

Article

Evolution and Structural Analyses of *Glossina morsitans* (Diptera; Glossinidae) Tetraspanins

Edwin K. Murungi ^{1,†,*}, Henry M. Kariithi ^{2,3,†,*}, Vincent Adunga ⁴, Meshack Obonyo ⁴ and Alan Christoffels ¹

¹ South African National Bioinformatics Institute (SANBI), University of the Western Cape, Private Bag X79, Bellville, Cape Town 7535, South Africa; E-Mail: alan.christoffels@uwc.ac.za

² Biotechnology Research Institute, Kenya Agricultural and Livestock Research Organization (KALRO), P.O. Box 57811, Kaptagat Rd, Nairobi 00200, Kenya

³ Laboratory of Virology, Wageningen University, Droevendaalsesteeg 1, Wageningen 6708 PB, The Netherlands

⁴ Department of Biochemistry and Molecular Biology, Egerton University, P.O. Box 536, Egerton 20115, Kenya; E-Mails: vowino@gmail.com (V.A.); obonyom@gmail.com (M.O.)

† These authors contributed equally to this work.

* Authors to whom correspondence should be addressed; E-Mails: eddkimm@gmail.com (E.K.M.); henry.kariithi@kalro.org or h_kariithi@yahoo.com (H.M.K.); Tel.: +254-789-716059 (E.M.); +254-706-561-535 (H.M.K.).

External Editor: Michael J. Stout

Received: 27 September 2014; in revised form: 23 October 2014 / Accepted: 2 November 2014 /

Published: 12 November 2014

Abstract: Tetraspanins are important conserved integral membrane proteins expressed in many organisms. Although there is limited knowledge about the full repertoire, evolution and structural characteristics of individual members in various organisms, data obtained so far show that tetraspanins play major roles in membrane biology, visual processing, memory, olfactory signal processing, and mechanosensory antennal inputs. Thus, these proteins are potential targets for control of insect pests. Here, we report that the genome of the tsetse fly, *Glossina morsitans* (Diptera: Glossinidae) encodes at least seventeen tetraspanins (GmTsps), all containing the signature features found in the tetraspanin superfamily members. Whereas six of the GmTsps have been previously reported, eleven could be classified as novel because their amino acid sequences do not map to characterized tetraspanins in the

available protein data bases. We present a model of the GmTsp by using GmTsp42Ed, whose presence and expression has been recently detected by transcriptomics and proteomics analyses of *G. morsitans*. Phylogenetically, the identified GmTsp segregate into three major clusters. Structurally, the GmTsp are largely similar to vertebrate tetraspanins. In view of the exploitation of tetraspanins by organisms for survival, these proteins could be targeted using specific antibodies, recombinant large extracellular loop (LEL) domains, small-molecule mimetics and siRNAs as potential novel and efficacious putative targets to combat African trypanosomiasis by killing the tsetse fly vector.

Keywords: tetraspanins; GmTsp; LEL; CD63; *Glossina morsitans*; *Trypanosoma*; phylogenetics; modeling; positive selection

1. Introduction

Tsetse flies (genus *Glossina*) are vectors of the unicellular flagellated trypanosome parasites (genus *Trypanosoma*) that cause African trypanosomiasis, a group of debilitating zoonotic neglected tropical diseases (NTDs) commonly referred to as sleeping sickness in humans and *nagana* in cattle [1]. African trypanosomiasis have been described as one of the “root causes of hunger and poverty” in sub-Saharan Africa [2]. Left untreated, these diseases can be fatal, with fatalities differing from one group of trypanosome to another [3,4]. The diseases are difficult to treat, and there are no efficacious vaccines. None of the available trypanocidal drugs are ideal: the most widely used drug, melarsoprol, is toxic and up to 10% of the patients die from the treatment itself [5,6]. Besides, the treatment schedules for these drugs are prolonged, excruciatingly painful, and requires continuous hospitalization [7]. Therefore, the control of the disease vector (tsetse) is of critical importance, and probably represents the most sustainable trypanosomiasis control method. Vector control using tsetse fly insecticide-impregnated traps [8], application of broad-spectrum insecticides [9], live baits [10], and the mating of virgin wild-type females with sexually-sterilized males [11] are some of the strategies applied to combat African trypanosomiasis. Although effective, some of these methods suffer several drawbacks such as increasing drug resistance/counterfeits [12,13], drug toxicity [5,6], and various environmental concerns such as loss of biodiversity and uncertainties on the fate of non-target organisms [14]. As such, identification of novel molecular targets that could be disrupted in the insect vector may provide new approaches of combating the diseases. One of the potential molecular targets is the tetraspanins superfamily (transmembrane 4 superfamily; TM4SF), a growing protein family of evolutionarily conserved integral membrane/surface proteins expressed in a wide range of multi-cellular organisms [15,16]. Available data suggest that pathogens (viruses, intracellular bacteria, and parasites) can “hijack” tetraspanins to gain entry into cells, for cytoplasmic trafficking after infection, and for final egress [17].

Tetraspanins are small (200–350 amino acid residues) type III surface glycoproteins with well defined structural motifs: short intracellular N- and C-termini connected by four transmembrane domains (TM1-TM4) with several conserved polar residues, which are interconnected by one small extracellular loop (SEL) and one large extracellular loop (LEL) [18,19]. The SEL domain contains

20–28 amino acid residues, while the LEL domain is made up of 76–131 amino acid residues [20]. The LEL domain is located between TM3 and TM4 and characteristically contains four to eight invariant conserved cysteine residues, two of which define a conserved Cys-Cys-Gly (CCG) motif and form intramolecular disulfide bonds crucial for structural integrity and functional specificities to tetraspanins [21–23]. More than 50% of tetraspanins also contain a Pro-x-x-Cys-Cys (PxxCC) motif, where “x” is any amino acid [24]. Tetraspanin proteins are synthesized in the endoplasmic reticulum (ER), and after palmitoylation, these proteins often form homodimers, which are subsequently transported to the cell surface to function as building blocks of large integrated signaling complexes or tetraspanin-enriched microdomains (TEMs) [25]. A combination of the above-mentioned features distinguishes tetraspanins from other four TM-domain proteins.

Tetraspanins are expressed in various cell types, and are implicated in a multitude of biological processes including signaling, cell adhesion, intracellular trafficking, and pathogen infections [20,26–28]. They facilitate these processes by a flare of rather promiscuous and unique associations with one another, and with a variety of non-tetraspanin integral macromolecules such as integrins [29], growth factor receptors/co-receptors, proteoglycans, complement-regulatory proteins, uroplakins, rhodopsin, members of the immunoglobulin superfamily, and others [30–32]. Via these interactions, tetraspanins form the TEM networks. The complexes arising from these protein-protein interactions have been proposed to play co-stimulatory roles in the activation of intracellular signaling pathways such as the N-terminal Jun Kinase pathway [33]. In animals, tetraspanins function as organizers of membrane signaling complexes, including cytoplasmic enzymes such as protein kinase C (PKC), phosphatidylinositol-4-kinase, and membrane components such as integrins and lipids [24,34]. Although the formation of the tetraspanins-PKC complex is independent of integrins, tetraspanins act as linker molecules by recruiting PKC into proximity with specific integrins.

Despite their biological importance, most tetraspanins have not been functionally explored because of their subtle and overlapping roles [35]. However, available data provide sufficient evidence that tetraspanins play important roles during bacterial [36], protozoal [37], and entomopathogenic fungal [38] infections. Since pathogens often “hijack” their hosts’ tetraspanin-organized natural cell processes (e.g., adhesion, internalization, vesicle trafficking, *etc.*) [39,40], the pathogen-tetraspanin interaction potentially offers novel therapeutic strategies against pathogenesis. Potentially, various *Trypanosoma* lifecycle stages could be targeted by tetraspanin-specific agents, such that anti-trypanocidal agents could be delivered to the parasite-containing vesicles. The advantage of such an approach is that tetraspanins are host-derived, implying that the possibility of the parasite developing resistance is much lower as opposed to the application of conventional trypanocidal drugs. Interrupting cellular processes that trypanosomes depend on for infection, multiplication/proliferation, and dissemination could greatly complement other available vector control strategies against African trypanosomiasis.

Due to their documented roles in pathogen-host interactions, tetraspanins are potential targets to control transmission of disease-causing parasites by insect vectors. For instance, insect tetraspanins could be targets for specific antibodies, recombinant LEL domains, small-molecule mimetics and siRNAs as potential strategies to combat diseases, in the interest of this work, African trypanosomiasis. Towards this end, we characterized the putative *Glossina morsitans* tetraspanins (abbreviated in this paper as GmTsp), and modeled for these proteins by using GmTsp42Ed. It is our opinion that the identification of tetraspanins will uncover efficacious novel insecticide targets against the tsetse fly.

2. Experimental

2.1. Sequential Retrieval, Ortholog Search and Bioinformatic Analyses of Tetraspanins

The tetraspanin gene sequences were retrieved from the UniProt database [41] for *Drosophila melanogaster* and *Musca domestica*, and from the VectorBase [42] for *Glossina morsitans*. To identify genes encoding tetraspanins in *G. morsitans*, the retrieved *D. melanogaster* annotated tetraspanins were used to construct a query protein set. BLASTp [43] was then used to search for all tetraspanin genes using a threshold setting as E-values $\leq 1e-4$. The putative *G. morsitans* tetraspanins obtained were then used to construct a Hidden Markov model (HMM) to exhaustively search the *G. morsitans* genome. To avoid false-positive hits that commonly arise during automated searches, the presence of the conserved LEL domain in each putative tetraspanin obtained was ascertained using InterPro version 48.0 [44]. The new nucleotide and deduced amino acid sequences of tetraspanins were analyzed from the recently published *G. morsitans* genomic data sets [45]. Transmembrane domains in the candidate sequences were then predicted using SMART [46] and THMHMM server v. 2.0 [47]. The TM helices of the candidate tetraspanins and their membrane-spanning segments were further discriminated using SOSUI [48], and the hydrophobicity quality of the protein sequences was predicted using ProtScale at ExPASy bioinformatics resource portal [49]. Further analyses of the identified *G. morsitans* putative tetraspanins were annotated using the Blast2GO v. 2.7.2 [50]. To investigate whether the retrieved *G. morsitans* putative tetraspanins are palmitoylated, we used the clustering and scoring strategy (CSS) algorithm (at high threshold), CSS-Palm v. 4.0, which predicts the likelihood of palmitoylation within inputted amino acid sequences [51].

