Thioester-containing proteins in the tsetse fly (*Glossina*) and their response to trypanosome infection

I. Matetovici and J. Van Den Abbeele

Unit of Veterinary Protozoology, Department of Biomedical Sciences, Institute of Tropical Medicine Antwerp (ITM), Antwerp, Belgium

Abstract

Thioester-containing proteins (TEPs) are conserved proteins with a role in innate immune immunity. In the current study, we characterized the TEP family in the genome of six tsetse fly species (*Glossina* spp.). Tsetse flies are the biological vectors of several African trypanosomes, which cause sleeping sickness in humans or nagana in livestock. The analysis of the tsetse TEP sequences revealed information about their structure, evolutionary relationships and expression profiles under both normal and trypanosome infection conditions. Phylogenetic analysis of the family showed that tsetse flies harbour a genomic expansion of specific TEPs that are not found in other dipterans. We found a general expression of all TEP genes in the alimentary tract, mouthparts and salivary glands. *Glossina morsitans* and *Glossina palpalis* TEP genes display a tissue-specific expression pattern with some that are markedly up-regulated when the fly is infected with the trypanosome parasite. A different TEP response was observed to infection with *Trypanosoma brucei* compared to *Trypanosoma congolense*, indicating that the tsetse TEP response is trypanosome-specific. These findings are suggestive for the involvement of the TEP family in tsetse innate immunity, with a possible role in the control of the trypanosome parasite.

Keywords: thioester-containing proteins, tsetse fly, trypanosome, innate immunity, vector–parasite interaction.

Introduction

Being exposed to a range of pathogenic organisms, insects rely solely on innate immune responses that serve as the first line of defence, protecting them through mechanisms that can be activated rapidly upon recognition of a foreign threat. An immune response is activated when pattern recognition receptors (PRRs) detect pathogen-associated molecules as ‘nonself’ and initiate a complex signal transduction cascade that finally leads to the production of specific effector molecules to control the microbial threat. PRRs include peptidoglycan-recognition proteins and Gram-negative-binding proteins that detect Gram-positive bacteria as well as fungi.

In addition to PRRs insects are equipped with secreted recognition molecules, opsonins, such as the thioester-containing proteins (TEPs; Blandin & Levashina, 2004). These are characterized by a highly reactive thioester motif that can bind covalently to pathogen target molecule. TEPs are a large protein family divided into four subfamilies: (1) vertebrate complement factors (C3–C5) and C3-like proteins, (2) alpha-2-macroglobulin (A2M), (3) insect TEP (iTEP) and (4) macroglobulin complement-related (MCR) proteins (Blandin & Levashina, 2004; Palmer & Jiggins, 2015). In the malaria vector *Anopheles gambiae*, TEP1 is involved in opsonization of various pathogens including *Plasmodium* spp. (Levashina *et al.*, 2001; Blandin *et al.*, 2008; Yassine *et al.*, 2012), and was demonstrated to play a key role in the removal of defective apoptotic sperm cells during spermatogenesis (Pompon & Levashina, 2015). *Aedes aegypti* macroglobulin complement-related factor (*Ae. aegypti_MCR*) was shown to possess antiviral activity against dengue virus infection by induction of antimicrobial peptides (Xiao *et al.*, 2014). In *Drosophila melanogaster*, TEP molecules display immunomodulatory responses to various pathogens, such as...
bacteria (Lagueux et al., 2000; Stroschein-Stevenson et al., 2006; Igboin et al., 2011; Dostalova et al., 2011; Shokal & Eleftherianos, 2017), fungi (Stroschein-Stevenson et al., 2006; Dostalova et al., 2017), nematodes (Arefin et al., 2014; Castillo et al., 2015) and parasitoids (Wertheim et al., 2005; Dostalova et al., 2017).

Besides the studies in mosquitoes and fruit flies there are only a handful of reports in other insects. Differential expression following microbial or parasitic infection has been reported for the house fly (Sackton et al., 2017) and for bumble bee workers (Erler et al., 2011). On the contrary, in the case of the tobacco hornworm (Gunnatina & Jiang, 2013) and the silkworm (Zhao et al., 2012), small or no changes were noted after microbial infection. Furthermore, in the honey bee genome one member of the A2M subfamily was identified, suggesting that the complement-like system may not be broadly conserved in insects (Palmer & Jiggins, 2015).

In the tsetse fly (Glossina spp.) we have recently demonstrated an increased expression of some TEPs in Trypanosoma brucei-infected salivary glands (Matetovici et al., 2016), which suggests a possible role of this protein family in controlling the parasite population in this micro-environment. Tsetse flies (Glossina spp.) are obligate blood-feeding insects and the exclusive biological vectors of different species of African trypanosomes in sub-Saharan Africa, protozoan parasites that cause devastating diseases in humans [human African trypanosomiasis (HAT) – sleeping sickness; caused by T. brucei complex] and livestock [animal African trypanosomiasis (AAT) – nagana; mainly caused by Trypanosoma congoense, Trypanosoma vivax]. The fly ingests the trypanosomes through feeding on an infected host. Once ingested, the parasite goes through a complex developmental cycle in the fly’s alimentary tract and mouthparts/salivary glands, depending on the trypanosome species (Rotureau & Van Den Abbeele, 2013), to achieve the final metacyclic infective state.

There are 31 species and subspecies of tsetse flies (Glossina) that are further subdivided into three groups (subgenera) – the Fusca, Morsitans and Palpalis groups – based on morphological characteristics (primarily genitalia structure), areal distribution, habitat and molecular data (Gooding & Kratsur, 2005; Kratsur, 2009). The Morsitans group (subgenus Glossina sensu stricto) includes tsetse fly vectors of HAT and AAT mainly in east and central Africa such as Glossina morsitans, Glossina pallidipes and Glossina austeni. The Palpalis group (subgenus Nemorhina) includes the major vectors of HAT in west and central Africa such as Glossina palpalis and Glossina fuscipes. Species of the Fusca group (subgenus Austenina) like Glossina brevipalpis have no significant medical or economic importance.

Most tsetse fly genomics and functional molecular research is focused on G. m. morsitans with only a few studies addressing the other tsetse fly vectors in a comparative approach (International Glossina Genome Initiative, 2014; Zhao et al., 2015; Macharia et al., 2016) owing to the limited availability of high-quality genomic information for these species. However, very recently the genomes of five tsetse fly species (beside G. m. morsitans) became publicly available, prompting us to perform a genomic and functional analysis of the tsetse thioester-containing proteins family. We show that the tsetse fly genomes encode between six to eight TEP genes, with differences between the species. Phylogenetic analysis of the family reveals that tsetse flies harbour some specific TEPs that are not found in any other dipterans. G. m. morsitans and G. palpalis gambiensis TEP genes display a tissue-specific expression pattern with some that are markedly up-regulated when the fly is infected with the trypanosome parasite. Interestingly, a different TEP response was observed to infection with T. brucei compared to T. congoense, indicating that the tsetse TEP response is trypanosome-specific. Taken together, our findings demonstrate a genomic expansion of tsetse fly-specific TEPs and are suggestive for the involvement of the TEP family in tsetse innate immunity with a possible role in the specific control of the trypanosome parasite in the fly.

