been named as “Bombyx mori” MinusC due to their homology with the *B. mori* MinusC sequences though its branch holds a bootstrap value of only 35%. This homology was determined using BLAST analysis and confirmed with the inParanoid eukaryotic ortholog database (O’Brien et al. 2005). Other subtype classifications of the Classic members were also similar to the clustering seen in the inParanoid database.

As shown in figure 5, the PlusC OBPs clustered as seven major phylogenetic clusters based on bootstrap cutoff value.
of 50%, but we further subdivided them into 11 subtypes (mplus1–mplus11). Indeed, though the interior node of mplus7–11 cluster hold a bootstrap value of 57%, we separated them as different subtypes because they clearly hold distinct sequence features (supplementary fig. S3b, Supplementary Material online). Furthermore, analysis of chromosomal localization of PlusC members from A. gambiae shows that mplus11 subtype members are specific to chromosome 3L while all other PlusC OBPs were specifically distributed on chromosome 2L. At this stage, it is difficult to interpret the molecular background behind this clustering.

The Classic subfamily members from the three genomes share an average sequence identity of 15.5%, while the PlusC OBPs share 17.3% average sequence identity. No distinct sequence features could be observed at the subfamily level (Classic, PlusC, and Atypical) because of high sequence divergence. Nevertheless, a close examination of the alignments for the different clusters which contain orthologous sequences from the three genomes within each subfamily indicates that the phylogenetic clusters established in this study tend to have specific sequence patterns (supplementary fig. S3a and b, Supplementary Material online). Some subgroups are

**Fig. 4.**—Unrooted phylogenetic tree of Classic OBPs in the three mosquito genomes and in *Drosophila melanogaster*. The *Anopheles gambiae*, *Aedes Aegypti*, and *Culex quinquefasciatus* members are colored in mustard, pink, and turquoise, respectively. The bootstrap values of the branches are indicated on the nodes in percentage values. The names of identified clusters inside the Classic OBPs subfamily are indicated on the branches. Detailed alignments of the members inside each cluster are provided in supplementary figure S3a, Supplementary Material online.
characterized by a very low average sequence identity like the B. mori MinusC subgroup within the Classic OBPs (21.5%), the mclassic9 (23.3%), or the mplus9 (24.3%), while other subgroups share significantly higher sequence identities like OS-E/OS-F (55.2%), Pbprp4 (60.2%), or mclassic4 (77.3%).

Sequence Specific Clustering of Two-Domain OBPs

The Atypical OBPs, unlike the Classic members, formed four major clusters based on bootstrap values which are named in this study matype1–matype4 (fig. 6) and showed distinct sequence features (supplementary fig. S3c, Supplementary Material online). The matype1 forms the smallest cluster among the four subtypes with two members from each genome, and this cluster is separated from the other three subtypes with high bootstrap values. The matype2 forms a distinctive type of Atypical members holding only a total of six cysteines (C1, C3, C4, C5, C1', and C6') out of the 12 conserved cysteines characteristic of the other subtypes of this subfamily (fig. 1 and supplementary fig. S3c).
Supplementary Material online). The matype2 still features to stand as a distinctive type with the presence of cysteines in the N-terminal domain lacking C2–C5 as previously described. The matype4 members unanimously hold a deletion of about 15 resides between the C1 and C2 which stands as the distinguishing feature of this subtype. The matype1 members are orthologous to AgamOBP39 that is located on chromosome 2R which is otherwise populated with Classic members supporting their close relation to the Classic members as observed in the phylogeny of the individual genomes. The matype2 members, intriguingly, share orthology with corresponding OBPs from *A. gambiae* that were mapped to chromosome X, whereas matype3 and matype4 members were sharing orthology with AgamOBPs distributed over chromosomes 3R and 3L.

**Figure 6.** Unrooted phylogenetic tree of Atypical odorant binding proteins in the three mosquito genomes. The *Anopheles gambiae*, *Aedes Aegypti*, and *Culex quinquefasciatus* members are colored in mustard, pink, and turquoise, respectively. The bootstrap values of the branches are indicated on the nodes in percentage values. The names of identified clusters inside the Atypical OBPs subfamily are indicated on the branches. Detailed alignments of the members inside each cluster are provided in supplementary figure S3c, Supplementary Material online.
Discussion

Evolutionary Aspects of OBP Gene Family in Mosquitoes

C. quinquefaciatus (Arensburger et al. 2010) and Aedes aegypti genomes (Nene et al. 2007) which code for 109 and 111 OBPs, respectively, have a significantly larger OBP genes repertoire than the related A. gambiae genome which harbors only 67 OBPs. A. gambiae belongs to the Anopheinae and Aedes aegypti and C. quinquefasciatus belong to the Culicinae subfamilies. These two subfamilies of mosquitoes are estimated to have diverged around ~120 My (Reidenbach et al. 2009). The increase in the number of genes indicates lineage-specific expansions of the OBP gene family among the mosquito species. The odorant binding gene family has been previously shown to adopt a birth and death model of evolution based on a number of factors which includes several gene gain and loss events in lineages, decrease in the number of orthology groups with increasing divergence times, and an uneven phylogenetic subfamily distribution across species (Vieira and Rozas 2011). Similar observations are made with respect to the OBP genes in Anopheles, Culex, and Aedes species, where a number of gene gain is observed in the Aedes and Culex species, uneven distribution of subfamilies is observed (where the MinusC Subfamily is absent in Anopheles), and the number of orthologous sequences are higher in the Aedes and Culex species and becomes lesser with respect to the Anopheles which is distant relatively to these two species comparatively. This provides further support to the already existing fact that the OBP gene family undergoes a birth and death model of evolution. Furthermore, the appearance of new genes and subfamilies in the Aedes and Culex could relate to the requirement of these genes for environmental adaptations by these species.

Evolution of MinusC Proteins in the Mosquito Genomes

The MinusC subfamily of OBPs was first identified in the Drosophila genome with some of its members lacking the second and fifth cysteine residues (Hekmat-Scafe et al. 2002) and later identified in other species which includes the Apis mellifera (Forét and Maleszka 2006) and B. mori (Gong et al. 2009 and Yoshizawa et al. 2011). In the case of mosquitoes, interestingly, MinusC OBPs are not present in the A. gambiae. But the MinusC OBPs appeared in Culicinae lineage of the mosquito OBPs. The close homology of these MinusC members with the B. mori MinusC OBPs suggests that they could have a common ancestor. Thus, it can be said that the MinusC OBPs appeared in the Endopterygota lineage of insects which is dated back to ~300 My and suggest that these OBPs appeared earlier in the evolution and not only in the Drosophilidae, Bombyx/Tribolium, and Apis lineages as believed earlier (Vieira and Rozas 2011). However, the absence of these OBPs in the A. gambiae is intriguing and suggests that they could have species-specific expansions. This in fact supports the birth and death model of evolution observed in the OBP family of proteins. Separately, the matype2 members belonging to the two-domain OBP subfamily, which retain only six cysteines, interestingly lack the C2 and C5 cysteines in the N-terminal domain. The absence of C2 and C5 cysteines being the characteristic feature of the MinusC proteins lays an important question on the evolutionary link between these members.

