



**HAL**  
open science

# A New Assessment of Thioester-Containing Proteins Diversity of the Freshwater Snail *Biomphalaria glabrata*

David Duval, Remi Pichon, Damien Lassalle, Maud Laffitte, Benjamin Gourbal, Richard Galinier

► **To cite this version:**

David Duval, Remi Pichon, Damien Lassalle, Maud Laffitte, Benjamin Gourbal, et al.. A New Assessment of Thioester-Containing Proteins Diversity of the Freshwater Snail *Biomphalaria glabrata*. *Genes*, 2020, 11, 10.3390/genes11010069 . hal-02431935

**HAL Id: hal-02431935**

**<https://hal.science/hal-02431935v1>**

Submitted on 8 Jan 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Article

# A New Assessment of Thioester-Containing Proteins Diversity of the Freshwater Snail *Biomphalaria glabrata*

David Duval, Remi Pichon, Damien Lassalle, Maud Laffitte, Benjamin Gourbal and Richard Galinier \*

IHPE, Université de Montpellier, CNRS, Ifremer, Université de Perpignan Via Domitia, 66860 Perpignan, France; david.duval@univ-perp.fr (D.D.); remi.pichon@univ-perp.fr (R.P.); damien.lassalle@univ-perp.fr (D.L.); maud.laffitte01@etu.umontpellier.fr (M.L.); benjamin.gourbal@univ-perp.fr (B.G.)

\* Correspondence: richard.galinier@univ-perp.fr

Received: 26 November 2019; Accepted: 20 December 2019; Published: 7 January 2020

**Abstract:** Thioester-containing proteins (TEPs) superfamily is known to play important innate immune functions in a wide range of animal phyla. TEPs are involved in recognition, and in the direct or mediated killing of several invading organisms or pathogens. While several TEPs have been identified in many invertebrates, only one TEP (named BgTEP) has been previously characterized in the freshwater snail, *Biomphalaria glabrata*. As the presence of a single member of that family is particularly intriguing, transcriptomic data and the recently published genome were used to explore the presence of other BgTEP related genes in *B. glabrata*. Ten other TEP members have been reported and classified into different subfamilies: Three complement-like factors (BgC3-1 to BgC3-3), one  $\alpha$ -2-macroglobulin (BgA2M), two macroglobulin complement-related proteins (BgMCR1, BgMCR2), one CD109 (BgCD109), and three insect TEP (BgTEP2 to BgTEP4) in addition to the previously characterized BgTEP that we renamed BgTEP1. This is the first report on such a level of TEP diversity and of the presence of macroglobulin complement-related proteins (MCR) in mollusks. Gene structure analysis revealed alternative splicing in the highly variable region of three members (BgA2M, BgCD109, and BgTEP2) with a particularly unexpected diversity for BgTEP2. Finally, different gene expression profiles tend to indicate specific functions for such novel family members.

**Keywords:** *Biomphalaria glabrata*; Thioester-containing protein; alternative splicing; immunity; Schistosome

## 1. Introduction

Thioester-containing proteins (TEP) are large secreted proteins playing central roles in the innate immune response [1,2]. TEP are characterized by the presence of a unique intrachain  $\beta$ -cysteinyl- $\gamma$ -glutamyl thioester bond (CGEQ) originally discovered in the human protease inhibitor,  $\alpha$ -2-macroglobulin (A2M), and complement C3 and C4 [3]. A2M are pan-protease suicide inhibitors that encapsulate targets after protease cleavage activation, leading to neutralization of targeted protease activity [4-8]; whereas complement factors deposit on target surfaces following activation, enabling the subsequent elimination or killing of the pathogens [9-11].

Phylogenetic analyses classified TEP proteins in two distinct groups, the complement factor and A2M depending on the presence of two domains, anaphylotoxin and C345C. The complement factor group is composed of vertebrate C3, C4, and C5 components and their orthologues (complement-like factors) in invertebrates. The A2M group comprises several closely related molecules:  $\alpha$ -2-macroglobulin (A2M), macroglobulin complement-related protein (MCR),

pregnancy zone protein (PZP), C3/PZP-like/A2M domain containing 8 (CPAMD8), antigen CD109 (CD109), and insect TEP (iTEP) [12].

Since the increasing availability of molecular databases for Metazoa (from Cnidarians to Mammalians), the thioester-encoding regions were found in many genes, that help in a better definition of the thioester-containing protein family. This ancestral family of molecules is present in all animal phyla except Porifera (sponges) [12]. However, each TEP type is not always recovered in all phyla. For example, seven members of this family are encoded in the human genome: Three complement factors (C3, C4, and C5), A2M, PZP, CD109, and CPAMD8. At the opposite, *Drosophila melanogaster* encodes solely six members, five iTEP, and one MCR [13-15]. The question of the origin and evolution of TEP was raised by several studies [1,12]. Indeed, iTEP was only found for Protostomia, MCR was, until now, only reported for arthropods, whereas complement factor TEP was never found in available insect genomes while present in all other arthropod groups. If vertebrate genomes are relatively well documented, data from invertebrates are too fragmentary to have a clear view of TEP gene distribution among them. Thus, additional data of TEP family member characterizations from organisms belonging to seldom documented phyla would be mandatory for a better understanding of TEP origin and evolution.