Results

Genomic organization of G. m. morsitans TEPs

To identify members of the thioester-containing protein family, the G. m. morsitans (GmorY1) genome was screened with available D. melanogaster TEP sequences (Dmel_TEP1–TEP6). Seven genes coding for putative TEPs were found: Gmm_TEP1 (GMOY010996), Gmm_TEP2 (GMOY010998), Gmm_TEP3 (GMOY008955), Gmm_TEP4 (GMOY001989), Gmm_TEP5, Gmm_TEP6 and Gmm_TEP7. Four of these TEP genes were previously annotated (International Glossina Genome Initiative, 2014) and three were newly identified. In fact, the latter were previously wrongly predicted as being part of the TEP1 (GMOY010996) gene. Based upon these findings, we newly annotated these genes as Gmm_TEP1, Gmm_TEP5 and Gmm_TEP6 (Supporting Information Table S1). Owing to gaps in the genome assembly, only a partial coding sequence of 2386 bp was available for Gmm_TEP6. We determined the remaining part of the sequence (1821 bp) by PCR and Sanger sequencing (Supporting Information Figs S1 and S2). The Gmm_TEP7 gene was identified de novo by querying the scaffold sequence (Gmor-Y1:sd7180000652159) for open reading frames (ORFs) with the ORF finder program available from the National Center for Biotechnology Information (NCBI; Supporting
Information Figs S1 and S2). Exon–intron structures were predicted using GenScan software and validated by visual inspection in Integrative Genomics Viewer software (Thorvaldsdottir et al., 2013) of tsetse fly RNA sequencing (RNA-seq) mapped reads (Matetovici et al., 2016). The coordinates of the seven G. m. morsitans genes encoding thioester-containing proteins are listed in Supporting Information Table S1. Five of the genes, Gmm_TEP1, Gmm_TEP2, Gmm_TEP5, Gmm_TEP6 and Gmm_TEP7, are located on the same genomic scaffold spanning a region of over 80 kb, with Gmm_TEP7, Gmm_TEP6, Gmm_TEP5 and Gmm_TEP1 tandemly arrayed on one strand, whereas Gmm_TEP2 is present on the opposite strand with an approximately 7.6 kb overlap with Gmm_TEP1 (Fig. 1). This clustering may have resulted from a duplication-inversion event, followed by recent tandem duplication. The other two genes, Gmm_TEP3 and Gmm_TEP4, are found at different locations in the genome (Fig. 1). The gene length varies considerably between different family members, from 7 to 20 kb (Fig. 1 and Supporting Information Table S1). Furthermore, the presence of a putative signal peptide is predicted for Gmm_TEP2, Gmm_TEP4, Gmm_TEP5, Gmm_TEP6 and Gmm_TEP7, indicating that they are probably secreted. The functional domain architecture of the tsetse TEPs was predicted by the SMART program (Fig. 2). Out of the seven TEPs, six contain a thioester motif, whereas Gmm_TEP4 displays a low-density lipoprotein receptor domain class A (LDLa) instead. Moreover, four protease-binding alpha-2-macroglobulin domains (A2M-N, A2M-N2, A2M and A2M_comp) and one alpha-2-macroglobulin receptor binding domain (A2M_recep) are present in all TEP sequences (Fig. 2). Gmm_TEP4 also contains a transmembrane domain.

Similar to D. melanogaster (Lagueux et al., 2000) three additional features are found to be characteristic for the G. m. morsitans TEPs: (1) a highly conserved region of about 40 amino acids residues harbouring the thioester motif (or LDLa in the case of Gmm_TEP4); (2) a variable region of approximately 120 residues between the A2M-N-2 and A2M domain and (3) a cysteine signature in the C-terminal part (Supporting Information Fig. S1).

Genomic organization of Glossina species TEPs

Recently, the genomes of five other tsetse fly species became available at VectorBase: G. brevipalpis (Fusca group), G. fuscipes and G. p. gambiensis (Palpalis group), G. pallidipes and G. austeni (Morsitans group). We screened these genomes using the G. m. morsitans homologue sequences. The identified and annotated TEP sequences are summarized in Supporting Information Table S1. Overall, the genomes of these five tsetse species have six to eight TEP genes, with G. brevipalpis having the highest and G. fuscipes the lowest number of genes. The genome organization is similar to the one described for G. m. morsitans with genes encoding for TEP1, TEP2, TEP5, TEP6, TEP7 and TEP8, TEP9 located on the same genomic scaffold (Fig. 3) and TEP3 and TEP4 at different locations. The TEP6 gene is not
present in the Palpalis group (G. fuscipes and G. p. gambiensis). Moreover, two copies of the Gpg_TEPI3 gene are found in the G. p. gambiensis genome. Two homologues, Gb_TEPI8 and Gb_TEPI9, are present only in the Fusca group (G. brevipalpis, Fig. 3). Unfortunately, owing to gaps in the currently available genome assembly, only partial gene sequences for TEP5, TEP6 and TEP7 are available for most of the tsetse species. An iterative mapping and assembly of publicly available tsetse species-specific RNA-seq reads as an attempt to close the gaps was unsuccessful. Indeed, the C-terminal of these three genes is highly conserved making it difficult to distinguish between the multi mapping reads (data not shown).

A potential signal peptide is predicted for the following proteins: TEP2 for the Palpalis group, TEP3 for all the species (except G. p. gambiensis GPPi047920), TEP4 for all species (except G. pallidipes), TEP5 for all species (except G. brevipalpis), TEP6 for G. pallidipes and G. m. morsitans, TEP7 for all species (except G. brevipalpis) and TEP8 for G. brevipalpis (Supporting Information Table S1).

**Phylogenetic analysis of tsetse fly TEPs**

The TEP amino acid sequences of the six tsetse fly species were aligned with known TEP sequences using the MUSCLE program. As described above, the tsetse fly genomes encode for TEP genes belonging to the subfamilies MCR and iTEP. Homologous sequences found in different dipterans or blood-feeding arthropods were used in the construction of the phylogenetic tree (Supporting Information Table S2). Only complete sequences were maintained in the final alignment. The phylogenetic tree was constructed based on a 2115 amino acid alignment from 114 sequences using the maximum likelihood (ML) method and LG substitution model (Fig. 4A).

The MCR group comprises genes encoding for the macroglobulin complement related proteins. As expected, the TEP4 sequences (MCR) of the six tsetse species are grouped together with the other insect MCRs, forming a sister group to the Stomoxys calcitrans and Musca domestica clade.