Atypical OBPs are Indeed Two-Domain OBPs

The increase in the number of Atypical OBPs in the three mosquito genome revealed important facets in this subfamily of proteins. We have shown that the Atypical OBPs so far identified only in the mosquitoes are indeed two-domain OBPs. This was also reported by Vieira and Rozas (2011), based on their phylogenetic analysis that they belong to the dimer OBP clade. We further provide evidence to this by characterizing each of the Atypical OBP domains by their closest homologue in the Classic OBP subfamily in their corresponding genomes. Very interestingly, the Classic OBP members, obtained as hits by each of these domains, were mainly found among the mclassic9, mclassic8, and Obp99a members (table 2). Atypical OBPs indeed are found to share closer phylogenetic proximity to OBPs from the mclassic9, mclassic8, and OBP99a phylogenetic clusters however not with significant bootstrap value because of sequence divergence (supplementary fig. S2a–c, Supplementary Material online). Moreover, Atypical gene clusters in A. gambiae are localized in close proximity to gene clusters that contains Classic OBPs from one of these three groups at the chromosomal level. On chromosome 2R, matype1 members AgamOBP39 and AgamOBP40 are localized at the level of the same gene cluster as the mclassic9 members AgamOBP11, AgamOBP12, and AgamOBP14. Likewise, on chromosome X, the matype2 OBPs AgamOBP34 to AgamOBP37 are localized proximal to AgamOBP8 and AgamOBP9 that belong to the Obp99a phylogenetic cluster. Similarly, on chromosome 3L, the matype3 AgamOBP31, AgamOBP44, and AgamOBP45 that form a gene cluster are in close proximity to AgamOBP22 which belongs to the mclassic8 group. Another interesting observation is that OBP members from these three phylogenetic clusters (mclassic9, mclassic8, and Obp99a) are closely related to the MinusC group of proteins in the Drosophila genome (fig. 4), and it has been established that the Drosophila Dimer OBPs B3cd and B3ef (Zhou et al. 2004), which are proteins that hold two OBP domains, are related to these Drosophilidae MinusC proteins.

The recent publication of a functional dimer in the C. quinquefasciatus genome (Mao et al. 2010) supports the current important speculations on Atypical members, indicating the importance of the presence of two-domain proteins in the binding of relatively large ligands. Thus, it is confirmed that the Atypical OBP members are indeed two-domain
### Table 2
Analysis of the Two Putative OBP Domains (N-term and C-term) of Atypical OBPs from Anopheles gambiae, Aedes Aegypti, and Culex quinquefasciatus

<table>
<thead>
<tr>
<th>Mosquito Atypical OBP</th>
<th>ID</th>
<th>N-term Phylogenetic Subgroup</th>
<th>C-term Phylogenetic Subgroup</th>
<th>Drosophila OBP Closest Homologues</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Phylogenetic E-value</td>
<td>Phylogenetic E-value</td>
<td></td>
</tr>
<tr>
<td>AaegOBP40</td>
<td>AAEL009597</td>
<td>matype1</td>
<td>AAEL005777</td>
<td>mclassic9a</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Obp99a</td>
<td>1e-10</td>
</tr>
<tr>
<td>AaegOBP41</td>
<td>AAEL009599</td>
<td>matype1</td>
<td>AAEL005772</td>
<td>mclassic9a</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Obp99a</td>
<td>1e-10</td>
</tr>
<tr>
<td>AGAP002190</td>
<td>AGAP000278</td>
<td>matype1</td>
<td>AGAP0002189</td>
<td>mclassic9b</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Obp99a</td>
<td>1e-10</td>
</tr>
<tr>
<td>AGAP002191</td>
<td>AGAP002188</td>
<td>matype1</td>
<td>AAagOBP9</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AGAP011647</td>
<td>AGAP010409</td>
<td>matype1</td>
<td>AGAP010787</td>
<td>mclassic8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CquiOBP85</td>
<td>2e-11</td>
</tr>
<tr>
<td>CPIJ015732</td>
<td>CPUJ010787</td>
<td>matype1</td>
<td>CPUJ010782</td>
<td>mclassic9b</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CquiOBP51</td>
<td>2e-17</td>
</tr>
<tr>
<td>AGAP000278</td>
<td>AGAP002189</td>
<td>matype1</td>
<td>AGAP002189</td>
<td>mclassic9b</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Obp99a</td>
<td>1e-10</td>
</tr>
<tr>
<td>AGAP000318</td>
<td>AAEL007003</td>
<td>matype2</td>
<td>AAEL007003</td>
<td>mclassic9a</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Obp99a</td>
<td>1e-07</td>
</tr>
<tr>
<td>AGAP000319</td>
<td>AAEL002617</td>
<td>matype2</td>
<td>AAEL002617</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Obp99a</td>
<td></td>
</tr>
<tr>
<td>AGAP000344</td>
<td>AAEL007003</td>
<td>matype2</td>
<td>AAEL007003</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>No group</td>
<td>5e-07</td>
</tr>
<tr>
<td>AGAP000350</td>
<td>AAEL001730</td>
<td>matype2</td>
<td>AAEL001730</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Obp99a</td>
<td></td>
</tr>
<tr>
<td>AGAP000377</td>
<td>AAEL004343</td>
<td>matype2</td>
<td>AAEL004343</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Obp99a</td>
<td>4e-06</td>
</tr>
<tr>
<td>AGAP001153</td>
<td>AAEL007003</td>
<td>matype2</td>
<td>AAEL007003</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>No group</td>
<td>1e-04</td>
</tr>
<tr>
<td>AGAP001174</td>
<td>AAEL004343</td>
<td>matype2</td>
<td>AAEL004343</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>mclassic9a</td>
<td>3e-05</td>
</tr>
<tr>
<td>AGAP001179</td>
<td>AAEL004343</td>
<td>matype2</td>
<td>AAEL004343</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>mclassic9a</td>
<td>3e-05</td>
</tr>
<tr>
<td>AGAP001189</td>
<td>AAEL004343</td>
<td>matype2</td>
<td>AAEL004343</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>mclassic9a</td>
<td>3e-05</td>
</tr>
<tr>
<td>AGAP004516</td>
<td>AAEL004343</td>
<td>matype2</td>
<td>AAEL004343</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>mclassic9a</td>
<td>3e-05</td>
</tr>
<tr>
<td>AGAP009433</td>
<td>AAEL004343</td>
<td>matype2</td>
<td>AAEL004343</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>mclassic9a</td>
<td>3e-05</td>
</tr>
<tr>
<td>AGAP013719</td>
<td>AAEL004343</td>
<td>matype2</td>
<td>AAEL004343</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>mclassic9a</td>
<td>3e-05</td>
</tr>
<tr>
<td>AGAP013720</td>
<td>AAEL004343</td>
<td>matype2</td>
<td>AAEL004343</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>mclassic9a</td>
<td>3e-05</td>
</tr>
<tr>
<td>AGAP014874</td>
<td>AAEL004343</td>
<td>matype2</td>
<td>AAEL004343</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>mclassic9a</td>
<td>3e-05</td>
</tr>
<tr>
<td>AGAP014876</td>
<td>AAEL00111730</td>
<td>matype2</td>
<td>AAEL00111730</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>AAagOBP9</td>
<td></td>
</tr>
</tbody>
</table>