2.2. Multiple Sequence Alignment and Phylogenetic Analysis

Putative tetraspanins sequences were aligned using MUSCLE [52] with default settings. Sites with alignment ambiguities were excluded manually using Jalview [53]. Maximum likelihood (ML) and Bayesian phylogenetic searches were performed using the Le and Gascuel (LG) model of amino acid substitution [54]. The model of sequence evolution prior to each analysis was determined using Prot-Test v. 3.2.1 [55]. The ML analysis was performed using the program PhyML v. 3.0 [56] with the following parameters: substitution model = LG; prop_invar = 0.0; gamma = empirical; nb_subst_cat = 4. Bootstraps for the ML were generated using 100 replicates of bootstrapping. Support for the nodes was assessed with 100 bootstrap replicates. Bayesian inference analysis was conducted with MrBayes v. 3.2 [57]. Ten million generations of MCMC simulation were used along with a burnin of 10,000 generations. This number of MCMC generations allowed for convergence of simulation chains and reduction of split frequencies to an acceptable level. Phylogenetic trees were built using the ML method and rendered using the interactive Tree Of Life (iTOL) v. 2.2.2 [58].

2.3. Positive Selection Analysis

Patterns of sequence change using nonsynonymous/synonymous (d_N/d_S) rate ratios were performed using five methods for detecting positive selection available from the DATAMONKEY [59] web server [59]. The five methods used in this study were: Single Likelihood Ancestor Counting model

(SLAC), the Fixed Effect Likelihood model (FEL), the Random Effect Likelihood model (REL), the Mixed Effects Model of Evolution (MEME) and the Fast Unbiased Bayesian AppRoximation (FUBAR) [60–62]. For these analyses, the best fitting nucleotide substitution model was determined through the automatic model selection tool available on the DATAMONKEY server. Prior to running the analyses, the dataset was screened for recombination using GARD [63]. Recombination can contribute to false inference of positive selection, causing a high rate of false positive detection. No evidence of recombination was found.

2.4. Protein Modeling

To obtain the appropriate template for homology modeling, the putative tetraspanin sequences were threaded through protein homology/analogy recognition engine platform (PHYRE) v 2.0 [64]. Homology models of tetraspanins were constructed using the program MODELLER v. 9.13 [65]. An alignment of query sequences with template proteins was used as input for modeling MODELLER, and 100 comparative models were generated for each putative tetraspanin sequence. The models were subsequently validated using MODELLER objective function and DOPE score, statistical parameters for the assessment of the model was done using the standard Modeler energy function. The model quality and accuracy were stereochemically ascertained by using the protein structure and verification tool, PROCHECK v. 3.5.4 [66,67]. The quality of the modeled structures were also validated by other structure verification servers such as VERIFY 3D (to analyze the compatibility of the 3D structure with the amino acid sequences [68]), and ERRAT (to statistically analyze non-bonded interactions between different atoms, whereby higher scores are indicative of higher quality [69]). The tetraspanin domains were mined from the Conserved Domain Database [70].

3. Results and Discussion

Tetraspanins are an evolutionarily conserved protein superfamily that have been investigated for their potential functions as “molecular facilitators” of cell growth, motility, signal transduction, and host-pathogen interactions [15,71]. The present study was conceived from the exhaustive collation of the available data reporting the existence of tetraspanin superfamily members in various organisms whose genomes have been sequenced [72]. Of particular interest in this study is the possibility that given the important roles that tetraspanins play, their disruption would putatively jettison the parasite transmitting cycle of *G. morsitans*. These proteins could therefore be potential targets aimed at mitigating the vectorial ability of *G. morsitans* in the control of African trypanosomiasis. Several studies have demonstrated the critical roles of tetraspanins during parasite infection in insect vectors. For instance, Jaramillo-Gutierrez *et al.* [73] showed that a knockdown of tetraspanins enhanced *Plasmodium* parasite infection in anopheline mosquitoes, potentially via blockade of immune cascades mediated by these proteins. With the recent availability of the *G. morsitans* genome sequence [45], it is interesting to identify and analyze tetraspanins in this insect. Therefore, here, we identified, structurally characterized and analyzed the evolution of tetraspanins in *G. morsitans in silico*.

3.1. *Glossina morsitans* Putative Tetraspanins

We identified and characterized a repertoire of tetraspanins in the genome of *Glossina morsitans* based on their definitive structural characteristics of the tetraspanin family of proteins. The presence of TM domains in the putative tetraspanins was determined using SMART and TMHMM 2.0 while the definitive extracellular domain was identified using InterPro. For the purposes of clarity in this study, we abbreviated the identified *G. morsitans* tetraspanins as “GmTsp”, an abbreviation used hereafter. Notably, the identified *G. morsitans* tetraspanins contained one LEL domain at the C-terminus, and four TM domains (Figure 1 and Table 1). We further evaluated the putative GmTspS by functional annotation using Blas2GO. Our analysis for the presence of the structural hallmarks of the tetraspanin protein family (see introduction) indicated that the *G. morsitans* genome contained twenty four TspS (here, we do not show the data for all the twenty four TspS). However, more stringent analysis (concentrating on the putative GmTspS that fulfill the requirement for four hydrophobic helices, SEL, LEL, presence of palmitoylation sites, and cytosolic orientation of the N- and C-terminal of the proteins) revealed that only seventeen of the putative GmTspS qualified as members of the tetraspanin superfamily. We confirmed that the N- and the C-terminal of the seventeen putative GmTspS are cytosolic with reference to the plasma membrane. Although we have not shown the data for all the putative GmTspS, we have shown the orientation of a representative of GmTspS using GmTsp42Ed in comparison with the human HsTspCD63 (see discussion later in this paper). Since the cytosolic orientation of these termini is important in cellular signaling, this result implies that the putative GmTspS are highly likely to have similar functions reported for other members of the tetraspanin superfamily. The presence of the structural hallmarks of the GmTspS is shown in Figure A1. Of the seventeen GmTspS, six were reported in the sialome [74] and transcriptome of *G. morsitans* [45], and their protein names and UniProt IDs are indicated in Table 1. The amino acid sequences of the remaining eleven of the seventeen GmTspS did not match to any sequences and/or information in the available protein databases other than in the VectorBase. These proteins appear as novel tetraspanins in the genome of *G. morsitans* because they only matched to EST sequences on VectorBase and no gene function for these sequences has been so far determined or inferred. For clarity, we arbitrarily named (and abbreviated) these eleven new tetraspanins as GmTsp 1 to 11 (see Table 1). Notably, all the identified GmTspS identified in this study fell within the size range (200–350) amino acid residues [16]) of the various tetraspanins reported to date.

Alignment of *G. morsitans* tetraspanin amino acid sequences alongside homologs from the house fly, the fruit fly, and human, revealed that the GmTspS have the two highly conserved features of tetraspanins: (i) the LEL harboring the CCG motif; and (ii) the four TM domains with some well-conserved residues (Figure 2). Other features characteristic of tetraspanins, *i.e.*, SEL, ICL and C-terminal region rich in charged/polar amino acids were observed as evident in Figure 2. Further, in addition to the CCG motif, the other additional motifs, *i.e.*, PxxCC and EGC were also present in the GmTspS (Figure 2). Notably, of the seventeen putative GmTspS that we identified, 64.7% ($n = 11$) contained the additional PxxCC within the LEL domain (Figure A1), which is in agreement with the documented data that more the 50% of TspS contain this motif [24]. Taken together, these findings show that *G. morsitans* tetraspanins have all the structural hallmarks of the typical tetraspanins found in other insects and

humans, and suggest that the putative *G. morsitans* tetraspanins identified in this study could in general be classified as members of the tetraspanin superfamily of proteins.

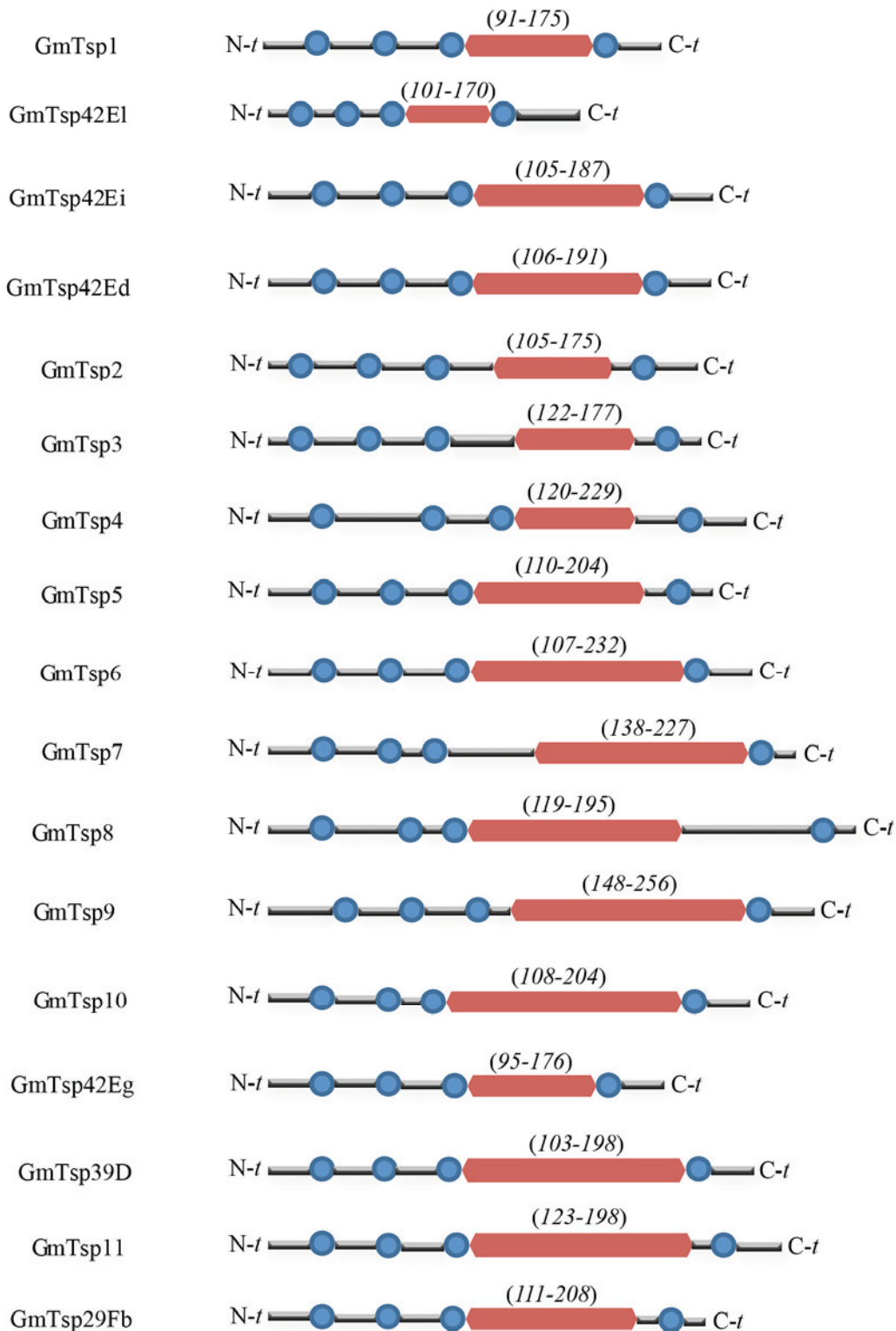
Table 1. Domain characteristics of seventeen GmTsp proteins in the genome of *G. morsitans*. The coordinates of the large extracellular loop (LEL) and the four transmembrane domains (TM) are shown in columns 4 and 5.