The iTEP group includes the ‘ancestral’ iTEP group (Bou Aoun et al., 2011) as well as several different species-specific expansion groups: mosquito, fruit flies, house fly, stable fly and tsetse fly. The ‘ancestral’ iTEP group was already characterized in Bou Aoun et al., (2011), Palmer & Jiggins (2015) and Sekiguchi & Nonaka (2015). In order not to overload the tree, we included only sequences from dipterans and blood-feeding arthropods. All six tsetse fly TEP2 orthologues are part of this group and cluster together with D. melanogaster_TEP3 and with the other Brachycera fly species. The TEP3–TEP9 tsetse sequences are clustered in a branch restricted to the Brachycera suborder, which is divided into three main groups. The first group contains sequences found in five species of the Tephritidae family (Rhagoletis zephyria, Ceratitis capitata, Bactrocera cucurbitae, Bactrocera dorsalis and Bactrocera oleae). The second group consists of D. melanogaster_TEP4, TEP3 in tsetse species (with two copies for G. p. gambiensis), one sequence for M. domestica, two sequences located in a tandem array for S. calcitrans and two paralogous sequences in a tandem array in four Tephritidae species. The third group contains D. melanogaster_TEP1 and _TEP2, one sequence for each of the four Tephritidae species and an increased expansion in Muscidae and Glossinidae. Here, the tsetse fly TEP1, TEP5, TEP6, TEP7, TEP8 and TEP9 are all grouped together and no orthologues were found in any of the analysed dipteran species.

To gain a better view on the evolution of the TEP family within the Glossinidae, a second phylogenetic tree was constructed. The tree included only the TEP sequences found in the tandem array specific to tsetse flies and was based on a 1700-nucleotide alignment using the ML method with general time reversible (GTR) as the substitution model (Fig. 4B), representing the A1M_N and A2M_N_2 domains. Taking into consideration the phylogeny as described above (Fig. 4A), tsetse

---

**Figure 3.** Genomic organization of thioester-containing protein (TEP) genes in Glossinidae. The coloured boxes denote exons, the lines in between are introns and the smaller white boxes represent untranslated regions; the arrows indicate the direction of transcription of each gene; the numbers between the genes represent the length in base pairs of the intergenic regions, with the negative number representing an overlapping region on the opposite strand. ?, unknown. The dotted line around the genes represent the length in base pairs of the intergenic regions, with the negative number representing an overlapping region on the opposite strand. [Colour figure can be viewed at wileyonlinelibrary.com]
Figure 4. (Figure caption on the next page.)
TEP5 and TEP7 were generated after the divergence of Glossinidae and Muscidae and before the split of the different Glossinidae groups. TEP8 and TEP9 most likely emerged from TEP5 duplication after the separation of the Fusca group, as these genes are found only in G. brevipalpis and absent in the other groups. Moreover, TEP6 was most likely duplicated from TEP7 in the Mor sitans group after the divergence of the Palpalis group.

Conservation of the CGEQ–thioester site

Specific to the thioester-containing proteins is the CGEQ–thioester site that plays a key role in the formation of a covalent bond to microbial surfaces. Variation in this site is linked with the inability of the protein to bind to microbes. Sequence analysis of the six Glossina homologues (Fig. 5) with a thioester motif indicated that the canonical thioester motif is present (Supporting Information Fig. S1) in TEP1, TEP5, TEP6 and TEP7 whereas TEP2 and TEP3 showed variation from the canonical sequence (Fig. 5 and Supporting Information Fig. S1).

One type of variation identified in the TEP2 homologues is the replacement of cysteine (first position) with serine in species belonging to the Morsitans group and with asparagine in the Palpalis group. A second type of substitution in TEP2 is observed for the Fusca group, where glutamine (position four) is replaced by a histidine (Fig. 5A). Another substitution that is present in all tsetse TEP3 sequences is that of glycine (second position) by alanine. This type of substitution is also found in flies belonging to the Tephritidae family (C. capitata, B. cucurbitae, B. dorsalis and B. oleae; Fig. 5B) as well in vertebrate homologues like the mouse C4 (Nonaka et al., 1985) and cattle C4 (Ren et al., 1993).

G. m. morsitans TEP expression in different tissues

To examine transcriptional profiles of tsetse fly TEPs we performed tissue-specific reverse transcriptase quantitative-PCR (RT-qPCR) using cDNAs prepared from posterior midgut, anterior midgut, proventriculus, salivary glands, mouthparts and fat body tissues from adult non-infected G. m. morsitans flies (Fig. 6). In all analysed tissues different TEP transcripts were detected. Of note, some TEP genes were abundantly expressed in specific tissues: (1) TEP3: mouthparts and to a lesser extent the fat body; (2) TEP6: mainly in the anterior midgut and to a lesser extent in the proventriculus; and (3) TEP7: mainly in the fat body. Interestingly, several TEPs were strongly expressed in the tsetse fly mouthparts.
To examine the transcriptional profiles of the TEP genes in trypanosome-infected tsetse flies, three different infection experiments were performed: (1) *G. m. morsitans* infected with *T. brucei*; (2) *G. m. morsitans* infected with *T. congolense* and (3) *G. p. gambiensis* infected with *T. brucei*. *T. brucei* and *T. congolense* have a similar life cycle in the tsetse fly midgut and proventriculus but the development of the final metacyclic infective stage occurs in a different tissue: *T. brucei* in the salivary glands, whereas *T. congolense* has its final development in the mouth parts of the fly. Therefore, the TEP expression analysis was performed in these tissues as they are relevant for the different parasite stages. Additionally, the fat body was also included in the analysis as it is a systemic immune response organ (although not directly infected with trypanosomes).

**Figure 6.** Tissue-related thioester-containing protein (TEP) expression in non-infected *Glossina morsitans* flies. The relative expression of TEP transcripts in tissues dissected from non-infected flies was determined by reverse transcriptase quantitative-PCR in relation to ribosomal protein 49 (GMOY001799) and GMOY006676 as the reference genes. MGp, midgut posterior region; MGa, midgut anterior region; PV, proventriculus; SG, salivary glands; MP, mouthparts; FB, fat body. Values are means with SD, n = 5. [Colour figure can be viewed at wileyonlinelibrary.com]
In some cases, tissue specific. In the *T. brucei*-infected posterior midgut, Gmm_TEP1 (1.8-fold change) and Gmm_TEP3 (twofold change) showed significantly increased expression. Gmm_TEP6 was significantly up-regulated in the *T. brucei*-infected anterior midgut, showing a more than fourfold increase. Gmm_TEP1 (1.8-fold change) and Gmm_TEP4 (1.9-fold change) were also found to be up-regulated. Gmm_TEP3 was highly expressed in the *T. brucei*-infected proventriculus, with a more than sixfold up-regulation. Five TEPs were significantly up-regulated in *T. brucei*-infected salivary glands. Gmm_TEP2 was highly and specifically expressed in this tissue, with a more than 32-fold change increase and the other four were Gmm_TEP1 (9.6-fold change), Gmm_TEP4 (3.8-fold change), Gmm_TEP5 (7.7-fold change) and Gmm_TEP7 (threefold change). None of the analysed TEP genes showed differential expression in the fat body (Fig. 7).