(continued)
Table 2 Continued

<table>
<thead>
<tr>
<th>ID</th>
<th>Mosquito Atypical OBP</th>
<th>Phylogenetic Subgroup</th>
<th>N-term</th>
<th>Phylogenetic Subgroup</th>
<th>E-value</th>
<th>C-term</th>
<th>Phylogenetic Subgroup</th>
<th>E-value</th>
<th>N-term</th>
<th>E-value</th>
<th>C-term</th>
<th>E-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGAP000641/44</td>
<td>mtype2</td>
<td>AGAP013182</td>
<td>ND</td>
<td>AgamOBP59</td>
<td>3e-09</td>
<td>AGAP002025</td>
<td>mclassic9b</td>
<td>4e-10</td>
<td>Pbp rp2</td>
<td>1e-04</td>
<td>Pbp rp1</td>
<td>5e-04</td>
</tr>
<tr>
<td>AGAP000642</td>
<td>mtype2</td>
<td>AGAP013182</td>
<td>ND</td>
<td>AgamOBP59</td>
<td>2e-08</td>
<td>AGAP002025</td>
<td>mclassic9b</td>
<td>2e-06</td>
<td>Obp56d</td>
<td>1e-05</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>AGAP000643</td>
<td>mtype2</td>
<td>AGAP013182</td>
<td>ND</td>
<td>AgamOBP59</td>
<td>8e-09</td>
<td>AGAP002025</td>
<td>mclassic9b</td>
<td>2e-06</td>
<td>Obp56d</td>
<td>1e-05</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>CPIJ001690</td>
<td>matype2</td>
<td>CPU009937</td>
<td>mclassic8</td>
<td>2e-07</td>
<td></td>
<td>—</td>
<td>—</td>
<td>Obp99a</td>
<td>4e-07</td>
<td>—</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>CquiOBP92</td>
<td>matype2</td>
<td>CPU009937</td>
<td>mclassic8</td>
<td>2e-05</td>
<td>CPU010782</td>
<td>mclassic9b</td>
<td>4e-02</td>
<td>Obp44a</td>
<td>4e-07</td>
<td>—</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>CPIJ003865</td>
<td>matype2</td>
<td>CPU009937</td>
<td>mclassic8</td>
<td>4e-06</td>
<td>CPU010782</td>
<td>mclassic9b</td>
<td>2e-02</td>
<td>Obp44a</td>
<td>7e-07</td>
<td>—</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>CquiOBP88</td>
<td>matype2</td>
<td>CPU009937</td>
<td>mclassic8</td>
<td>3e-07</td>
<td>CPU010782</td>
<td>mclassic9b</td>
<td>1e-04</td>
<td>Obp44a</td>
<td>1e-09</td>
<td>Obp99b</td>
<td>1e-03</td>
<td></td>
</tr>
<tr>
<td>CPIJ003866</td>
<td>matype2</td>
<td>CPU009937</td>
<td>mclassic8</td>
<td>1e-09</td>
<td>CPU010782</td>
<td>mclassic9b</td>
<td>4e-02</td>
<td>ObP99a</td>
<td>9e-07</td>
<td>—</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>CquiOBP90</td>
<td>matype2</td>
<td>CPU009937</td>
<td>mclassic8</td>
<td>2e-06</td>
<td>CPU017326</td>
<td>ObP99a</td>
<td>6e-02</td>
<td>Obp44a</td>
<td>4e-04</td>
<td>Obp56g</td>
<td>4e-03</td>
<td></td>
</tr>
<tr>
<td>CPIJ017163</td>
<td>matype2</td>
<td>CPU009937</td>
<td>mclassic8</td>
<td>6e-06</td>
<td>CPU010789</td>
<td>ObP953</td>
<td>2e-02</td>
<td>Obp44a</td>
<td>2e-04</td>
<td>—</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>CquiOBP99</td>
<td>matype2</td>
<td>CPU009937</td>
<td>mclassic8</td>
<td>7e-10</td>
<td>CPU010782</td>
<td>mclassic9b</td>
<td>1e-02</td>
<td>Obp99c</td>
<td>2e-04</td>
<td>—</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>CPIJ017166</td>
<td>matype2</td>
<td>CPU009937</td>
<td>mclassic8</td>
<td>1e-03</td>
<td>CPU016343</td>
<td>mclassic9b</td>
<td>3e-02</td>
<td>Obp44a</td>
<td>1e-04</td>
<td>—</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>CquiOBP95</td>
<td>matype2</td>
<td>CPU009937</td>
<td>mclassic8</td>
<td>1e-03</td>
<td>CPU001365</td>
<td>ObP987</td>
<td>Pbp rp1</td>
<td>1e-02</td>
<td>Obp44a</td>
<td>4e-04</td>
<td>Obp56d</td>
<td>7e-05</td>
</tr>
<tr>
<td>CPIJ017167</td>
<td>matype2</td>
<td>CPU016951</td>
<td>Obp19a</td>
<td>AgamOBP57</td>
<td>6e-04</td>
<td>CPU016343</td>
<td>mclassic9b</td>
<td>3e-02</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>CquiOBP96</td>
<td>matype2</td>
<td>CPU009937</td>
<td>mclassic8</td>
<td>8e-03</td>
<td>CPU010789</td>
<td>mclassic7</td>
<td>2e-02</td>
<td>Obp44a</td>
<td>3e-03</td>
<td>Obp99b</td>