Protein Name (Description *)	UniProt/VectorBase ID	Length [aa]	LEL Domain Coordinates	TM Domain Coordinates
GmTsp1 (Protein late bloomer-like)	GMOY003645-PA [‡]	214	91–175	12–34, 41–63, 70–92, 181–203
GmTsp42E1	D3TMF8	209	110–170	13–32, 42–64, 69–88, 176–198
GmTsp 42Ei	D3TL43	228	105–187	13–35, 50–72, 79–101, 189–211
GmTsp42Ed	D3TMA1	229	106–191	12–34, 49–71, 84–106, 194–216
GmTsp2 (Tetraspanin EC2/Peripherin)	GMOY003647-PA [‡]	220	105–175	7–29, 44–66, 73–95, 187–209
GmTsp3 (Protein late bloomer-like)	GMOY003648-PA [‡]	220	122–177	7–28, 48–70, 75–92, 189–211
GmTsp4 (Tetraspanin, isoform A)	GMOY003747-PA [‡]	269	120–229	15–37, 66–88, 98–120, 235–257
GmTsp5 (CD151 antigen-like protein, isoform x2)	GMOY007608-PA [‡]	234	110–204	19–41, 56–78, 90–112, 208–230
GmTsp6 (Tetraspanin family integral membrane protein, isoform B)	D3TLU9	268	107–232	12–34, 54–76, 83–105, 233–255
GmTsp7 (Tetraspanin, isoform C)	GMOY010508-PA [‡]	283	138–227	13–35, 55–77, 84–106, 231–253
GmTsp8 (Tetraspanin-5-like, isoform x1)	GMOY004352-PA [‡]	327	119–195	21–43, 63–85, 92–114, 290–312
GmTsp9 (Tetraspanin-33-like, isoform x1)	GMOY011478-PA [‡]	295	148–256	39–61, 81–103, 115–137, 257–279
GmTsp10 (CD63 antigen-like protein)	GMOY000619-PA [‡]	241	108–204	13–35, 57–79, 86–108, 206–228
GmTsp42Eg	D3TQ76	218	95–176	13–35, 40–62, 75–97, 180–202
GmTsp39D	GMOY010261-PA [‡]	234	103–198	13–35, 48–70, 82–104, 201–223
GmTsp11 (Transmembrane 4 protein, isoform C)	GMOY009229-PA [‡]	285	123–198	12–35, 50–72, 85–107, 207–229
GmTsp29Fb	D3TQ22	303	111–208	20–42, 57–79, 91–113, 215–237

* The description of the proteins were performed using Blast2GO v 2.7.2 [50]; [‡] These proteins did not match to sequences and/or information in biological databases other than the VectorBase. The apparently new *G. morsitans* tetraspanins have been arbitrarily abbreviated as GmTsp 1 to 11.

Generally, protein palmitoylation is a special class of covalent post-translational modification which reversibly mediates various cellular processes. In particular, palmitoylation ensures not only surface hydrophobicity and membrane affinity, but also plays important roles in modulating the trafficking, stability, sorting, protein-protein and protein-lipid interactions of the modified proteins [75,76]. Various studies have demonstrated palmitoylation of tetraspanins on the cytoplasmic cysteine residues proximal to the plasma membranes contributes to the organization of these proteins into TEMs [77–79], underpinning the role of palmitoylation in functional integrity of tetraspanins. The results of palmitoylation prediction of GmTsp revealed that, as typical of tetraspanins [77–79], the predicted palmitoylation sites in the GmTsp involve multiple Cysteine residues adjacent to the borders between the cytoplasmic and TM domains (compare data presented in Figure A1 and Table 1). It should however be noted that the palmitoylation sites in the GmTsp shown in the Figure A1 are predicted, implying that some of these sites may not be truly palmitoylated, especially those that are predicted to be on the cysteine residues in the CCG and the PxxCC motifs.

Figure 1. Domain architecture of *Glossina morsitans* putative tetraspanins. Predicted domain structures of the seventeen *G. morsitans* are indicated. Transmembrane and large extracellular domain (LEL) domains are depicted as blue round circles and magenta rectangular blocks, respectively. The numbers in the parentheses indicate the coordinates (amino acid residues) of the LEL. For details of the coordinates of the TMs, compare this figure with data presented in Table 1 and in Figure A1. N-*t* and C-*t* represents the N- and C-termini of the GmTsp proteins, respectively.



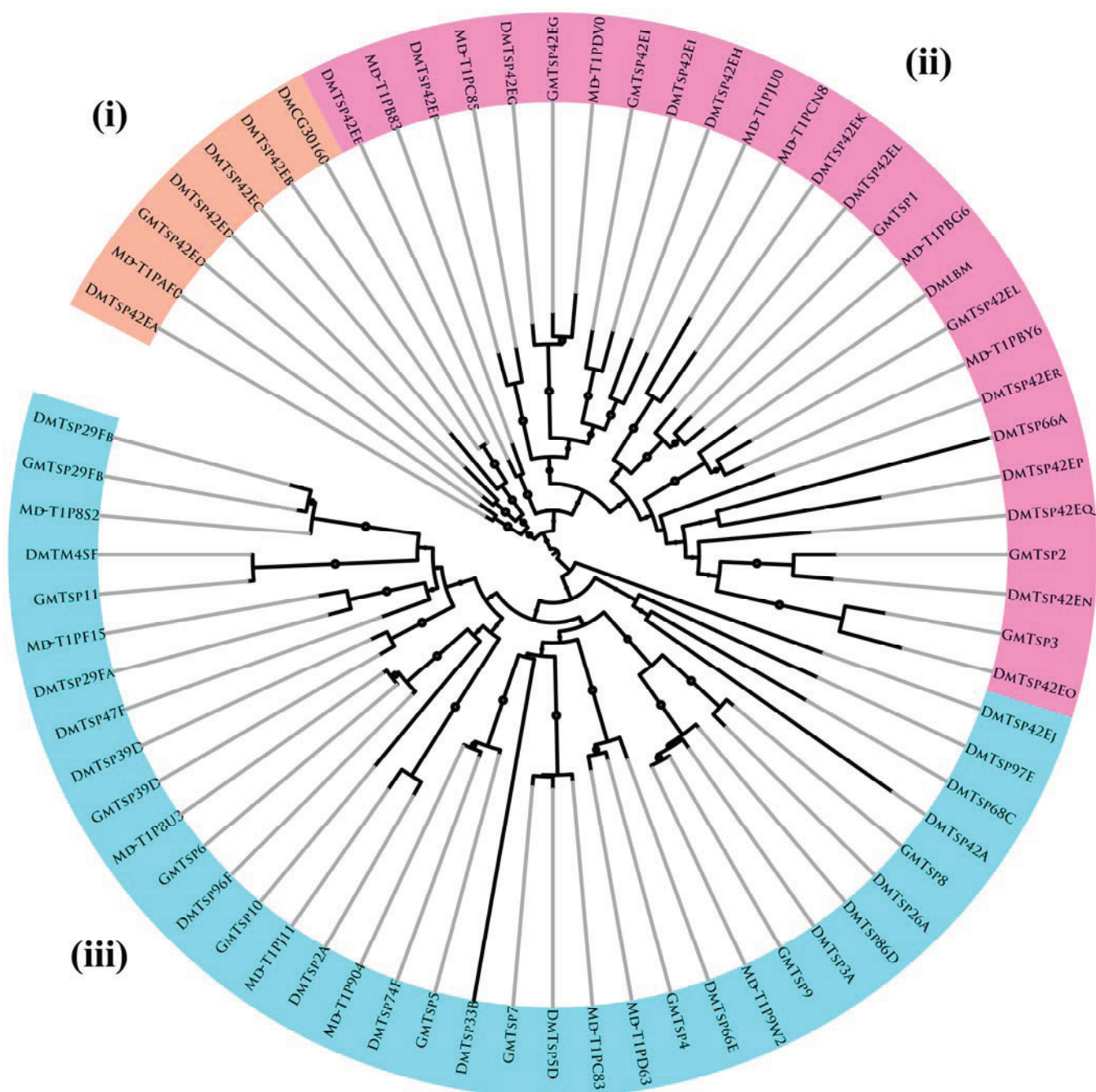
3.2. Evolutionary and Phylogenetic Analysis

To explore phylogenetic relationship among tetraspanins in different insect species, a phylogenetic tree was constructed including tetraspanins from *D. melanogaster* [56] and *M. domestica*. We initially included human (*H. sapiens*) tetraspanins in the evolutionary analysis to determine if they co-evolved with the insects' tetraspanins. Tetraspanins were originally identified in humans as putative mediators of tumor progression [80]. In this paper, we have not shown data on the analysis that included the human tetraspanins. The unrooted phylogenetic tree (Figure 3) was generated by the alignment of full-length protein sequences of 76 tetraspanin proteins and was inferred using both Bayesian inference and maximum likelihood (ML) methods. Tree nodes supported by high posterior probability and bootstrap values for the two methods were considered robust. A list of all amino acid sequences used for phylogenetic analysis is provided in Supplemental File S1.