**TEP expression in *T. congolense*-infected G. m. morsitans flies**

To investigate whether the observed differential TEP expression in *G. m. morsitans* flies was a general response to trypanosome infection or specific to the *T. brucei* parasite, a similar TEP expression analysis was performed in *T. congolense*-infected and age-matched non-infected flies (Fig. 8). In the posterior midgut, two genes showed an increase in expression in the presence of the parasite, Gmm_TEP3 (2.3-fold change) and Gmm_TEP5 (2.2-fold change). In the anterior midgut only Gmm_TEP6 was significantly up-regulated in the presence of *T. congolense*, displaying a more than fourfold increase. Additionally, Gmm_TEP1 presented an up-regulation of more than twofold in the mouthparts. No differential expression was noted in the proventriculus, salivary glands and fat body in *T. congolense*-infected flies.

**TEP expression in *T. brucei*-infected G. p. gambiensis flies**

TEP gene expression was also determined in another tsetse fly vector, *G. p. gambiensis*, which was infected with the *T. brucei* parasite and compared with age-matched non-infected flies (Fig. 9). As described above, the genome of *G. p. gambiensis* harbours six TEP...
genes (Fig. 3), there is no TEP6 orthologue. Three TEP genes were found to be significantly up-regulated following *T. brucei* infection: *Gpg_TEP2* (2.5-fold change) and *Gpg_TEP5* (1.9-fold change) in the proventriculus and *Gpg_TEP4* (1.8-fold change) in the salivary glands. No differential expression was observed in the posterior and anterior midgut or fat body of *T. brucei*-infected flies.

### Discussion

In the present study, we report for the first time a detailed genomic characterization of the TEP family from six tsetse fly species and a possible role of this protein family in the interaction with the trypanosome parasite population. Overall, depending on the species, the genome of the tsetse fly contains six to eight genes encoding for TEP proteins belonging to two specific TEP subfamilies ie insect TEP and MCR proteins (Fig. 4A), a similar repertoire as observed in *D. melanogaster* (Lagueux *et al*., 2000). The majority of *Glossina* TEP genes are located on the same scaffold and analysis of the gene structure (Fig. 3) suggests an expansion through segmental duplication (Hurles, 2004). A similar feature has been observed in the *An. gambiae* mosquito genome, where the 15 TEP genes are organized in four clusters and one isolated locus (Christophides *et al*., 2002). Furthermore, phylogenetic analysis revealed that this gene duplication gave rise to an expansion of specific tsetse TEPs that are not found in any other dipteran (Fig. 4A). Many similar genus-specific gene duplications can be found in the Brachycera suborder. Gene family duplication is one of the shaping forces in the evolution of an organism but understanding the factors maintaining paralogues in a genome is a difficult task (Innan & Kondrashov, 2010). One feature that may drive paralogue maintenance is host–pathogen interaction. Indeed, it has been observed that immune genes are experiencing rapid evolution in many organisms, armouring hosts with defence mechanisms to detect and fight pathogens. Such expansion in the TEP protein family has been documented in the tsetse fly and stable fly (this paper, Fig. 4A), mosquitoes (Little & Cobbe, 2005; Oubbard *et al*., 2008), fruit fly (Sackton *et al*., 2007) and in the house fly (Sackton *et al*., 2017). Interestingly,
these species are either blood feeders (like mosquitoes, tsetse and stable fly) or various pests such as fruit flies and house fly, species with a feeding habit that exposes them to a rather pathogen-unpredictable environment. In contrast, insects feeding on plants show no such TEP gene expansion in their genome; eg various species of bees (Apis, Bombus) have only three TEP genes (Bou Aoun et al., 2011; Palmer & Jiggins, 2015), the silk worm (Bombyx mori) has four genes (Zhao et al., 2012) and tobacco hornworm (Manduca sexta) two genes (Gunaratna & Jiang, 2013).

In arthropods, blood-feeding behaviour evolved independently more than 20 times and in insects at least five times at the order level (Mans, 2011). Consequently, in these organisms specific mechanisms and unique protein repertoires have evolved to acquire and process efficiently the bloodmeal, as well as the microbial challenge associated with it (ie ingestion of microbial contaminants during blood feeding on the host skin; Ribeiro & Arca, 2009). Expansion of the TEP protein family observed in various blood-feeding insects could probably be an adaptation to their respective life styles. As an exception, sand fly genomes (Lutzomyia longipalpis and Phlebotomus papatas) encode only three TEPs: one MCR gene and two iTEP genes (Fig. 4A).

The TEP genomic arsenal in the tsetse species investigated here showed variations between subgenera (groups). Both Palpalis group (G. p. gambiensis, G. fuscipes) and Fusca group flies (G. brevipalpis) lack the TEP6 gene whereas the latter also showed two additional TEPs (TEP8 and TEP9) that were not found in the other genomes. Whether TEP expansion in the tsetse groups is somehow associated with their biological differences regarding feeding preference or ability to transmit the trypanosome parasites remains to be determined. In our phylogenetic analysis (Fig. 4A), Gb_TEP8 and Gb_TEP9 formed a separate clade, indicating that they emerged from a Gb_TEP5 duplication. It would be interesting to see if this expansion is Fusca specific or restricted only to G. brevipalpis.

The main characteristic of the TEP protein family is the presence of a thioester covalent bond between the sulphydryl group of cysteine and the carbonyl group of glutamine (Tack et al., 1980; Law & Dodds, 1997). This thioester site (CGEQ) is a functionally essential site that allows the formation of a covalent bond to microbial...
This molecule is capable of binding *An. gambiae* and therefore might not have a microbe-binding capacity. Of the tsetse TEP2 proteins can form a thioester bond by a histidine. Overall, these results suggest that none replacement with asparagine. In addition, in the Fusca on in the Palpalis group a new mutation resulted in its replacement with serine and later in the separation of the Morsitans group, where two nucleotides resulted in its replacement with asparagine. In addition, in the Fusca group (*G. brevipalpis*) the glutamine residue is replaced by a histidine. Overall, these results suggest that none of the tsetse TEP2 proteins can form a thioester bond and therefore might not have a microbe-binding capacity. *An. gambiae* TEP1 has been shown to be a key molecule in the immune response against malaria parasites. This molecule is capable of binding *Plasmodium* ookinetes in the midgut epithelium basal lamina and targeting them for lysis (Blandin *et al.*, 2004). Interestingly, the presence of the thioester motif is not always a prerequisite for pathogen binding. *D. melanogaster* MCR was shown to specifically bind to the *Candida albicans* surface, and to subsequent promote its phagocytosis (Stroschein-Stevenson *et al.*, 2006). Moreover, not all TEPs directly interact with the pathogen surface. *Ae. aegypti* MCR was reported to control dengue virus infection in an indirect manner, by induction of antimicrobial peptides after its recruitment of a scavenger receptor-C (Xiao *et al.*, 2014). However, whether the TEP-based complement-like system plays a role in tsetse fly defence against trypanosomes remains to be determined.