<td>2e-03</td>
<td></td>
</tr>
<tr>
<td>CPIJ017169</td>
<td>matype2</td>
<td>CPU009937</td>
<td>mclassic8</td>
<td>3e-02</td>
<td>CPU016343</td>
<td>mclassic9b</td>
<td>6e-03</td>
<td>Obp44a</td>
<td>6e-03</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>CquiOBP100</td>
<td>matype2</td>
<td>CPU009937</td>
<td>mclassic8</td>
<td>8e-04</td>
<td>CPU016343</td>
<td>mclassic9a</td>
<td>2e-06</td>
<td>Obp56d</td>
<td>8e-04</td>
<td>Obp99a</td>
<td>9e-08</td>
<td></td>
</tr>
<tr>
<td>CPIJ017170</td>
<td>matype2</td>
<td>CPU009937</td>
<td>mclassic3a</td>
<td>1e-05</td>
<td>AAEL004343</td>
<td>mclassic9a</td>
<td>3e-06</td>
<td>Obp56d</td>
<td>3e-05</td>
<td>Obp99a</td>
<td>5e-07</td>
<td></td>
</tr>
<tr>
<td>CquiOBP94</td>
<td>matype2</td>
<td>AAEL002596</td>
<td>AAegOBP9</td>
<td>mclassic3a</td>
<td>8e-04</td>
<td>AAEL004343</td>
<td>mclassic9</td>
<td>3e-06</td>
<td>Obp56d</td>
<td>3e-05</td>
<td>Obp99a</td>
<td>5e-07</td>
</tr>
<tr>
<td>AAEEL006385</td>
<td>matype3</td>
<td>AAEL002617</td>
<td>mclassic3a</td>
<td>1e-05</td>
<td>AAEL004343</td>
<td>mclassic9a</td>
<td>3e-06</td>
<td>Obp56d</td>
<td>3e-05</td>
<td>Obp99a</td>
<td>5e-07</td>
<td></td>
</tr>
<tr>
<td>AaeOBP33</td>
<td>matype3</td>
<td>AAEL002617</td>
<td>mclassic3a</td>
<td>1e-05</td>
<td>AAEL004343</td>
<td>mclassic9a</td>
<td>3e-06</td>
<td>Obp56d</td>
<td>3e-05</td>
<td>Obp99a</td>
<td>5e-07</td>
<td></td>
</tr>
<tr>
<td>AAEEL006393</td>
<td>matype3</td>
<td>AAEL002617</td>
<td>mclassic3a</td>
<td>1e-05</td>
<td>AAEL004343</td>
<td>mclassic9a</td>
<td>3e-06</td>
<td>Obp56d</td>
<td>3e-05</td>
<td>Obp99a</td>
<td>5e-07</td>
<td></td>
</tr>
<tr>
<td>AaeOBP28</td>
<td>matype3</td>
<td>AAEL002617</td>
<td>mclassic3a</td>
<td>1e-05</td>
<td>AAEL004343</td>
<td>mclassic9a</td>
<td>3e-06</td>
<td>Obp56d</td>
<td>3e-05</td>
<td>Obp99a</td>
<td>5e-07</td>
<td></td>
</tr>
<tr>
<td>AAEEL006396</td>
<td>matype3</td>
<td>AAEL002617</td>
<td>mclassic3a</td>
<td>1e-05</td>
<td>AAEL004342</td>
<td>mclassic9a</td>
<td>2e-03</td>
<td>Obp56d</td>
<td>1e-05</td>
<td>Obp99a</td>
<td>4e-05</td>
<td></td>
</tr>
<tr>
<td>AaeOBP31</td>
<td>matype3</td>
<td>AAEL002617</td>
<td>mclassic3a</td>
<td>1e-05</td>
<td>AAEL004342</td>
<td>mclassic9a</td>
<td>2e-03</td>
<td>Obp56d</td>
<td>1e-05</td>
<td>Obp99a</td>
<td>4e-05</td>
<td></td>
</tr>
<tr>
<td>AAEEL006398</td>
<td>matype3</td>
<td>AAEL001730</td>
<td>mclassic8</td>
<td>2e-02</td>
<td>AAEL011730</td>
<td>mclassic9</td>
<td>1e-06</td>
<td>Obp56d</td>
<td>9e-03</td>
<td>Obp99a</td>
<td>1e-06</td>
<td></td>
</tr>
<tr>
<td>AaeOBP32</td>
<td>matype3</td>
<td>AAEL001730</td>
<td>mclassic8</td>
<td>1e-06</td>
<td>AAEL001730</td>
<td>mclassic9</td>
<td>1e-08</td>
<td>Obp99a</td>
<td>2e-05</td>
<td>Obp99b</td>
<td>5e-04</td>
<td></td>
</tr>
<tr>
<td>AGAP010648</td>
<td>matype3</td>
<td>AGAP002025</td>
<td>mclassic9b</td>
<td>1e-10</td>
<td>AGAP002025</td>
<td>mclassic9b</td>
<td>1e-08</td>
<td>Obp99a</td>
<td>2e-05</td>
<td>Obp99b</td>
<td>5e-04</td>
<td></td>
</tr>
<tr>
<td>AgamOBP44</td>
<td>matype3</td>
<td>AGAP002025</td>
<td>mclassic9b</td>
<td>1e-12</td>
<td>AGAP002025</td>
<td>mclassic9b</td>
<td>1e-12</td>
<td>Obp99a</td>
<td>8e-07</td>
<td>Obp99b</td>
<td>6e-08</td>
<td></td>
</tr>
<tr>
<td>AGAP010649</td>
<td>matype3</td>
<td>AGAP013182</td>
<td>ND</td>
<td>AgamOBP59</td>
<td>5e-09</td>
<td>AGAP002025</td>
<td>mclassic9b</td>
<td>2e-06</td>
<td>Obp56d</td>
<td>1e-05</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>AgamOBP31</td>
<td>matype3</td>
<td>AGAP013182</td>
<td>ND</td>
<td>AgamOBP59</td>
<td>1e-11</td>
<td>AGAP002189</td>
<td>mclassic9b</td>
<td>3e-11</td>
<td>Obp99b</td>
<td>9e-05</td>
<td>Obp99a</td>
<td>3e-10</td>
</tr>
</tbody>
</table>