As shown in Figure 3, three distinct clusters were observed: (i) A seven-member cluster consisting of one from *G. morsitans* (GmTsp42Ed), three from *D. melanogaster* (DmTsp42Ea, DmTsp42Ed, DmTsp42Ec, DmTsp42Eb and DmCG30160) and one from *M. domestica* (Md-T1PAF0); (ii) A twenty seven-member cluster with six *G. morsitans* (GmTsp42Eg, GmTsp42Ei, GmTsp1, GmTsp42EI, GmTsp2 and GmTsp3), fourteen *D. melanogaster* (DmTsp42Eo, DmTsp42En, DmTsp42Eq, DmTsp42Ep, DmTsp66A, DmTsp42Er, Dmlbm, DmTsp42EI, DmTsp42Ek, DmTsp42Eh, DmTsp42Ei, DmTsp42Eg, DmTsp42Ef and DmTsp42Ee) and seven *M. domestica* (Md-TIPB83, Md-TIPC85, Md-TIPDV0, Md-TIPJU0, Md-TIPCN8, Md-TIPBG6 and Md-TIPBY6); (iii) A thirty five-member cluster made up of nine *G. morsitans* tetraspanins (GmTsp11, GmTsp39D, GmTsp6, GmTsp10, GmTsp5, GmTsp7, GmTsp4, GmTsp9 and GmTsp8), eighteen *D. melanogaster* (DmTsp29Fb, DmTM4SF, DmTsp29Fa, DmTsp47F, DmTsp39D, DmTsp96F, DmTsp2A, DmTsp74F, DmTsp33B, DmTsp5D, DmTsp66E, DmTsp3A, DmTsp86D, DmTsp26A, DmTsp42A, DmTsp68C, DmTsp97E and DmTsp42Ej) and eight *M. domestica* tetraspanins (Md-T1P8S2, Md-T1PF15, Md-T1P8U3, Md-T1PJ11, Md-T1P904, Md-T1PC83, Md-T1PD63 and Md-T1P9W2).

In general, the clustering pattern observed in Figure 3 is indicative of gene duplications, gene loss and species-specific expansion in the various clades, as evidenced by the sub-clades in the three major clades. Further, within the three main clades, there are nine sub-clades displaying 1:1:1 orthologous relationship. The conservation of tetraspanins in the three insect species points to a probable critical role for the gene. Only a few clades show evidence of putative species-specific expansion. Overall, in most (sub-) clades there is evidence of gene loss. Several orthologous relationships among the putative insect tetraspanins are evident. The wide presence of tetraspanins in almost all organisms indicates that they have experienced a long evolutionary history [82]. Based on the phylogenetic relationships and ancestral origin of tetraspanins, it is possible that they have evolved from a single or a few ancestral genes by gene duplication and divergence and this evolution is impacted by gene loss and positive selection on coding sequence [16,83].

Figure 3. Phylogenetic analysis of tetraspanin homologs based on amino acid sequences: The phylogenetic tree shows clustering of the TspS into three main clades and nine sub-clades, which display 1:1:1 orthologous relationships of these proteins. The phylogenetic tree was constructed using the ML method implemented in PhyML [81] using the LG model of amino acid substitution [54]. The seventeen putative tsetse fly (*G. morsitans*) tetraspanins identified in this study (see Table 1) were phylogenetically compared to homologs in the fruit fly (*D. melanogaster*) and housefly (*M. domestica*). Nodes with bootstrap support values >80% are marked with solid circles. Gm, *Glossina morsitans*; Dm, *Drosophila melanogaster*; Md, *Musca domestica*. **Note:** for the abbreviation of the TspS in this figure, all letters are in upper case; the letters written in lower case in the main text are shown here in decreased font size.



3.3. Structural Analysis

Of the seventeen *G. morsitans* putative tetraspanins identified in this study, we selected one, Tetraspanin 42Ed (GmTsp42Ed; UniProt ID # D3TMA1) for 3-D structural modeling. The main reason for selection of GmTsp42Ed was because, in addition to being detected at the transcript level [45,74], there is evidence that this protein is also expressed in the proteome of *Trypanosoma brucei*-infected *G. morsitans* salivary glands [84], indicative of a possible important role for this tetraspanin in *G. morsitans* vectorial ability. The modeling of GmTsp42Ed was performed using multiple templates (1G8Q, 4JKV, and 2M7Z) as identified in the threading sequence-structure engine PHYRE. Comparative homology modeling by satisfaction of spatial restraints using MODELLER resulted in a cylindrical and very compact 3-D model (Figure 4). The model has seven α -helices and two anti-parallel β -strands. Four helices (color coded in Figure 4) form the hydrophobic transmembrane domains while the remaining three are found in the extracellular LEL domain. Two anti-parallel β -strands are inserted between TM3 and the LEL helices resulting in a break in the continuity of the helical conformation. The anti-parallel β -strands probably form the non-conserved sub-domain of the LEL domain [85]. Both N- and C-terminal regions of GmTsp42Ed adopt somewhat ordered conformations. For the N-terminus, this is perhaps due to the palmitoylation of a cysteine 3 residue in this region. To further confirm whether the N- and C-termini of the GmTsp42Ed are oriented to the cytosolic side of the plasma membrane, we compared the amino acid sequence of this Tsp with the human HsTspCD63. The result of this comparative analysis revealed that, similar to the human homolog, the two termini of GmTsp42Ed are indeed cytosolic, and the degree of similarity between the two proteins was very high (Figure 4B,C).

We stereochemically ascertained the quality and accuracy of the GmTsp42Ed using the PROCHECK protein structure validation and verification package. The Ramachandran plot obtained showed that 87.8%, 7.3%, 2.9%, and 2.0% of residues fell within the most favored regions, additionally favored regions, generously allowed regions and the disallowed regions, respectively (Figure 5). This makes a combined percentage of 98.0% of the residues in the favored and allowed regions for the model making it stereochemically robust [86]. Again, using GmTsp42Ed as a representative of *G. morsitans* tetraspanins, this result appears to confirm that in general, the GmTsp belong to the tetraspanin superfamily.

3.4. Positive Selection Analysis of *G. morsitans* Tetraspanins

Mutation and selection have different effects on nonsynonymous (amino-acid-replacement) and synonymous (silent) substitution rates (d_N and d_S , respectively), and are therefore means of understanding the dynamics of molecular sequence evolution [87]. Models of variable d_N/d_S ratios among sites provide important insights into the functional constraints at different amino acid sites, and are to detect sites under positive selection [88]. A $d_N/d_S \gg 1.0$ is considered as a convincing indicator of positive selection [89]. Analysis of selection pressures exerted on tetraspanins in *G. morsitans*, *D. melanogaster* and *M. domestica* revealed that several sites are under positive selection based on statistical significance tests as assessed by various models such as MEME, FEL, IFEL, REL and FUBAR (See Supplemental File S2). Unlike the other methods, MEME methodology can identify both

episodic and persistent positive selection because it allows the distribution of the dN/dS ratio to vary from site to site and also from branch to branch at a site. Therefore, the additional positively selected codons identified by MEME and not by the other approaches, are likely to have been subject to episodes of positive selection. Overall, these results suggest that positive selection is an important contributor in the evolution of tetraspanins.

Figure 4. Cartoon representation of the modeled GmTsp42Ed structure: (A) 3-D representation of GmTsp42Ed model, in which the N- and the C-termini are shown as NH2 and COOH, respectively. The four transmembrane domains are color-coded as red, purple, orange and yellow for TM1 to TM4, respectively. The LEL domain which is located between TM3 and TM4 has three α -helices and two anti-parallel β -strands and is shown in cyan; (B,C) 2-D representation of the orientation of the cytoplasmic (shown in blue), non-cytoplasmic (small extracellular loop (SEL) and LEL domains) (shown in green), and TM1-4 helices (shown in yellow) of *G. morsitans* GmTsp42Ed and the human tetraspanin (HsTspCD63), respectively. The alignments of GmTsp42Ed and HsTspCD63 are shown at the bottom of the domain orientations. The dotted lines in Panels (B) and (C) represent a hypothetical cellular plasma membrane.