Overlapping phylogenetic data with the results of differential gene expression studies is one of the first steps in understanding the possible function of the corresponding proteins. Therefore, we first investigated the expression profiles of the different TEP genes in various tissues of non-infected adult *G. m. morsitans* flies. All of the TEP genes were found to be expressed (at variable levels) in different parts of the alimentary tract and systemically in the fat body. Remarkably, strong expression of some TEPs was found to be tissue-specific i.e *Gmm_TEP3* (mouthparts, fat body), *Gmm_TEP6* (anterior midgut) and *Gmm_TEP7* (fat body) (Fig. 6). In *D. melanogaster* adult flies, the tissue-specific TEP expression pattern was found to be consistent with a role in innate immunity as TEP genes were expressed in haemocytes, in the fat body and in some barrier epithelia (Bou Aoun *et al.*, 2011). The elevated expression of TEP genes in the alimentary tract of tsetse fly could also be suggestive of an innate immune function with an involvement in the inactivation of a variety of microorganisms that are ingested when feeding on a host, followed by neutralization and hindering their proliferation/invasion to other tissues.

In addition, our data suggest a differential expression of TEPs in different parts of the tsetse alimentary tract in response to the trypanosome parasite during its complex developmental cycle in the fly (Rotureau & Van Den Abbeele, 2013). The significantly increased expression of *G. m. morsitans* TEPs upon *T. brucei* or *T. congolense* infection in the different tissues may indicate their involvement in the control of the parasite in the different tsetse micro-environments. Remarkably, a trypanosome-specific TEP response was observed, indicating the existence of a differential parasite-recognition mechanism.

In the posterior midgut, three TEP genes were observed to be up-regulated. *Gmm_TEP3* had increased expression both in response to *T. brucei* and *T. congolense* whereas two other genes showed a trypanosome-specific response: *Gmm_TEP1* to *T. brucei* and *Gmm_TEP5* to *T. congolense*. From the posterior midgut, it has been suggested that the established procyclic trypanosomes cross the peritrophic matrix, and migrate to the anterior part of the midgut. Here, we found significant up-regulation of *Gmm_TEP6* in response to both parasites, whereas *Gmm_TEP4* was up-regulated only in response to *T. brucei*. Differential expression of TEP genes in *T. brucei*-infected midgut was also observed in a macroarray analysis (Lehane *et al.*, 2003), where two genes corresponding to *Gmm_TEP6* and *Gmm_TEP3* were found to be up- and down-regulated, respectively.

From the anterior midgut trypanosomes invade the proventriculus (cardia), where several differentiation steps take place. Beside its role in the synthesis of the peritrophic matrix components (Lehane, 1997), this tissue can regulate the expression of some immune effectors like antimicrobial peptides and reactive oxygen/nitrogen species (Hao *et al.*, 2003). Here, *Gmm_TEP3* was found to be highly up-regulated in response to *T. brucei* but not to *T. congolense*. *T. congolense* parasites migrate to the mouthparts (cibarium and proboscis) and attach to the chitinous lining where they develop into an infective metacyclic form. Currently, knowledge on immune-related tsetse–trypanosome interactions in the cibarium and proboscis is lacking. Therefore, our finding of *Gmm_TEP1* up-regulation in response to *T. congolense* in the mouthparts is the first of its kind. In contrast to *T. congolense*, *T. brucei* parasites do not establish in the tsetse mouthparts but migrate to the
salivary glands, where they attach to the gland epithelium and complete their development into the infective metacyclic form. Consistent with our previously reported RNA-seq results (Matetovici et al., 2016), there is strong up-regulation of different TEPs (especially TEP2) in the T. brucei-infected salivary gland. None of these TEP genes were found to be differentially expressed in the salivary glands from T. congolense-infected flies (data not shown), suggesting that the TEP response is locally induced by the parasite in the gland. In both T. brucei- and T. congolense-infected flies none of the TEP genes were up-regulated in the fat body. A similar result was observed in the D. melanogaster abdominal fat body where no expression of TEP genes was detected before or after systemic immune challenge (Bou Aoun et al., 2011).

Of note is the significant up-regulation of Gmm_TEP4 (MCR), especially in the salivary glands, of T. brucei-infected flies. In D. melanogaster, the MCR has been shown to have multiple roles in innate immunity (Stroschein-Stevenson et al., 2006; Shokal et al., 2017) and in the formation and maintenance of the septate junctions (Batz et al., 2014; Hall et al., 2014). In T. brucei-infected salivary glands, the parasites tightly attach with their flagellum to the epithelial lining (Vickerman et al., 1988), thereby possibly damaging the septate junctions. Increased expression of Gmm_TEP4, together with the other eight genes encoding for septate junction proteins (Matetovici et al., 2016), would consequently be needed to repair and maintain the structural integrity of this tissue. Gmm_TEP4 up-regulation in salivary glands was not observed in T. congolense-infected flies (data not shown), again demonstrating the specificity of this parasite-related response.

The Morsitans group flies are known as good vectors of animal pathogenic trypanosomes (especially T. congolense and T. brucei brucei) and of the human-pathogenic T. brucei rhodesiense, in contrast to the Palps group flies, which are more refractory to trypanosome infection despite being the main vectors of the human T. brucei gambiense parasite. Our study revealed that the TEP expression profile in response to a trypanosome infection (T. brucei brucei) is strongly dissimilar in the two tsetse fly species, with only few genes up-regulated in G. p. gambiensis flies in sharp contrast to G. m. morsitans flies. Significant differences between these two species in the midgut immune-related response following trypanosome infection have been previously reported (Hamidou Soumana et al., 2014). Similar to our results, no TEP genes were found to be differentially expressed in a recent comparative transcriptome analysis of T. b. gambiense-infected midgut (Hamidou Soumana et al., 2015). Altogether, these results suggest differences in the innate immune responses of these two vectors to trypanosomes and strengthen the need for more comparative studies between the tsetse species.

The analysis of the tsetse TEP sequences revealed interesting information about their structure, evolutionary relationships and expression profiles under both normal and trypanosome infection conditions. We demonstrated the occurrence of a genomic expansion of specific tsetse TEPs that are not found in other dipterans. Moreover, we found general expression of all TEP genes in the alimentary tract, mouthparts and salivary glands, and a trypanosome-specific and tissue-related TEP response in the tsetse fly. These findings are suggestive for the involvement of the TEP family in tsetse innate immunity, with a possible role in controlling the parasite density in the infected tissues and/or acting in tissue damage repair.

**Experimental procedures**

**Sequence retrieval, identification and annotation of tsetse fly TEPs**

Genome sequences of six tsetse species, G. m. morsitans, G. pallidipes, G. austeni, G. fuscipes, G. p. gambiensis and G. brevipalpis, were obtained from the VectorBase database (https://www.vectorbase.org/; genome version is described in Supporting Information Table S1).