(continued)
<table>
<thead>
<tr>
<th>ID</th>
<th>Phylogenetic Subgroup</th>
<th>Phylogenetic Subgroup</th>
<th>Phylogenetic Subgroup</th>
<th>Phylogenetic Subgroup</th>
<th>Phylogenetic Subgroup</th>
<th>Phylogenetic Subgroup</th>
<th>Phylogenetic Subgroup</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPIJ009038</td>
<td>matype3</td>
<td>mclass8</td>
<td>9e-08</td>
<td>mclass8</td>
<td>7e-08</td>
<td>Obp56d</td>
<td>2e-03</td>
</tr>
<tr>
<td>CquiOBP87</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CPIJ017342</td>
<td>matype3</td>
<td>mclass8</td>
<td>2e-08</td>
<td>mclass8</td>
<td>6e-07</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>CquiOBP93</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AAEL000796</td>
<td>matype4</td>
<td>mclass8</td>
<td>8e-04</td>
<td>mclass9a</td>
<td>6e-07</td>
<td>Obp56i</td>
<td>—</td>
</tr>
<tr>
<td>AaegOBP96</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AAEL000821</td>
<td>matype4</td>
<td>mclass8</td>
<td>1e-05</td>
<td>mclass9a</td>
<td>3e-05</td>
<td>Obp56g</td>
<td>1e-04</td>
</tr>
<tr>
<td>AaegOBp6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AAEL000827</td>
<td>matype4</td>
<td>No group</td>
<td>4e-03</td>
<td>mclass9a</td>
<td>6e-07</td>
<td>Obp56i</td>
<td>—</td>
</tr>
<tr>
<td>AaegOBP84</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AAEL000831</td>
<td>matype4</td>
<td>mclass8</td>
<td>1e-02</td>
<td>mclass9a</td>
<td>3e-05</td>
<td>Obp56g</td>
<td>1e-04</td>
</tr>
<tr>
<td>AaegOBP85</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AAEL000833</td>
<td>matype4</td>
<td>mclass7</td>
<td>4e-03</td>
<td>mclass9a</td>
<td>8e-05</td>
<td>Obp56i</td>
<td>—</td>
</tr>
<tr>
<td>AaegOBP7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AAEL000835</td>
<td>matype4</td>
<td>mclass8</td>
<td>8e-03</td>
<td>mclass9a</td>
<td>6e-07</td>
<td>Obp56i</td>
<td>—</td>
</tr>
<tr>
<td>AaegOBP97</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AAEL000837</td>
<td>matype4</td>
<td>—</td>
<td>4e-03</td>
<td>—</td>
<td>—</td>
<td>Obp51a</td>
<td>1e-02</td>
</tr>
<tr>
<td>AaegOBP112</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AAEL001487</td>
<td>matype4</td>
<td>mclass8</td>
<td>2e-04</td>
<td>mclass9a</td>
<td>1e-06</td>
<td>Obp56i</td>
<td>—</td>
</tr>
<tr>
<td>AaegOBP114</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AAEL003311</td>
<td>matype4</td>
<td>mclass8</td>
<td>7e-05</td>
<td>mclass9a</td>
<td>2e-04</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>AaegOBP111</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AAEL003315</td>
<td>matype4</td>
<td>mclass8</td>
<td>2e-08</td>
<td>mclass9a</td>
<td>2e-04</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>AaegOBP16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AAEL003511</td>
<td>matype4</td>
<td>mclass8</td>
<td>2e-04</td>
<td>mclass9a</td>
<td>1e-06</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>AaegOBP87</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AAEL003513</td>
<td>matype4</td>
<td>mclass8</td>
<td>5e-07</td>
<td>mclass9a</td>
<td>1e-07</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>AaegOBP100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AAEL003525</td>
<td>matype4</td>
<td>mclass8</td>
<td>2e-03</td>
<td>mclass9a</td>
<td>4e-07</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>AaegOBP101</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AAEL003538</td>
<td>matype4</td>
<td>mclass8</td>
<td>5e-07</td>
<td>mclass9a</td>
<td>1e-07</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>AaegOBP102</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AAEL004856</td>
<td>matype4</td>
<td>mclass8</td>
<td>3e-05</td>
<td>mclass9a</td>
<td>6e-06</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>AaegOBP86</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AAEL010714</td>
<td>matype4</td>
<td>mclass8</td>
<td>3e-05</td>
<td>mclass9a</td>
<td>2e-07</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>AaegOBP45</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AAEL010718</td>
<td>matype4</td>
<td>mclass8</td>
<td>5e-07</td>
<td>mclass9a</td>
<td>2e-06</td>
<td>Obp56d</td>
<td>6e-03</td>
</tr>
<tr>
<td>AaegOBP44</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AAEL010872</td>
<td>matype4</td>
<td>mclass8</td>
<td>9e-05</td>
<td>mclass9a</td>
<td>5e-05</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>AaegOBP46</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AAEL010874</td>
<td>matype4</td>
<td>mclass8</td>
<td>3e-06</td>
<td>mclass9a</td>
<td>3e-06</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>AaegOBP88</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AAEL010875</td>
<td>matype4</td>
<td>mclass8</td>
<td>3e-06</td>
<td>mclass9a</td>
<td>2e-05</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>AaegOBP103</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CPIJ000653</td>
<td>matype4</td>
<td>—</td>
<td>—</td>
<td>mclass9b</td>
<td>2e-06</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>CquiOBP84</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CPIJ008154</td>
<td>matype4</td>
<td>—</td>
<td>—</td>
<td>mclass9b</td>
<td>7e-05</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

(continued)
OBPs which were previously observed in Drosophila as Dimer OBPs and that they no more stand specific to the mosquito genomes as reported earlier (Xu et al. 2003). Furthermore, the matype2 members which carry a presence of only 6 cysteines in the place of 12 cysteines as in the other two-domain OBPs is suggestive of a possible adaptation in the fold with 3 disulphide bonds in place of 6 disulphide bonds in the other types. The astound distribution of these matype2 OBP genes from A. gambiae on the X chromosome further increases the speculative importance of these proteins in the blood feeding mechanism by female mosquitoes. Interestingly, most of the members of the two-domain OBP subfamily are reported as differentially expressed with respect to blood time series which adds to the importance of these proteins in host recognition (Dissanayake et al. 2010).