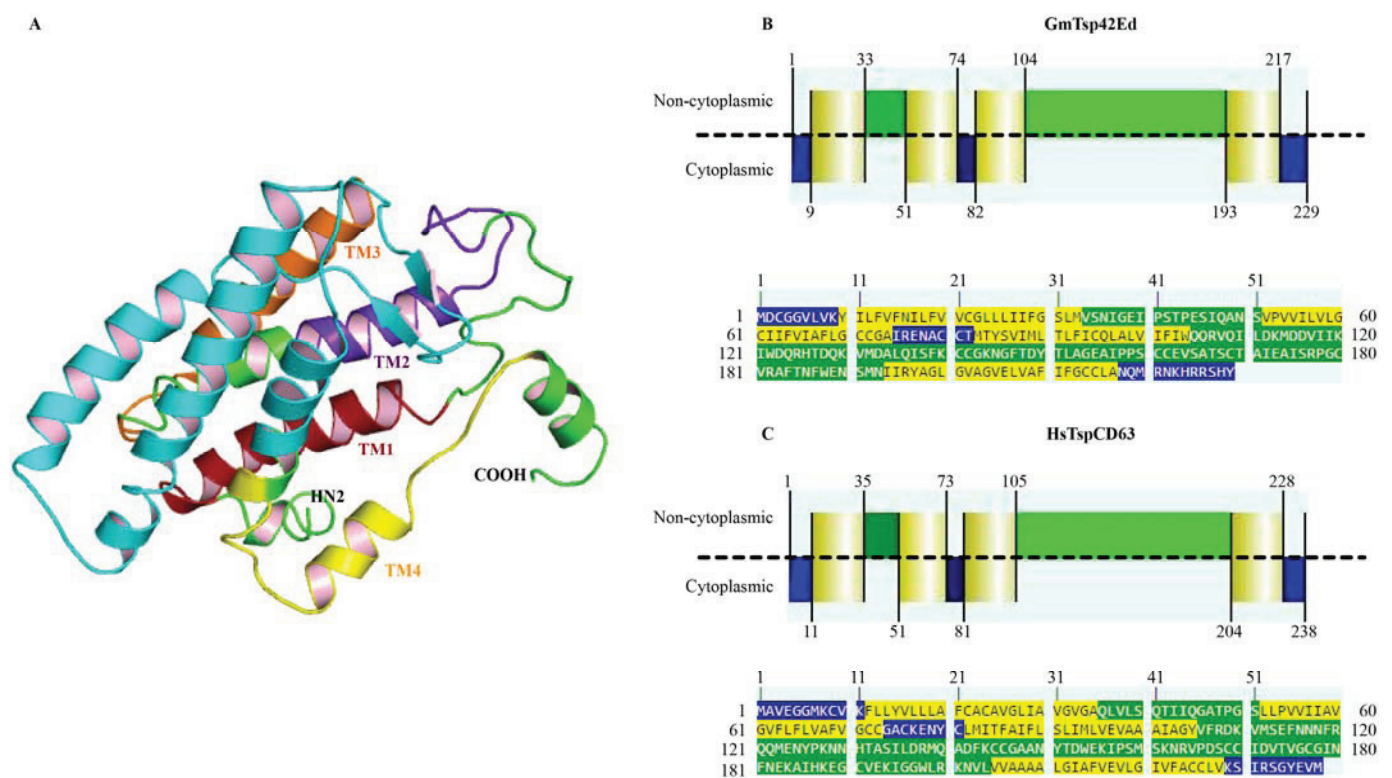
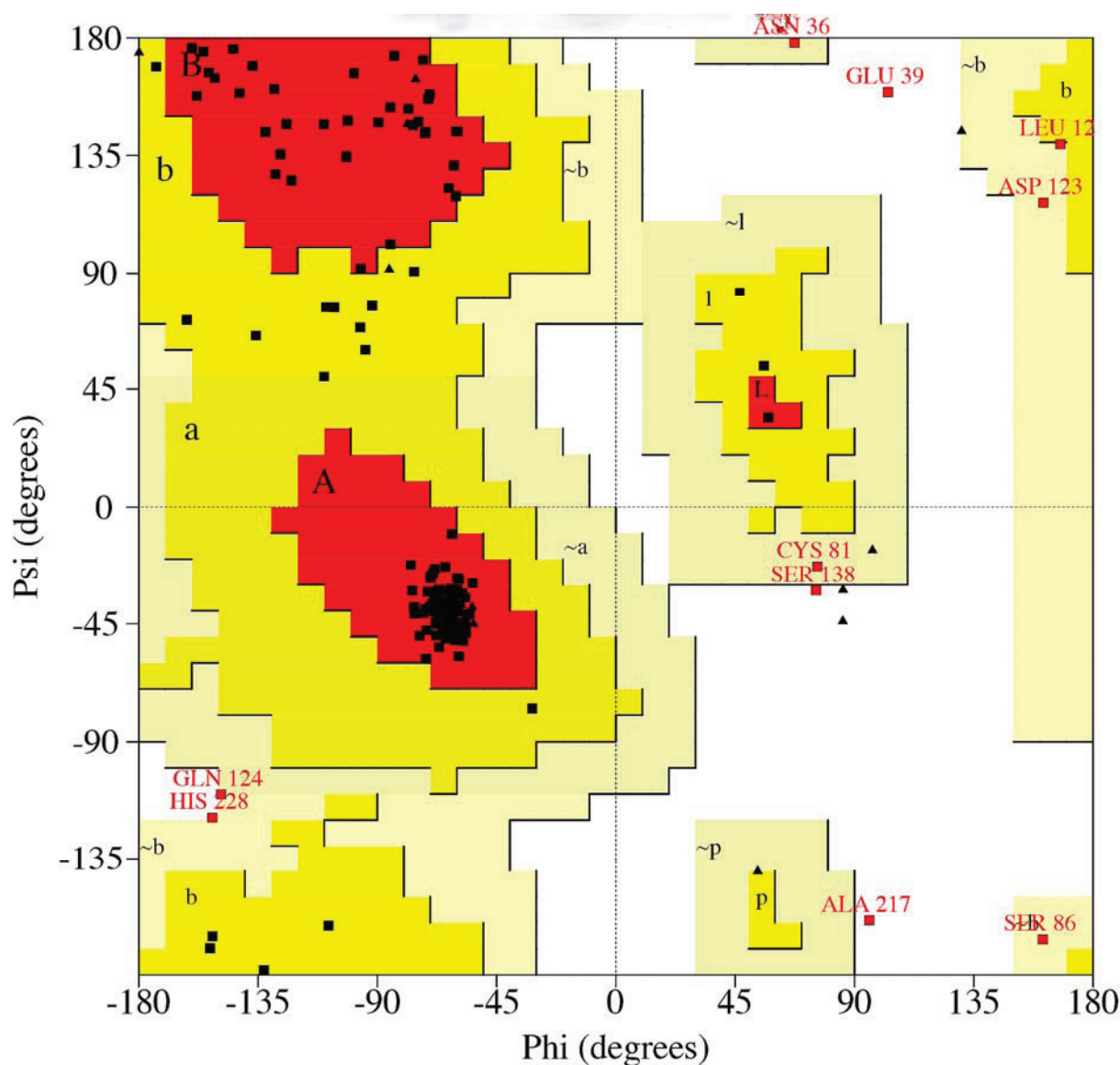


Figure 5. Ramachandran plots for GmTsp42Ed: The plot statistics are indicated for all the non-glycine and non-proline residues that fell within the favored regions. Shown in red, yellow, pale yellow, and white, respectively, are amino acid residues in most favored, additionally allowed, generously allowed, and disallowed regions. Based on the analysis of 118 structures of resolution of at least 2.0 Angstroms and R-factor no greater than 20%, a good model is expected to have $\geq 90\%$ in the most favored regions. The plots were generated using PROCHECK v. 3.5.4 [66].



Plot statistics

Residues in most favoured regions [A,B,L]	180	87.8%
Residues in additional allowed regions [a,b,l,p]	15	7.3%
Residues in generously allowed regions [~a,~b,~l,~p]	6	2.9%
Residues in disallowed regions	4	2.0%

Number of non-glycine and non-proline residues	205	100.0%
Number of end-residues (excl. Gly and Pro)	2	
Number of glycine residues (shown as triangles)	16	
Number of proline residues	6	

Total number of residues	229	

3.5. Targeting Tetraspanins as Potential Therapeutics against African Trypanosomes

Due to a lack or failure of conventional vaccines and drugs to combat insect-transmitted pathogens, the identification of novel therapeutic targets against infections is under intensive research. For African trypanosomiasis, there is need to address the problem of drug resistance and/or counterfeits [12,13], and a lack of efficacious vaccines [5,6] against this group of NTDs. An approach to this problem is to target cellular processes of the vector rather than targeting the *Trypanosoma* parasites. As described in Section 1 of this article, recent data have indicated that members of the tetraspanin superfamily have potential to provide such a novel approach. Specific tetraspanins family members have been shown to play roles in pathogen infections by selectively associating with the pathogens at multiple infection stages from the initial cellular attachment to the egress of mature pathogens (See Table 1 in Ref. [90] for tetraspanin family members with reported links to pathologies). Although the precise roles of TEMs during pathogen infections have not been extensively investigated, several studies have demonstrated that tetraspanin-based drugs (mimicry) can disrupt normal biological functions. For instance, inhibition of the binding of hepatitis C virus envelop glycoprotein EC to its receptor, CD81 [91]. Further, Spoden *et al.* [92] provided evidence that tetraspanin-specific antibodies and siRNAs inhibited both the cell entry and subsequent infection of the human papilloma virus type 16. Similar finding have been reported in the case of the infection of human immunodeficiency virus [93–95]. In addition to targeting host-encoded tetraspanins, other independent studies have explored the application of antibodies against recombinant TSPs from parasites. For instance, Tran *et al.* [96] demonstrated that recombinant antibodies to SmTSP-1/2 (cloned from the trematode *Schistosoma mansoni*) significantly reduced the parasite loads after challenge with *S. mansoni*. Silencing of *Sm-tsp-1* or *Sm-tsp-1* resulted in malformation of the parasite tegument and tetraspanins-depleted parasites were found to be defective in survival in the host [97], implying that Tsps have integral structural roles in the development and maintenance of the parasite's tegument. Similar to their application in the control of schistosomiasis, Tsps have the potential as vaccine candidates against filariasis. Although not yet fully worked out, Dakshinamoorthy *et al.* (2013) have recently demonstrated that antibodies to BmTSP LEL and WbTSP LEL (cloned from the nematodes *Brugia malayi* and *Wuchereria bancrofti*, respectively) conferred significant protection to mice that were challenged with filarial worms [98]. Future research in the case of *Trypanosoma* parasites should focus on the analysis of tetraspanins expressed by these parasites to explore their potential as vaccine candidates. Taken together, although we did not analyze *Trypanosoma*-specific tetraspanins in this article, the data presented in our study provide proof of principle that targeting tetraspanins, either vector- or parasite-encoded, is a promising strategy to inhibit specific stages of pathogen infection.