Exon–intron structures were predicted using GenSCAN software (http://genes.mit.edu/GENSCAN.html). The signal peptides of tsetse fly TEP amino acid sequences were predicted via the SIGNALP program (http://www.cbs.dtu.dk/services/SignalP/) and the protein domains by the SMART program (http://smart.embl-heidelberg.de/).

**Alignments and phylogenetic analysis**

Sequences of the TEPs from D. melanogaster were retrieved from FlyBase (http://flybase.org/). Sequences belonging to Muscidae, Culicidae, Psychodidae, Cimicidae, Reduviidae, Sarcoptidae and Ixodidae species were retrieved from VectorBase by blastp searches using the annotated D. melanogaster sequences. The Tephritidae sequences were retrieved from VectorBase by blastp searches using the NCBI protein database after they were identified by NCBI blastp searches (https://blast.ncbi.nlm.nih.gov/Blast.cgi) using annotated D. melanogaster sequences. The obtained protein sequences (Supporting Information Table S2) were used to create a multiple alignment with the MUSCLE algorithm (http://www.ebi.ac.uk/Tools/msa/muscle/). The phylogenetic tree was constructed using the ML method implemented in PhyML (Guindon et al., 2010), with the best-fitting evolutionary model and 100 bootstrap replications. The trees were displayed and annotated with Interactive tree of life (ITOL) v. 3 (Letunic & Bork, 2016).

**Tsetse fly infection and tissue collection**

Male G. morsitans morsitans from the colony at the Institute of Tropical Medicine (Antwerp, Belgium) were used in all experiments (Elsen et al., 1993). The G. pallipalpis gambiensis flies...
emerged from pupae obtained from the tsetse fly colony of the Institute of Zoology of the Slovak Academy of Sciences (Bratislava). All experimental flies were maintained at 26 ± 0.5°C and 65 ± 5% relative humidity.

The T. brucei brucei AnTAR1 (derived from EATRO 1125; Le Ray et al., 1977) and T. congolense MSOROM7 (Tihon et al., 2017) strains were used for the tsetse fly infection experiments. Freshly emerged flies were fed 24–48 h after their eclosion with a trypanosome-infected bloodmeal supplemented with 10 mM reduced L-glutathione to enhance midgut infection rates in the tsetse fly (MacLeod et al., 2007). For this, parasitized blood was harvested with heparin from cyclophosphamide-immune suppressed mice (Endoxan®, Baxter, Germany) at 6–7 days postinfection and mixed with defibrinated horse blood (E&O Laboratories Limited, London, UK) to obtain around 2 × 10^6 bloodstream form trypanosomes/ml in the initial bloodmeal. Afterwards, the flies were maintained for 4 weeks and fed every 2–3 days on non-infected defibrinated horse blood using an artificial membrane feeding system. Three different tsetse fly–trypanosome infection protocols were performed: (1) G. m. morsitans–T. brucei, (2) G. m. morsitans–T. congolense and (3) G. p. gambiensis–T. brucei.

Twenty-eight days after the infective bloodmeal, T. brucei salivary gland infected flies were selected by inducing probing on prewarmed (37°C) glass slides, which were microscopically examined for the presence of metacyclic trypanosomes (modification of the method of Burtt, 1946). Immediately after screening the flies were fed and maintained for another 72 h before dissection. To select for the T. congolense-infected flies, individuals were microscopically examined during dissection and only infected tissues were collected. For each of the three fly infection groups, a control group of age-matched and non-infected tsetse flies was maintained on the same feeding regime. From all these experimental groups the following tissues were collected by dissection: salivary glands (pool of 20); proventriculus (pool of 10); anterior midgut (pool of three); posterior midgut (pool of three) and fat body (pool of three).

RNA isolation and TEP tissue expression profiling by RT-qPCR

Total RNA was isolated from the pooled tissues using TRIzol® reagent (Life Technologies, Rockville, MD, USA) according to the manufacturer’s instructions. Samples were DNaseI treated (Ambion, Life Technologies, Rockville, USA) and first-strand cDNA was reverse transcribed using oligo(dT)15 primer and Transcriptor Reverse Transcriptase (Roche, Indianapolis, IN, USA), following the manufacturer’s instructions. RT-qPCR reactions of 20 μl were performed with a SensiMIX™SYBR® No-ROX kit (Bioline, London, UK) and 0.5 μM of each primer (Supporting Information Table S3). They were run on a Light Cycler 480 system (Roche) with the following thermal cycling conditions: 10 min at 95°C (polymerase activation) followed by 40 cycles of 10 s/95°C (denaturation), 10 s/60°C (annealing) and 30 s/72°C (elongation). For each tissue five biological replicates were used. Data were analysed using BioGazelR Qbase™ version 1.5 software (Biogazelle NV, Belgium) in order to evaluate reference gene stability and to obtain normalized values for the tested genes in the different tissue samples.

G. m. morsitans TEP6 cloning

The missing Gmm_TEP6 sequence was obtained by PCR using the following primer set: pET22b_Tep6_Fw 5'-aaggggtgggtgggtgggtggTGTGAACATGACCAAGGCTTCA TCTGGC-3' and pET22b_TEP6_Rev 5'-cttaaagagagatacat atgTGGTGTCTTCTTCTGAGTACGATCCACTGTA-3' (vector sequence is in lowercase and gene-specific sequence in uppercase). The PCR reaction was set up as follows: 1× Phusion Master Mix (New England Biolabs, Beverly, MA, USA), 100 ng cDNA from T. brucei-infected midgut (obtained as described above) and 0.5 μM of each primer. The PCR cycling conditions were as follows: 1 min at 98°C for initial denaturation, followed by 30 cycles of 15 s at 98°C (denaturation), 30 s at 72°C (annealing) and 60 s at 72°C (elongation), and another 10 min at 72°C (final elongation). The amplification product was cloned into linearized Xhol and Ndel restriction enzymes pET22b vector (Novagen, Madison, WI, USA) using a NEBuilder HiFi DNA Assembly Cloning kit (New England Biolabs) and transformed into NEB® 5-alpha Competent E. coli cells. Four positive clones were Sanger sequenced using T7 promoter/terminator primers. After sequencing a new internal primer (Tep6_Fw: AACCGCTTATGTGGCTCGTT) was designed and the plasmid was resequenced, covering in way the entire gap.

Acknowledgements

We wish to acknowledge the technical assistance of Jos Van Hees and to thank Peter Takáč from the Institute of Zoology of the Slovak Academy of Sciences (Bratislava) for providing us the G. p. gambiensis flies. This work was funded by the Interver University Attraction Pole program (P7/41) (Belspo) and by a Structural Research Funding offered by the Institute of Tropical Medicine (SOFI-TRIPARTITE) grant. We also would like to thank Wes Warren and his team at the McDonnell Genome Institute at Washington University, Saint Louis, for the genomic library construction, sequencing and data curation of the different tsetse species, supported by 584NIH-NHGRI grant 5U54HG00307907 to Richard K. Wilson. We wish to acknowledge the Vectorbase database for hosting the tsetse species genomes, supported by NAIID NIH grant HHSN272201400029C and the International Glossina Genome Initiative (IGGI).