Ecological adaptations might have driven the need for the observed expansion in two-domain OBP gene repertoire in the Aed. aegyptii and C. quinquefasciatus genome when compared with A. gambiae. Our observations indicate that this expansion most probably occurred through gene duplication events in localized genome regions which lead to the observed gene clusters. We hence hypothesize that two distinct mechanisms could underlie the emergence of Atypical genes in mosquitoes. The observations made in A. gambiae genome sustain the first hypothesis that two-domain OBPs might have originated from gene duplicates of mclassic9, mclassic8,

### Table 2

<table>
<thead>
<tr>
<th>ID</th>
<th>Mosquito Atypical OBP</th>
<th>N-term</th>
<th>Phylogenetic Subgroup</th>
<th>E-value</th>
<th>C-term</th>
<th>Phylogenetic Subgroup</th>
<th>E-value</th>
<th>Mosquito Classic OBP</th>
<th>Closest Homologues</th>
<th>Droso phila OBP</th>
<th>Closest Homologues</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPIJ008155</td>
<td>matype4</td>
<td>CPU010789</td>
<td>mclassic7</td>
<td>2e-03</td>
<td>CPU016343</td>
<td>mclassic9b</td>
<td>1e-06</td>
<td></td>
<td></td>
<td>Obp99a</td>
<td>2e-04</td>
</tr>
<tr>
<td>CquiOBP79</td>
<td></td>
<td>CquiOBP53</td>
<td></td>
<td></td>
<td>CquiOBP63</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CPIJ008156</td>
<td>matype4</td>
<td>CPU010789</td>
<td>mclassic7</td>
<td>4e-02</td>
<td>CPU010782</td>
<td>maclassic9b</td>
<td>2e-11</td>
<td></td>
<td></td>
<td>Obp99a</td>
<td>2e-08</td>
</tr>
<tr>
<td>CquiOBP80</td>
<td></td>
<td>CquiOBP53</td>
<td></td>
<td></td>
<td>CquiOBP46</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CPIJ008157</td>
<td>matype4</td>
<td>CPU010787</td>
<td>mclassic9a</td>
<td>1e-02</td>
<td>CPU010782</td>
<td>maclassic9b</td>
<td>1e-09</td>
<td></td>
<td></td>
<td>Obp99a</td>
<td>6e-08</td>
</tr>
<tr>
<td>CquiOBP76</td>
<td></td>
<td>CquiOBP51</td>
<td></td>
<td></td>
<td>CquiOBP46</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CPIJ008158</td>
<td>matype4</td>
<td>CPU016343</td>
<td>mclassic9b</td>
<td>1e-08</td>
<td>CPU016343</td>
<td>mclassic9b</td>
<td>8e-07</td>
<td>Obp99a</td>
<td>1e-05</td>
<td>Obp99a</td>
<td>3e-08</td>
</tr>
<tr>
<td>CquiOBP77</td>
<td></td>
<td>CquiOBP63</td>
<td></td>
<td></td>
<td>CquiOBP63</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CPIJ008159</td>
<td>matype4</td>
<td>CPU009937</td>
<td>mclassic8</td>
<td>6e-03</td>
<td>CPU016343</td>
<td>mclassic9b</td>
<td>4e-10</td>
<td></td>
<td></td>
<td>Obp99a</td>
<td>6e-09</td>
</tr>
<tr>
<td>CquiOBP78</td>
<td></td>
<td>CquiOBP44</td>
<td></td>
<td></td>
<td>CquiOBP63</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CPIJ008160</td>
<td>matype4</td>
<td>CPU009937</td>
<td>mclassic8</td>
<td>1e-02</td>
<td>CPU016343</td>
<td>mclassic9b</td>
<td>1e-09</td>
<td></td>
<td></td>
<td>Obp44a</td>
<td>3e-03</td>
</tr>
<tr>
<td>CquiOBP81</td>
<td></td>
<td>CquiOBP44</td>
<td></td>
<td></td>
<td>CquiOBP63</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CPIJ008161</td>
<td>matype4</td>
<td>CPU010789</td>
<td>mclassic7</td>
<td>2e-02</td>
<td>CPU016343</td>
<td>mclassic9b</td>
<td>2e-09</td>
<td></td>
<td></td>
<td>Obp99b</td>
<td>3e-06</td>
</tr>
<tr>
<td>CquiOBP82</td>
<td></td>
<td>CquiOBP53</td>
<td></td>
<td></td>
<td>CquiOBP63</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AAEI008640</td>
<td>—</td>
<td>AAEI011730</td>
<td>mclassic8</td>
<td>2e-08</td>
<td>AAEI011730</td>
<td>mclassic8</td>
<td>1e-05</td>
<td></td>
<td></td>
<td>Obp99c</td>
<td>3e-07</td>
</tr>
<tr>
<td>AaegOBP113</td>
<td></td>
<td>AaegOBP81</td>
<td></td>
<td></td>
<td>AaegOBP81</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AAEI014430</td>
<td>—</td>
<td>AAEI007003</td>
<td>No group</td>
<td>2e-08</td>
<td>AAEI011730</td>
<td>mclassic8</td>
<td>2e-05</td>
<td>Obp99b</td>
<td>1e-07</td>
<td>Obp44a</td>
<td>1e-06</td>
</tr>
<tr>
<td>AaeOBP58</td>
<td></td>
<td>AaeOBP80</td>
<td></td>
<td></td>
<td>AAEI007003</td>
<td>mclassic8</td>
<td>9e-09</td>
<td>Obp99b</td>
<td>1e-07</td>
<td>Obp99b</td>
<td>3e-03</td>
</tr>
<tr>
<td>AAEI014431</td>
<td>—</td>
<td>AAEI011730</td>
<td>mclassic8</td>
<td>4e-12</td>
<td>AAEI004342</td>
<td>mclassic9a</td>
<td>9e-09</td>
<td>Obp99b</td>
<td>1e-07</td>
<td>Obp99b</td>
<td>3e-03</td>
</tr>
<tr>
<td>AaeOBP1110</td>
<td></td>
<td>AaeOBP81</td>
<td></td>
<td></td>
<td>AAEI011730</td>
<td>mclassic8</td>
<td>9e-09</td>
<td>Obp99b</td>
<td>1e-07</td>
<td>Obp99b</td>
<td>3e-03</td>
</tr>
<tr>
<td>AGAP000580</td>
<td>—</td>
<td>AGAP002189</td>
<td>mclassic9b</td>
<td>2e-06</td>
<td>AGAP002025</td>
<td>mclassic9b</td>
<td>4e-05</td>
<td>Obp99b</td>
<td>6e-05</td>
<td>Obp99c</td>
<td>2e-06</td>
</tr>
<tr>
<td>AgamOBP38</td>
<td></td>
<td>AgamOBP14</td>
<td></td>
<td></td>
<td>AGAP002025</td>
<td>mclassic9b</td>
<td>8e-07</td>
<td>Obp99a</td>
<td>1e-06</td>
<td>Obp99c</td>
<td>1e-03</td>
</tr>
<tr>
<td>AGAP000638</td>
<td>—</td>
<td>AGAP010409</td>
<td>mclassic8</td>
<td>2e-11</td>
<td>AGAP002025</td>
<td>mclassic9b</td>
<td>8e-07</td>
<td>Obp99a</td>
<td>1e-06</td>
<td>Obp99c</td>
<td>1e-03</td>
</tr>
<tr>
<td>AgamOBP32</td>
<td></td>
<td>AgamOBP22</td>
<td></td>
<td></td>
<td>AGAP002025</td>
<td>mclassic9b</td>
<td>2e-07</td>
<td>Obp99a</td>
<td>1e-06</td>
<td>Obp99c</td>
<td>1e-03</td>
</tr>
<tr>
<td>AGAP000640</td>
<td>—</td>
<td>AGAP010409</td>
<td>mclassic8</td>
<td>2e-11</td>
<td>AGAP002025</td>
<td>mclassic9b</td>
<td>7e-11</td>
<td></td>
<td></td>
<td>Obp44a</td>
<td>2e-04</td>
</tr>
<tr>
<td>AgamOBP33</td>
<td></td>
<td>AgamOBP22</td>
<td></td>
<td></td>
<td>AGAP002025</td>
<td>mclassic9b</td>
<td>2e-08</td>
<td>Obp99a</td>
<td>5e-05</td>
<td>Obp99c</td>
<td>7e-05</td>
</tr>
<tr>
<td>AGAP0005182</td>
<td>—</td>
<td>AGAP0031382</td>
<td>ND</td>
<td>8e-07</td>
<td>AGAP002025</td>
<td>mclassic9b</td>
<td>9e-09</td>
<td></td>
<td></td>
<td>Obp44a</td>
<td>2e-04</td>
</tr>
<tr>
<td>AgamOBP41</td>
<td></td>
<td>AgamOBP59</td>
<td></td>
<td></td>
<td>AGAP002025</td>
<td>mclassic9b</td>
<td>2e-08</td>
<td>Obp99a</td>
<td>5e-05</td>
<td>Obp99c</td>
<td>7e-05</td>
</tr>
<tr>
<td>AGAP009065</td>
<td>—</td>
<td>AGAP0031382</td>
<td>ND</td>
<td>6e-10</td>
<td>AGAP002025</td>
<td>mclassic9b</td>
<td>1e-09</td>
<td></td>
<td></td>
<td>Obp44a</td>
<td>2e-04</td>
</tr>
<tr>
<td>AgamOBP42</td>
<td></td>
<td>AgamOBP59</td>
<td></td>
<td></td>
<td>AGAP002025</td>
<td>mclassic9b</td>
<td>1e-09</td>
<td></td>
<td></td>
<td>Obp44a</td>
<td>2e-04</td>
</tr>
<tr>
<td>AGAP009402</td>
<td>—</td>
<td>AGAP010409</td>
<td>mclassic8</td>
<td>5e-11</td>
<td>AGAP002189</td>
<td>Obp44a</td>
<td>6e-16</td>
<td>Obp99a</td>
<td>4e-06</td>
<td>Obp99a</td>
<td>2e-09</td>
</tr>
<tr>
<td>AgamOBP43</td>
<td></td>
<td>AgamOBP22</td>
<td></td>
<td></td>
<td>AGAP002189</td>
<td>mclassic9b</td>
<td>9e-09</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note.—The table shows top hits results of the BLAST search among all mosquito Classic OBPs and Drosophila OBPs after splitting the Atypical proteins into their two respective putative domains.
or Obp99a related members and their subsequent gene fusion leading to Atypical genes co-clustered with their Classic counterparts. The observations made in Aed. aegyptii and C. quinquefasciatus support the second complementary hypothesis whereby the Atypical genes have undergone further gene duplications probably in response to ecological constraints in these mosquito lineages.