4. Conclusions

This work identified seventeen putative tetraspanins in the genome of the tsetse fly, *G. morsitans*, eleven of which appear to be novel as their proteins sequences did not match to any tetraspanin superfamily member either determined or inferred, rather, their sequences only matched to EST sequences in the VectorBase. The presence of the hallmark tetraspanin features, including the LEL domain harboring the CCG/PxxCC motifs, four TM domains with some well-conserved residues, the

SEL, ICL and C-terminal region rich in charged/polar amino acids suggest that the putative *G. morsitans* tetraspanins identified in this study are *bona fide* members of the tetraspanin superfamily of proteins. However, there is need to experimentally validate the functionality of these tetraspanins in *Glossina*. Nevertheless, the results presented here constitute a platform for the expansion of future exploration into the biological roles of *G. morsitans* tetraspanins, and their potential as candidates for anti-*Trypanosoma* mitigation strategies. This is based on the fact that tetraspanins are not merely passive targets for pathogen infection, but appear to have more fundamental roles during initial attachment (cell-to-cell fusion) to host cells, cellular trafficking, and egress. This is a subject of our immediate further investigations for the identification and characterization of *Trypanosoma*-encoded tetraspanins as these are potential targets for the development of vaccines against the parasite.

Acknowledgments

This work was supported by South African National Bioinformatics Institute, Egerton University, Wageningen University, and Kenya Agricultural and Livestock Research Organization.

Author Contributions

Edwin K. Murungi and Henry M. Kariithi designed the research, performed the experiments, analyzed data and wrote the paper. Vincent Adunga, Meshack Obonyo and Alan Christoffels contributed to writing the paper, critical comments and suggestions. All the authors read and approved the final manuscript.

Appendix

Figure A1. Distribution of the structural features of seventeen *G. morsitans* putative tetraspanins detected in this study: The start and end of the domains/motifs are indicated by the coordinates of the amino acid residues (shown with black numbers with vertical bars). The palmitoylation sites are shown in blue bold letters; the transmembrane domains (TM1–4) are depicted in light blue; the SEL motifs are in pink; the CCG domains are shown in red; the LEL motifs are shown in light green, while the PxxCC motifs are in dark blue. Notice that six of the GmTsp1s do not contain the PxxCC motif.

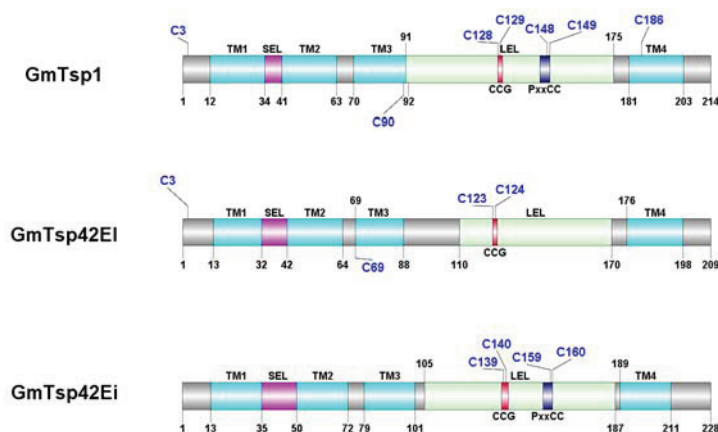
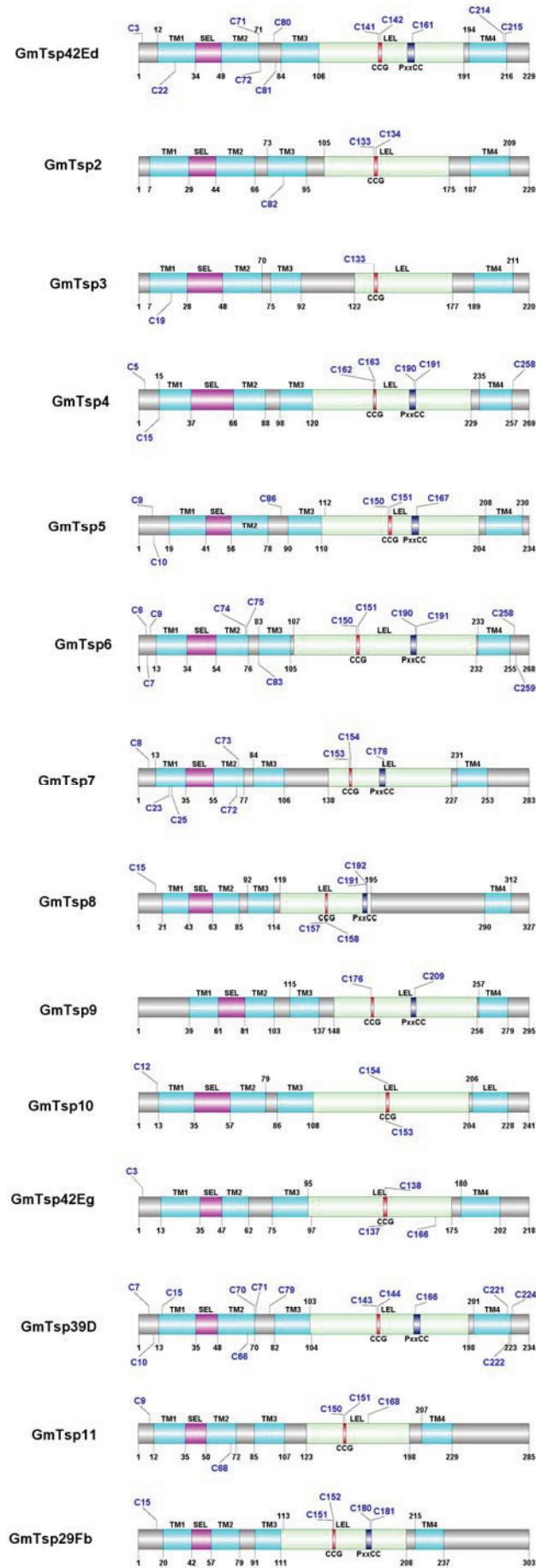


Figure A1. Cont.



Conflicts of Interest

The authors declare no conflict of interest.

References

1. Maudlin, I. Transmission of African trypanosomiasis: Interactions among tsetse immune system, symbionts, and parasites. In *Advances in Disease Vector Research*, 7th ed.; Harris, K., Ed.; Springer: New York, NY, USA, 1991; pp. 117–148.
2. Vreysen, M.J.B. Prospects for area-wide integrated control of tsetse flies (Diptera: Glossinidae) and trypanosomiasis in sub-Saharan Africa. *Rev. Soc. Entomol. Argent.* **2006**, *65*, 1–21.
3. Brun, R.; Blum, J.; Chappuis, F.; Burri, C. Human African trypanosomiasis. *Lancet* **2010**, *375*, 148–159.
4. Chappuis, F.; Loutan, L.; Simarro, P.; Lejon, V.; Büscher, P. Options for field diagnosis of human African trypanosomiasis. *Clin. Microbiol. Rev.* **2005**, *18*, 133–146.
5. Burri, C.; Nkunku, S.; Merolle, A.; Smith, T.; Blum, J.; Brun, R. Efficacy of new, concise schedule for melarsoprol in treatment of sleeping sickness caused by *Trypanosoma brucei gambiense*: A randomised trial. *Lancet* **2000**, *355*, 1419–1425.
6. Burri, C. Chemotherapy against human African trypanosomiasis: Is there a road to success? *Parasitology* **2010**, *137*, 1987–1994.
7. Matovu, E.; Seebeck, T.; Enyaru, J.C.; Kaminsky, R. Drug resistance in *Trypanosoma brucei* spp., the causative agents of sleeping sickness in man and nagana in cattle. *Microbes Infect.* **2001**, *3*, 763–770.
8. Brightwell, B.; Dransfield, R. Odour attractants for tsetse: *Glossina austeni*, *G. brevipalpis* and *G. swynnertoni*. *Med. Vet. Entomol.* **1997**, *11*, 297–299.
9. Allsopp, R. Options for vector control against trypanosomiasis in Africa. *Trends Parasitol.* **2001**, *17*, 15–19.
10. Rowlands, G.J.; Leak, S.G.A.; Mulatu, W.; Nagda, S.M.; Wilson, A.; D'Ieteren, G.D.M. Use of deltamethrin “pour-on” insecticide for the control of cattle trypanosomiasis in the presence of high tsetse invasion. *Med. Vet. Entomol.* **2001**, *15*, 87–96.
11. Vreysen, M.J.; Saleh, K.M.; Ali, M.Y.; Abdulla, A.M.; Zhu, Z.R.; Juma, K.G.; Dyck, V.A.; Msangi, A.R.; Mkonyi, P.A.; Feldmann, H.U. *Glossina austeni* (Diptera: Glossinidae) eradicated on the island of Unguja, Zanzibar, using the sterile insect technique. *J. Econ. Entomol.* **2000**, *93*, 123–135.
12. Barrett, M.P.; Vincent, I.M.; Burchmore, R.J.; Kazibwe, A.J.; Matovu, E. Drug resistance in human African trypanosomiasis. *Future Microbiol.* **2011**, *6*, 1037–1047.
13. Geerts, S.; Holmes, P.H.; Eisler, M.C.; Dially, O. African bovine trypanosomiasis: The problem of drug resistance. *Trends Parasitol.* **2001**, *17*, 25–28.
14. Du Toit, R. Trypanosomiasis in Zululand and the control of tsetse flies by chemical means. *Onderstepoort. J. Vet. Res.* **1954**, *26*, 317–385.
15. Hemler, M.E. Tetraspanin proteins mediate cellular penetration, invasion, and fusion events and define a novel type of membrane microdomain. *Annu. Rev. Cell Dev. Biol.* **2003**, *19*, 397–422.