References

Arefin, B., Kucerova, L., Dobes, P., Markus, R., Strnad, H., Wang, Z. et al. (2014) Genome-wide transcriptional analysis of Drosophila larvae infected by entomopathogenic nematodes shows involvement of complement, recognition and extracellular matrix proteins. J Innate Immun 6: 192–204.

Batz, T., Forster, D. and Luschnig, S. (2014) The transmembrane protein Macroglobulin complement-related is essential for septate junction formation and epithelial barrier function in Drosophila. Development 141: 899–908.

Blandin, S. and Levashina, E.A. (2004) Thioester-containing proteins and insect immunity. Mol Immunol 40: 903–908.

© 2018 The Authors. Insect Molecular Biology published by John Wiley & Sons Ltd on behalf of Royal Entomological Society, 27, 414–428.
Thioester-containing proteins in the tsetse fly

Blandin, S., Shiao, S.H., Moita, L.F., Janse, C.J., Waters, A.P., Kafatos, F.C. et al. (2004) Complement-like protein TEP1 is a determinant of vectorial capacity in the malaria vector Anopheles gambiae. Cell 116: 661–670.

Blandin, S.A., Marois, E. and Levashina, E.A. (2008) Antimalarial responses in Anopheles gambiae: from a complement-like protein to a complement-like pathway. Cell Host Microbe 3: 364–374.

Bou Aoun, R., Hetru, C., Troxler, L., Doucet, D., Ferrandon, D. and Matt, N. (2011) Analysis of thioester-containing proteins during the innate immune response of Drosophila melanogaster. J Innate Immun 3: 52–64.

Burtt, E. (1946) Salivation by Glossina morsitans on to glass slides; a technique for isolating infected flies. Ann Trop Med Parasitol 40: 141–144.

Castillo, J.C., Creasy, T., Kumar, P., Shetty, A., Shokal, U., Tallon, L.J. et al. (2015) Drosophila anti-nematode and antibacterial immune regulators revealed by RNA-Seq. BMC Genomics 16: 519.

Christophides, G.K., Zdobnov, E., Barillas-Mury, C., Birney, E., Blandin, S., Blass, C. et al. (2002) Immunity-related genes and gene families in Anopheles gambiae. Science 298: 159–165.

Dostalova, A., Rommelaere, S., Poidevin, M. and Lemaître, B. (2017) Thioester-containing proteins regulate the Toll pathway and play a role in Drosophila defence against microbial pathogens and parasitoid wasps. BMC Biol 15: 79.

Eisen, P., Van Hees, J. and De Lil, E. (1993) L’historique et les conditions d’elevage des lignees de glossines (Diptera, Glossinidae) maintenues a l’Institut de Medecine tropicale Prince Leopold d’Anvers. J Afr Zool 107: 439–449.

Erlér, S., Popp, M. and Lattorff, H.M. (2011) Dynamics of immune system gene expression upon bacterial challenge and wounding in a social insect (Bombus terrestris). PLoS ONE 6: e18126.

Gooding, R.H. and Krafsur, E.S. (2005) Tsetse genetics: contributions to biology, systematics, and control of tsetse flies. Annu Rev Entomol 50: 101–123.

Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W. and Gascuel, O. (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol 59: 307–321.

Gunaratna, R.T. and Jiang, H. (2013) A comprehensive analysis of the Manduca sexta immunotranscriptome. Dev Comp Immunol 39: 388–398.

Hall, S., Bone, C., Oshima, K., Zhang, L., Mcgraw, M., Lucas, B. et al. (2014) Macroglobulin complement-related encodes a protein required for sepal junction organization and paracellular barrier function in Drosophila. Development 141: 889–898.

Hamidou Soumana, I., Tchicaya, B., Chuchana, P. and Geiger, A. (2014) Midgut expression of immune-related genes in Glossina palpalis gambiensis challenged with Trypanosoma brucei gambiens. Front Microbiol 5: 609.

Hamidou Soumana, I., Klopp, C., Ravel, S., Nabihoudine, I., Tchicaya, B., Parrinello, H. et al. (2015) RNA-seq de novo assembly reveals differential gene expression in Glossina palpalis gambiensis infected with Trypanosoma brucei gambiens vs. non-infected and self-cured files. Front Microbiol 6: 1259.

Hao, Z., Kasumba, I. and Aksoy, S. (2003) Proventriculus (cardia) protein to a complement-like pathway. Insect Biochem Mol Biol 33: 1155–1164.

Hurles, M. (2004) Gene duplication: the genomic trade in spare parts. PLoS Biol 2: E206.

Igbin, C.O., Tordoff, K.P., Moeschberger, M.L., Griffen, A.L. and Leys, E.J. (2011) Porphyromonas gingivalis-host interactions in a Drosophila melanogaster model. Infect Immun 79: 449–458.

Innan, H. and Kondrashov, F. (2010) The evolution of gene duplications: classifying and distinguishing between models. Nat Rev Genet 11: 97–108.

International Glossina Genome Initiative. (2014) Genome sequence of the tsetse fly (Glossina morsitans): vector of African trypanosomiasis. Science 344: 380–386.

Krafsur, E.S. (2009) Tsetse flies: genetics, evolution, and role as vectors. Infect Genet Evol 9: 124–141.

Lagueux, M., Perroudou, E., Levashina, E.A., Capovilla, M. and Hoffmann, J.A. (2000) Constitutive expression of a complement-like protein in toll and JAK gain-of-function mutants of Drosophila. Proc Natl Acad Sci USA 97: 11427–11432.

Law, S.K. and Dodds, A.W. (1997) The internal thioster and the covalent binding properties of the complement proteins C3 and C4. Protein Sci 6: 263–274.

Le Ray, D., Barry, J.D., Easton, C. and Vickerman, K. (1977) First tsetse fly transmission of the “AnTat” serodeme of Trypanosoma brucei. Ann Soc Belg Med Trop 57: 369–381.

Lehane, M.J. (1997) Peritrophic matrix structure and function. Annu Rev Entomol 42: 525–550.

Lehane, M.J., Aksoy, S., Gibson, W., Kerhornou, A., Berriman, M., Hamilton, J. et al. (2003) Adult midgut expressed sequence tags from the tsetse fly Glossina morsitans morsitans and expression analysis of putative immune response genes. Genome Biol 4: R63.

Letunic, I. and Bork, P. (2016) Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Res 44: W242–W245.

Levashina, E.A., Moita, L.F., Blandin, S., Vriend, G., Lagueux, M. and Kafatos, F.C. (2001) Conserved role of a complement-like protein in phagocytosis revealed by dsRNA knockout in cultured cells of the mosquito, Anopheles gambiae. Cell 104: 709–718.

Lih, M., Gao, P., Yuan, C., Zhang, Y. and Cao, S. (2015) Identification of a new complement-related protein in Tsetse flies. Insect Mol Biol 24: 599–605.