Our analysis hence sustains the proposition that the Atypical OBP genes to be renamed two-domain OBP proteins. Their future structural characterization and ligand binding profiling would be of significant importance in deciphering their contribution in olfaction in mosquitoes.

**Materials and Methods**

**Sequence Searches**

The database of the predicted protein sequences of the three mosquito genomes A. gambiae (A. gambiae annotation, AgamP3.4), Aed. aegypti (Aed. aegypti annotation, AaegL1.1), and C. quinquefasciatus (C. quinquefasciatus annotation, CpipJ1.2) were downloaded from the VectorBase (Lawson et al. 2009) version 3.4 (http://www.vectorbase.org, last accessed January 9, 2013) and Ensembl Genomes (Hubbard et al. 2009). The putative OBPs in the three mosquito species were identified using 10 Drosophila query sequences which belong to three different subfamilies Classic/General OBPs, PlusC, and MinusC OBPs using a PSI-BLAST (Altschul et al. 1997) run of 10 query sequences with an E-value cutoff of 3e−10 (Vieira et al. 2007) and an alignment length cutoff of 75% with respect to the query sequence. At this level, all of the previously identified members in the three genomes were identified with identification of a few additional members. A second run of PSI-BLAST was initiated with the hits from the previous runs. Using this protocol it was possible to not only pick up all the members of OBPs reported so far (Vogt 2002; Xu et al. 2003; Zhou et al. 2004, 2008; Pelletier and Leal 2009, 2011; Vieira and Rozas 2011) but also a remarkable number of additional members. The additional sequences were checked for the presence of a signal peptide using the SignalP server (Petersen et al. 2011), the PBP/GOBP domain using CD-Search (Marchler-Bauer and Bryant 2004) in the case of classic OBPs, and alignment of the new sequences with their subfamily members in case of Atypical and PlusC proteins. The D7 proteins which were identified using this method but which are considered as a distinct family of proteins related to the OBPs were also retained for further analysis and used as an outgroup in the construction of phylogenetic trees. The orthologous sequences were identified based on the reciprocal best hit approach using BLAST (Moreno-Hagelsieb and Latimer 2008). The newly added sequences were named according to the naming conventions used in the earlier reports (Vogt 2002; Xu et al. 2003; Zhou et al. 2004, 2008; Pelletier and Leal 2009).