16. Huang, S.; Tian, H.; Chen, Z.; Yu, T.; Xu, A. The evolution of vertebrate tetraspanins: gene loss, retention, and massive positive selection after whole genome duplications. *BMC Evol. Biol.* **2010**, *10*, e306.
17. Monk, P.N.; Partridge, L.J. Tetraspanins: Gateways for infection. *Infect. Disord. Drug. Targets.* **2012**, *12*, 4–17.
18. Yunta, M.; Oliva, J.L.; Barcia, R.; Horejsi, V.; Angelisova, P.; Lazo, P.A. Transient activation of the c-Jun N-terminal kinase (JNK) activity by ligation of the tetraspan CD53 antigen in different cell types. *Eur. J. Biochem.* **2002**, *269*, 1012–1021.
19. Patnaik, B.B.; Kang, S.M.; Seo, G.W.; Lee, H.J.; Patnaik, H.H.; Jo, Y.H.; Tindwa, H.; Lee, Y.S.; Lee, B.L.; Kim, N.J.; *et al.* Molecular cloning, sequence characterization and expression analysis of a CD63 homologue from the coleopteran beetle, *Tenebrio molitor*. *Int. J. Mol. Sci.* **2013**, *14*, 20744–20767.
20. Boucheix, C.; Rubinstein, E. Tetraspanins. *Cell Mol. Life. Sci.* **2001**, *58*, 1189–1205.
21. Stipp, C.S.; Kolesnikova, T.V.; Hemler, M.E. Functional domains in tetraspanin proteins. *Trends Biochem. Sci.* **2003**, *28*, 106–112.
22. Kovalenko, O.V.; Metcalf, D.G.; DeGrado, W.F.; Hemler, M.E. Structural organization and interactions of transmembrane domains in tetraspanin proteins. *BMC Struct. Biol.* **2005**, *5*, e11.
23. Charrin, S.; Jouannet, S.; Boucheix, C.; Rubinstein, E. Tetraspanins at a glance. *J. Cell Sci.* **2014**, *127*, 3641–3648.
24. Berditchevski, F. Complexes of tetraspanins with integrins: More than meets the eye. *J. Cell Sci.* **2001**, *114*, 4143–4151.
25. Kovalenko, O.V.; Yang, X.; Kolesnikova, T.V.; Hemler, M.E. Evidence for specific tetraspanin homodimers: Inhibition of palmitoylation makes cysteine residues available for cross-linking. *Biochem. J.* **2004**, *377*, 407–417.
26. Fradkin, L.G.; Kamphorst, J.T.; DiAntonio, A.; Goodman, C.S.; Noordermeer, J.N. Genome-wide analysis of the *Drosophila* tetraspanins reveals a subset with similar function in the formation of the embryonic synapse. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 13663–13668.
27. Charrin, S.; Le, N.F.; Oualid, M.; Billard, M.; Faure, G.; Hanash, S.M.; Boucheix, C.; Rubinstein, E. The major CD9 and CD81 molecular partner: identification and characterization of the complexes. *J. Biol. Chem.* **2001**, *276*, 14329–14337.
28. Hemler, M.E. Tetraspanin functions and associated microdomains. *Nat. Rev. Mol. Cell Biol.* **2005**, *6*, 801–811.
29. Hemler, M.E. Integrin associated proteins. *Curr. Opin. Cell Biol.* **1998**, *10*, 578–585.
30. Stipp, C.S.; Orlicky, D.; Hemler, M.E. FPRP, a major, highly stoichiometric, highly specific. *J. Biol. Chem.* **2001**, *276*, 4853–4862.
31. Stipp, C.S.; Kolesnikova, T.V.; Hemler, M.E. EWI-2 is a major CD9 and CD81 partner and member of a novel Ig protein subfamily. *J. Biol. Chem.* **2001**, *276*, 40545–40554.
32. Levy, S.; Shoham, T. Protein-protein interactions in the tetraspanin web. *Physiology (Bethesda)* **2005**, *20*, 218–224.

33. Lagaudrière-Gesbert, C.; Naour, F.L.; Lebel-Binay, S.; Billard, M.; Lemichez, E.; Boquet, P.; Boucheix, C.; Conjeaud, H.; Rubinstein, E. Functional analysis of four tetraspans, CD9, CD53, CD81, and CD82, suggests a common role in costimulation, cell adhesion, and migration: Only CD9 upregulates HB-EGF activity. *Cell. Immunol.* **1997**, *182*, 105–112.
34. Zhang, F.; Kotha, J.; Jennings, L.K.; Zhang, X.A. Tetraspanins and vascular functions. *Cardiovasc. Res.* **2009**, *83*, 7–15.
35. Charrin, S.; Le, N.F.; Silvie, O.; Milhiet, P.E.; Boucheix, C.; Rubinstein, E. Lateral organization of membrane proteins: Tetraspanins spin their web. *Biochem. J.* **2009**, *420*, 133–154.
36. Tham, T.N.; Gouin, E.; Rubinstein, E.; Boucheix, C.; Cossart, P.; Pizarro-Cerda, J. Tetraspanin CD81 is required for *Listeria monocytogenes* invasion. *Infect. Immun.* **2010**, *78*, 204–209.
37. Silvie, O.; Greco, C.; Franetich, J.F.; Dubart-Kupperschmitt, A.; Hannoun, L.; van Gemert, G.J.; Sauerwein, R.W.; Levy, S.; Boucheix, C.; Rubinstein, E.; *et al.* Expression of human CD81 differently affects host cell susceptibility to malaria sporozoites depending on the *Plasmodium* species. *Cell Microbiol.* **2006**, *8*, 1134–1146.
38. Luo, S.; He, M.; Cao, Y.; Xia, Y. The tetraspanin gene *MaPls1* contributes to virulence by affecting germination, appressorial function and enzymes for cuticle degradation in the entomopathogenic fungus, *Metarhizium acridum*. *Environ. Microbiol.* **2013**, *15*, 2966–2979.
39. Hassuna, N.; Monk, P.N.; Moseley, G.W.; Partridge, L.J. Strategies for targeting tetraspanin proteins: Potential therapeutic applications in microbial infections. *BioDrugs* **2009**, *23*, 341–359.
40. Van Spriel, A.B.; Figdor, C.G. The role of tetraspanins in the pathogenesis of infectious diseases. *Microbes Infect.* **2010**, *12*, 106–112.
41. The UniProt Consortium. The Universal Protein Resource (UniProt) 2009. *Nucleic Acids Res.* **2009**, *37*, D169–D174.
42. Megy, K.; Emrich, S.J.; Lawson, D.; Campbell, D.; Dialynas, E.; Hughes, D.S.T.; Koscielny, G.; Louis, C.; MacCallum, R.M.; Redmond, S.N.; *et al.* VectorBase: improvements to a bioinformatics resource for invertebrate vector genomics. *Nucleic Acids Res.* **2012**, *40*, D729–D734.
43. Altschul, S.F.; Madden, T.L.; Schäffer, A.A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D.J. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **1997**, *25*, 3389–3402.
44. Hunter, S.; Apweiler, R.; Attwood, T.K.; Bairoch, A.; Bateman, A.; Binns, D.; Bork, P.; Das, U.; Daugherty, L.; Duquenne, L.; *et al.* InterPro: The integrative protein signature database. *Nucleic Acids Res.* **2009**, *37*, D211–D215.
45. International Glossina Genome Initiative. Genome sequence of the tsetse fly (*Glossina morsitans*): Vector of African trypanosomiasis. *Science* **2014**, *344*, 380–386.
46. Schultz, J.; Milpetz, F.; Bork, P.; Ponting, C.P. SMART, a simple modular architecture research tool: identification of signaling domains. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 5857–5864.
47. Krogh, A.; Larsson, B.; von Heijne, G.; Sonnhammer, E.L.L. Predicting transmembrane protein topology with a hidden markov model: Application to complete genomes. *J. Mol. Biol.* **2001**, *305*, 567–580.

48. Hirokawa, T.; Boon-Chieng, S.; Mitaku, S. SOSUI: Classification and secondary structure prediction system for membrane proteins. *Bioinformatics* **1998**, *14*, 378–379.
49. Gasteiger, E.; Hoogland, C.; Gattiker, A.; Duvaud, S.; Wilkins, M.R.; Appel, R.D.; Bairoch, A. Protein identification and analysis tools on the ExPASy server. In *The proteomics protocols handbook*, 1st ed.; Walker, J.M., Ed.; Humana Press: Totowa, NJ, USA, 2005; pp. 571–607.
50. Conesa, A.; Gotz, S.; Garcia-Gomez, J.M.; Terol, J.; Talon, M.; Robles, M. Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **2005**, *21*, 3674–3676.
51. Ren, J.; Wen, L.; Gao, X.; Jin, C.; Xue, Y.; Yao, X. CSS-Palm 2.0: An updated software for palmitoylation sites prediction. *Prot. Eng. Des. Sel.* **2008**, *21*, 639–644.
52. Edgar, R.C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic. Acids Res.* **2004**, *32*, 1792–1797.
53. Waterhouse, A.M.; Procter, J.B.; Martin, D.M.; Clamp, M.; Barton, G.J. Jalview Version 2: A multiple sequence alignment editor and analysis workbench. *Bioinformatics* **2009**, *25*, 1189–1191.
54. Whelan, S.; Goldman, N. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol. Biol. Evol.* **2001**, *18*, 691–699.
55. Darriba, D.; Taboada, G.L.; Doallo, R.; Posada, D. Prot-Test 3: Fast selection of best-fit models of protein evolution. *Bioinformatics* **2011**, *27*, 1164–1165.
56. Guindon, S.; Delsuc, F.; Dufayard, J.F.; Gascuel, O. Estimating maximum likelihood phylogenies with PhyML. *Methods Mol. Biol.* **2009**, *537*, 113–137.
57. Ronquist, F.; Teslenko, M.; van der Mark, P.; Ayres, D.L.; Darling, A.; Höhna, S.; Larget, B.; Liu, L.; Suchard, M.A.; Huelsenbeck, J.P. MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* **2012**, *61*, 539–542.
58. Letunic, I.; Bork, P. Interactive Tree Of Life (iTOL): An online tool for phylogenetic tree display and annotation. *Bioinformatics* **2007**, *23*, 127–128.
59. Delpont, W.; Poon, A.F.; Frost, S.D.; Kosakovsky Pond, S.L. Datamonkey 2010: A suite of phylogenetic analysis tools for evolutionary biology. *Bioinformatics* **2010**, *26*, 2455–2457.
60. Kosakovsky Pond, S.L.; Frost, S.D.W. Not so different after all: A comparison of methods for detecting amino acid sites under selection. *Mol. Biol. Evol.* **2005**, *22*, 1208–1222.
61. Murrell, B.; Wertheim, J.O.; Moola, S.; Weighill, T.; Scheffler, K.; Kosakovsky Pond, S.L. Detecting individual sites subject to episodic diversifying selection. *PLoS Genet.* **2012**, *8*, e1002764.
62. Murrell, B.; Moola, S.; Mabona, A.; Weighill, T.; Sheward, D.; Kosakovsky Pond, S.L.; Scheffler, K. FUBAR: A fast, unconstrained Bayesian AppRoximation for inferring selection. *Mol. Biol. Evol.* **2013**, *30*, 1196–1205.
63. Kosakovsky Pond, S.L.; Posada, D.; Gravenor, M.B.; Woelk, C.H.; Frost, S.D.W. GARD: A genetic algorithm for recombination detection. *Bioinformatics* **2006**, *22*, 3096–3098.
64. Kelley, L.A.; Sternberg, M.J. Protein structure prediction on the Web: A case study using the Phyre server. *Nat. Protoc.* **2009**, *4*, 363–371.