Liu, W., Xie, Y., Ma, J., Luo, X., Nie, P., Zuo, Z. et al. (2015) Identification of a new complement-related protein in Tsetse flies. Insect Mol Biol 24: 599–605.

Macleod, E.T., Mann, B.J. (2011) Evolution of vertebrate hemostatic and inflammatory control mechanisms in blood-feeding arthropods. J Innate Immun 3: 41–51.

Matetovici, I., Caljon, G. and Van Den Abbeele, J. (2016) Tsetse fly tolerance to T. brucei infection: transcriptome analysis of...
trypanosome-associated changes in the tsetse fly salivary gland. *BMC Genomics* 17: 971.

Nonaka, M., Nakayama, K., Yeul, Y.D. and Takahashi, M. (1985) Complete nucleotide and derived amino acid sequences of the fourth component of mouse complement (C4). Evolutionary aspects. *J Biol Chem* 260: 10936–10943.

Obbard, D.J., Callister, D.M., Jiggins, F.M., Soares, D.C., Yan, G. and Little, T.J. (2008) The evolution of TEP1, an exceptionally polymorphic immunity gene in *Anopheles gambiae*. *BMC Evol Biol* 8: 274.

Palmer, W.J. and Jiggins, F.M. (2015) Comparative genomics reveals the origins and diversity of arthropod immune systems. *Mol Biol Evol* 32: 2111–2129.

Pompon, J. and Levashina, E.A. (2015) A new role of the mosquito complement-like cascade in male fertility in *Anopheles gambiae*. *PLoS Biol* 13: e1002255.

Ren, X.D., Dodds, A.W. and Law, S.K. (1993) The thioester and isotypic sites of complement component C4 in sheep and cattle. *Immunogenetics* 37: 120–128.

Ribeiro, J. and Arca, B. (2009) From sialomes to the sialoviruse: an insight into salivary potion of blood-feeding insects. *Adv Insect Physiol* 37: 59–118.

Rotureau, B. and Van Den Abbeele, J. (2013) Through the dark continent: African trypanosome development in the tsetse fly. *Front Cell Infect Microbiol* 3: 53.

Sackton, T.B., Lazzaro, B.P., Schlenke, T.A., Evans, J.D., Hultmark, D. and Clark, A.G. (2007) Dynamic evolution of the innate immune system in *Drosophila*. *Nat Genet* 39: 1461–1468.

Sackton, T.B., Lazzaro, B.P. and Clark, A.G. (2017) Rapid expansion of immune-related gene families in the house fly, *Musca domestica*. *Mol Biol Evol* 34: 857–872.

Sekiguchi, R. and Nonaka, M. (2015) Evolution of the complement system in protostomes revealed by de novo transcriptome analysis of six species of Arthropoda. *Dev Comp Immunol* 50: 58–67.

Shokal, U. and Elefftherianos, I. (2017) The *Drosophila* Thioester containing Protein-4 participates in the induction of the cellular immune response to the pathogen *Photorhabdus*. *Dev Comp Immunol* 76: 200–208.

Shokal, U., Kopydlovskiy, H. and Elefftherianos, I. (2017) The distinct function of Tep2 and Tep6 in the immune defense of *Drosophila melanogaster* against the pathogen *Photorhabdus*. *Virulence* 12: 1–15.

Stroschein-Stevenson, S.L., Foley, E., O’Farrell, P.H. and Johnson, A.D. (2006) Identification of *Drosophila* gene products required for phagocytosis of *Candida albicans*. *PLoS Biol* 4: e4.

Tack, B.F., Harrison, R.A., Janatova, J., Thomas, M.L. and Prahl, J.W. (1980) Evidence for presence of an internal thioester bond in third component of human complement. *Proc Natl Acad Sci USA* 77: 5764–5768.

Thorvaldsdottir, H., Robinson, J.T. and Mesirov, J.P. (2013) Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* 14: 179–192.

Tihon, E., Imamura, H., Dujardin, J.C., Van Den Abbeele, J. and Van Den Broeck, F. (2017) Discovery and genomic analyses of hybridization between divergent lineages of *Trypanosoma congolense*, causative agent of Animal African Trypanosomiasis. *Mol Ecol* 26: 6524–6538.

Vickerman, K., Tetley, L., Hendry, K.A. and Turner, C.M. (1988) Biology of African trypanosomes in the tsetse fly. *Biol Cell* 64: 109–119.

Wertheim, B., Kraaijeveld, A.R., Schuster, E., Blanc, E., Hopkins, M., Fletcher, S.D. et al. (2005) Genome-wide gene expression in response to parasitoid attack in *Drosophila*. *Genome Biol* 6: R94.

Xiao, X., Liu, Y., Zhang, X., Wang, J., Li, Z., Pang, X. et al. (2014) Complement-related proteins control the flavivirus infection of *Aedes aegypti* by inducing antimicrobial peptides. *PLoS Pathog* 10: e1004027.

Yassine, H., Kamareddine, L. and Osta, M.A. (2012) The mosquito melanization response is implicated in defense against the entomopathogenic fungus *Beauveria bassiana*. *PLoS Pathog* 8: e1003029.

Zhao, P., Dong, Z., Duan, J., Wang, G., Wang, L., Li, Y. et al. (2012) Genome-wide identification and immune response analysis of serine protease inhibitor genes in the silkworm, *Bombyx mori*. *PLoS One* 7: e31168.

Zhao, X., Silva, T.L.a.e., Cronin, L., Savage, A.F., O’Neill, M., Nerima, B. et al. (2015) Immunogenicity and serological cross-reactivity of saliva proteins among different tsetse species. *PLoS Negl Trop Dis* 9: e0004038.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher’s website:

**Table S1.** Glossina spp. thioester-containing protein genes information (sheet 1) and tsetse species genome data (sheet 2).

**Table S2.** Species used in the phylogenetic tree.

**Table S3.** Reverse transcriptase quantitative-PCR primers used for thioester-containing protein (TEP) expression analysis. *Glossina morsitans* TEP genes (sheet 1); *G. m. morsitans*–*Trypanosoma brucei* reference genes (sheet 2); *G. p. gambiensi* TEP genes (sheet 3) and *G. p. gambiensi*–*T. brucei* reference genes (sheet 4).

**Figure S1.** Multiple sequence alignment of *G. m. morsitans* insect thioester-containing proteins (iTEPs). The conserved functional domains including four proteinase-binding alpha-2-macroglobulin (α2M) domains (A2M_N, A2M_N2, A2M and A2M_comp), one A2M receptor binding domain (A2M_receptor) and a thioester domain (Thioester) are marked on the top of the alignment with a blue line. The signal peptide sequence at the N-terminal is in italics and the putative thioester motif (GCGERQ) is highlighted in black. The thioester associated histidine (His) is indicated by an arrow at position 1991. Six cysteines located at 1392–1530 in the thioester-containing protein sequences.

**Figure S2.** *G. m. morsitans* thioester-containing protein sequences.