**Multiple Sequence Alignment**

The multiple sequence alignment forms the basis for any analysis of a family of proteins and it is highly necessary to obtain an accurate alignment. The error rate in the alignment increases with the increase in divergence of the proteins. Structure-based alignments in turn are considered to be the most accurate forms of alignments and hence, in this study, the structure alignment was used in constructing the alignments. The structure alignment was constructed using 10 OBPs in the OBP gene family using COMPARER (Sali and Blundell 1990). However, the use of the structure alignment as profiles was restricted to seven members in the case of OBPs and two members for the D7 family due to the limited number of structural data (data not shown). The OBPs and the D7 sequences were aligned to their respective structure alignments as profiles, and a combined alignment of the two family of proteins was constructed using the profile-profile alignment option using ClustalX (Thompson et al. 1994, 1997; Jeanmougin et al. 1998). The alignments were truncated based on the structure alignment on the N-terminal end which corresponds to the signal peptide region that has a high substitution rate; however, the C-terminal ends were retained due to the presence of an extended C-terminal in the case of Atypical subfamily members of the OBP family. This method was applied for aligning the sequences in all the three different genomes. Alignments for the different subclasses were constructed with sequences from all the three mosquito genomes and in the case of Classic subfamily, along with Drosophila sequences. The alignment of the Atypical and PlusC subclasses of OBPs were however not based on the structure alignment.

**Phylogenetic Analysis**

The phylogenetic trees were inferred using the Neighbor-Joining method (Saitou and Nei 1987) in MEGA 4.0 (Tamura et al. 2007). The percentage of replicate trees in which the associated sequences cluster together in the bootstrap test (1000 replicates) are shown next to the branches of the bootstrap consensus trees (Felsenstein 1985) and branches with <50% bootstrap cutoff were collapsed. The evolutionary distances were computed using the Poisson correction method (Zuckerkandl and Pauling 1965) and are in the units of number of amino acid substitutions per site. All positions containing alignment gaps and missing data were eliminated only in pairwise sequence comparisons (pairwise deletion option).

The trees were rooted at the branches of the D7 family of proteins which was considered as an outgroup (supplementary fig. S2a–c, Supplementary Material online).

The trees of the different subclasses (figs. 4–6) used for the comparative analysis of the different genomes were analyzed as unrooted trees. The phylogenetic trees were inferred using the Neighbor-Joining method (Saitou and Nei 1987) in MEGA 4.0 (Tamura et al. 2007). The percentage of replicate trees in
which the associated sequences cluster together in the bootstrap test (1000 replicates) are shown next to the branches of the bootstrap consensus trees (Felsenstein 1985) and branches with <50% bootstrap cutoff were collapsed for the PlusC and Atypical OBP trees. The branches were not collapsed for the Classic OBP tree, however the subtype definition was still based on 50% bootstrap cutoff.

Orthology, Paralogy, Chromosomal Mapping, and Tentative Syntenic Analysis

OBP orthologs have been identified using the reciprocal BLAST hit approach (Moreno-Hagelsieb and Latimer 2008) which is widely used in the detection of orthologs. The inParanoid database (O’Brien et al. 2005) was used to examine the inparalogous relationship between OBPs. Assembled genome data was only available for A. gambiae at the date of this work in the Ensembl Genome (Hubbard et al. 2009) and VectorBase (Lawson et al. 2009). The chromosomal locations of OBPs from A. gambiae were identified using this data. The genome data of Aed. aegypti and C. quinquefasciatus as featured todate in Ensembl Genomes and VectorBase are not yet assembled and were used to map the OBP genes in these genomes at the supercontigs level. The exact chromosomal locations are known for only about 10% of their supercontigs among which very few harbor OBP genes. Orthologous OBP genes identified as described above were used to establish putative synteny between chromosomal segments from A. gambiae and supercontigs from the other two Culicinæa species. The genes were mapped to their respective location on the chromosome or supercontigs (supplementary fig. S1a–e, Supplementary Material online). The chromosomes of A. gambiae was used as reference and were represented as a yellow bar and the contigs of Aedes and Culex are represented in purple and green, respectively. The direct three-way (1:1:1) orthology relationships among the three genomes are represented as green lines. The two-way (1:1) orthology relationships between two species are represented as black lines, and the inparalogous relationships are represented as red lines. The figures of the chromosomal mapping were drawn to scale using Adobe illustrator CS5.

Atypical Domain Analysis

The two constitutive PBP/GOBP OBP domains of Atypical OBPs were further characterized for their relationship with Classic or PlusC OBPs. For each Atypical OBP, the boundary between the N-term and C-term PGP/GOBP domains was manually delimited. This was performed by subjecting the full-length sequence to Pfam (Finn et al. 2010) and Conserved Domain Database (Marchler-Bauer et al. 2011) and was further validated by analyzing their cysteine profiles. Each N-term and C-term domain hence delimited was then subjected to a PSI-BLAST search (E-value cutoff value of $10^{-2}$) against a database that contains all OBPs from the same mosquito species in an attempt to find their putative distantly related single-domain OBPs. A similar search was performed against a database of Drosophila OBPs.

Supplementary Material

Supplementary figures S1–S3 and tables S1–S4 are available at Genome Biology and Evolution online (http://www.gbe.oxfordjournals.org/).

Acknowledgments

R.S. is thankful to University of La Réunion for supporting this collaborative research through continuous invited professorship support since 2005. The authors are thankful to Manipal University, NCBS (TIFR), and the University of La Réunion for bioinformatics infrastructural support. The authors thank Prof. N. Srinivasan (Indian Institute of Science), Prof. Matthew (NCBS), Prof. Frédéric Cadet (University of La Réunion), Prof. Vinh Tran (University of Nantes), and Prof. Sankaramarikrishnan (IIIT, Kanpur) for useful discussions regarding this work. The authors are also thankful to Olivier Cadet, Swapnil Mahajan, and Nicolas Fontaine for technical assistance toward setting up the mOBPdb database. This work was in part supported by a grant from Conseil Régional de La Réunion, the French Ministry of Research, and the European Union in the framework of the GRI Phase III project. M.M. was supported by an international PhD fellowship from Conseil Régional de La Réunion in the framework of the joint dual-studentship program between Manipal University and University of La Réunion.

Literature Cited


Associate editor: Yoshito Nimura