65. Sali, A.; Blundell, T.L. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **1993**, *234*, 779–815.
66. Laskowski, R.A.; MacArthur, M.W.; Moss, D.S.; Thornton, J.M. PROCHECK: A program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* **1993**, *26*, 283–291.
67. Morris, A.L.; MacArthur, M.W.; Hutchinson, E.G.; Thornton, J.M. Stereochemical quality of protein structure coordinates. *Proteins* **1992**, *12*, 345–364.
68. Eisenberg, D.; Luthy, R.; Bowie, J.U. VERIFY3D: Assessment of protein models with three-dimensional profiles. *Methods. Enzymol.* **1997**, *277*, 396–404.
69. Colovos, C.; Yeates, T.O. Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Sci.* **1993**, *2*, 1511–1519.
70. Marchler-Bauer, A.; Zheng, C.; Chitsaz, F.; Derbyshire, M.K.; Geer, L.Y.; Geer, R.C.; Gonzales, N.R.; Gwadz, M.; Hurwitz, D.I.; Lanczycki, C.J.; *et al.* CDD: Conserved domains and protein three-dimensional structure. *Nucleic Acids Res.* **2013**, *41*, D348–D352.
71. Adell, T.; Gamulin, V.; Perovic-Ottstadt, S.; Wiens, M.; Korzhev, M.; Muller, I.M.; Muller, W.E. Evolution of metazoan cell junction proteins: The scaffold protein MAGI and the transmembrane receptor tetraspanin in the demosponge *Suberites domuncula*. *J. Mol. Evol.* **2004**, *59*, 41–50.
72. Todres, E.; Nardi, J.B.; Robertson, H.M. The tetraspanin superfamily in insects. *Insect Mol. Biol.* **2000**, *9*, 581–590.
73. Jaramillo-Gutierrez, G.; Rodrigues, J.; Ndikuyeze, G.; Povelones, M.; Molina-Cruz, A.; Barillas-Mury, C. Mosquito immune responses and compatibility between *Plasmodium* parasites and anopheline mosquitoes. *BMC Microbiol.* **2009**, *9*, e154.
74. Alves-Silva, J.; Ribeiro, J.M.C.; van den Abbeele, J.; Attardo, G.; Hao, Z.; Haines, L.R.; Soares, M.B.; Berriman, M.; Aksoy, S.; Lehane, M.J. An insight into the sialome of *Glossina morsitans morsitans*. *BMC Genomics* **2010**, *11*, e213.
75. Greaves, J.; Chamberlain, L.H. Palmitoylation-dependent protein sorting. *J. Cell Biol.* **2007**, *176*, 249–254.
76. Linder, M.E.; Deschenes, R.J. Palmitoylation: Policing protein stability and traffic. *Nat. Rev. Mol. Cell Biol.* **2007**, *8*, 74–84.
77. Berditchevski, F.; Odintsova, E.; Sawada, S.; Gilbert, E. Expression of the palmitoylation-deficient CD151 weakens the association of alpha 3 beta 1 integrin with the tetraspanin-enriched microdomains and affects integrin-dependent signaling. *J. Biol. Chem.* **2002**, *277*, 36991–37000.
78. Charrin, S.; Manie, S.; Oualid, M.; Billard, M.; Boucheix, C.; Rubinstein, E. Differential stability of tetraspanin/tetraspanin interactions: Role of palmitoylation. *FEBS Lett.* **2002**, *516*, 139–144.
79. Yang, X.; Claas, C.; Kraeft, S.K.; Chen, L.B.; Wang, Z.; Kreidberg, J.A.; Hemler, M.E. Palmitoylation of tetraspanin proteins: Modulation of CD151 lateral interactions, subcellular distribution, and integrin-dependent cell morphology. *Mol. Biol. Cell* **2002**, *13*, 767–781.
80. Hotta, H.; Ross, A.H.; Huebner, K.; Isobe, M.; Wendeborn, S.; Chao, M.V.; Ricciardi, R.P.; Tsujimoto, Y.; Croce, C.M.; Koprowski, H. Molecular cloning and characterization of an antigen associated with early stages of melanoma tumor progression. *Cancer Res.* **1988**, *48*, 2955–2962.
81. Guindon, S.; Gascuel, O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* **2003**, *52*, 696–704.

82. Wang, F.; Vandepoele, K.; van Lijsebettens M. Tetraspanin genes in plants. *Plant Sci.* **2012**, *190*, 9–15.
83. Garcia-Espana, A.; Chung, P.J.; Sarkar, I.N.; Stiner, E.; Sun, T.T.; Desalle, R. Appearance of new tetraspanin genes during vertebrate evolution. *Genomics* **2008**, *91*, 326–334.
84. Kariithi, H.M.; van Den, A.; Boeren, S.; Caljon, G.; İkbal A.İ.; Murungi, K.E.; Abd-Alla, M.M.A.; Vlak, J.M. Laboratory of Virology, Wageningen University: Wageningen, The Netherlands; Unpublished data, 2014.
85. Seigneuret, M. Complete predicted three-dimensional structure of the facilitator transmembrane protein and hepatitis C virus receptor CD81: Conserved and variable structural domains in the tetraspanin superfamily. *Biophys. J.* **2006**, *90*, 212–227.
86. Reddy, Ch.S.; Vijayasathy, K.; Srinivas, E.; Sastry, G.M.; Sastry, G.N. Homology modeling of membrane proteins: A critical assessment. *Comput. Biol. Chem.* **2006**, *30*, 120–126.
87. Ohta, T. Synonymous and nonsynonymous substitutions in mammalian genes and the nearly neutral theory. *J. Mol. Evol.* **1995**, *40*, 56–63.
88. Nielsen, R.; Yang, Z. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics* **1998**, *148*, 929–936.
89. Kreitman, M.; Akashi, H. Molecular evidence for natural selection. *Annu. Rev. Ecol. Syst.* **1995**, *26*, 403–422.
90. Martin, F.; Roth, D.M.; Jans, D.A.; Pouton, C.W.; Partridge, L.J.; Monk, P.N.; Moseley, G.W. Tetraspanins in viral infections: A fundamental role in viral biology? *J. Virol.* **2005**, *79*, 10839–10851.
91. Van Compernelle, S.E.; Wiznycia, A.V.; Rush, J.R.; Dhanasekaran, M.; Baures, P.W.; Todd, S.C. Small molecule inhibition of hepatitis C virus E2 binding to CD81. *Virology* **2003**, *314*, 371–380.
92. Spoden, G.; Freitag, K.; Husmann, M.; Boller, K.; Sapp, M.; Lambert, C.; Florin, L. Clathrin- and caveolin-independent entry of human papillomavirus type 16: Involvement of tetraspanin-enriched microdomains (TEMs). *PLoS One* **2008**, *3*, e3313.
93. Ho, S.H.; Martin, F.; Higginbottom, A.; Partridge, L.J.; Parthasarathy, V.; Moseley, G.W.; Lopez, P.; Cheng-Mayer, C.; Monk, P.N. Recombinant extracellular domains of tetraspanin proteins are potent inhibitors of the infection of macrophages by human immunodeficiency virus type 1. *J. Virol.* **2006**, *80*, 6487–6496.
94. Nydegger, S.; Khurana, S.; Kremmentsov, D.N.; Foti, M.; Thali, M. Mapping of tetraspanin-enriched microdomains that can function as gateways for HIV-1. *J. Cell Biol.* **2006**, *173*, 795–807.
95. Von Lindern, J.J.; Rojo, D.; Grovit-Ferbas, K.; Yeramian, C.; Deng, C.; Herbein, G.; Ferguson, M.R.; Pappas, T.C.; Decker, J.M.; Singh, A.; *et al.* Potential role for CD63 in CCR5-mediated human immunodeficiency virus type 1 infection of macrophages. *J. Virol.* **2003**, *77*, 3624–3633.
96. Tran, M.H.; Pearson, M.S.; Bethony, J.M.; Smyth, D.J.; Jones, M.K.; Duke, M.; Don, T.A.; McManus, D.P.; Correa-Oliveira, R.; Loukas, A. Tetraspanins on the surface of *Schistosoma mansoni* are protective antigens against schistosomiasis. *Nat. Med.* **2006**, *12*, 835–840.
97. Tran, M.H.; Freitas, T.C.; Cooper, L.; Gaze, S.; Gatton, M.L.; Jones, M.K.; Lovas, E.; Pearce, E.J.; Loukas, A. Suppression of mRNAs encoding tegument tetraspanins from *Schistosoma mansoni* results in impaired tegument turnover. *PLoS Pathog.* **2010**, *6*, e1000840.

98. Dakshinamoorthy, G.; Munirathinam, G.; Stoicescu, K.; Reddy, M.V.; Kalyanasundaram, R. Large extracellular loop of tetraspanin as a potential vaccine candidate for filariasis. *PLoS One*. **2013**, *8*, e77394.

© 2014 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).