Among invertebrates, TEPs were most deeply studied in insect group, especially in flies and mosquitos, with *Anopheles gambiae* TEP1 (AgTEP1) being the only TEP for which a crystal structure was resolved [16]. Several TEPs were also characterized in other arthropod species, like in the Tick *Ixodes ricinus*, for which most of the TEP classes, including C3-like complement factors, A2M, MCR, and iTEP were described [17]. In *D. melanogaster*, iTEP and MCR are induced upon bacterial, fungal, and parasitic challenges [1,14,15,18-20]. In mosquito, AgTEP1 was shown to be able to directly bind malaria parasite and to promote their elimination [21,22], while *Aedes aegypti* iTEP have been shown to regulate *Flavivirus* infection [23,24]. Finally, the Mollusca phylum has been seldom explored, CD109 molecule was characterized in the cephalopod *Euprymna scolopes* [25], two bivalve species were reported to contain A2M, iTEP, and C3-like complement factor [26-28], and more recently five TEP members belonging to C3-like complement factors, A2M, and CD109 classes have been identified in the gastropod *Littorina littorea* [29]. However, only few data are available concerning the functional role of such TEP proteins in mollusc phylum.

The freshwater snail *Biomphalaria glabrata* (Bg) is the intermediate host of the trematode parasite *Schistosoma mansoni*, the agent of human intestinal schistosomiasis. As such, many studies have been conducted on the immunobiological interactions between these two protagonists [30-32]. In this host-parasite interaction was defined the compatible/incompatible status of the so-called compatibility polymorphism [33-36] that resulted from the co-evolutionary arm race between host and parasite [37]. Compatibility polymorphism can be defined as follow: A specific *B. glabrata* strain harbors variable degrees of susceptibility toward different parasite strains, whereas a defined *S. mansoni* strain presents different levels of virulence/infectivity toward different *B. glabrata* strains. Several studies have thus investigated key molecules of snail immune system and parasite virulence factors in order to decipher the molecular dialogue between *B. glabrata* and *S. mansoni*, and to explain this compatibility polymorphism phenotype [30,38-43]. Among the snail immune determinants playing a role in the host-parasite interplay, proteomic analyses have demonstrated the presence of a thioester-containing protein (BgTEP) within an immune complex formed by parasite antigen and snail immune recognition receptor [39,44]. A recent study characterized this BgTEP as being similar to insect TEP, secreted by hemocytes, expressed following immune challenges, and able to bind to different pathogen teguments including the parasite *S. mansoni* [45]. Binding of full-length and processed forms toward such pathogens, suggests important immune functions and potential opsonizing role [45]. In addition, a protease inhibitor activity of this BgTEP following proteolytic cleavage was previously provided by several biochemical studies, and more particularly against a cysteine proteinase secreted by *S. mansoni* [46-48]. Altogether, these results suggested a potential dual function of BgTEP as protease inhibitor and potential opsonin involved in snail immune response [45].

In the present study, we used transcriptomic data, as well as the *B. glabrata* recently published genome [49], to explore the diversity of TEP family members in the snail *B. glabrata*. We described 10 additional members of TEP family, based on the presence of thioester and A2M domains. We subsequently performed a phylogenetic analysis to classify these molecules and analyzed their exon-intron structures. We also determined the tissue expression pattern of BgTEP genes by quantitative PCR.

## 2. Materials and Methods

### 2.1. Ethical Statements

Our laboratory holds permit # A66040 for experiments on animals, which was obtained from the French Ministry of Agriculture and Fisheries and the French Ministry of National Education, Research, and Technology. The housing, breeding, and care of the utilized animals followed the ethical requirements of our country. The experimenter possesses an official certificate for animal experimentation from both of the above-listed French ministries (Decree # 87–848, 19 October 1987). The various protocols used in this study have been approved by the French veterinary agency of the DRAAF Languedoc-Roussillon (Direction Régionale de l'Alimentation, de l'Agriculture et de la Forêt), Montpellier, France (authorization # 007083).

### 2.2. Database Mining, Gene Identification, and Sequence Analysis

The transcriptomes corresponding either to naïve [50] (NCBI BioProject PRJNA213050) or to *S. mansoni* infected [51] snails from Brazil (BgBre1) (Recife, Brazil), and whole-genome sequence databases of *B. glabrata* [49] were screened to search for the presence of A2M (pfam00207) and/or thioester-containing (cd02897) domains to identify TEP related genes. When only partial transcripts were obtained from transcriptomic data for a given gene, the lacking portions of cDNA were confirmed by standard PCR amplification to complete cDNA sequences.

Then, genomic structures of the corresponding genes were determined using VectorBase database (<https://www.vectorbase.org/organisms/biomphalaria-glabrata>).

Open reading frames for each gene and their encoded amino acid sequences were determined using Translate tool from the ExPasy Bioinformatics Resource Portal of the Swiss Institute of Bioinformatics (<https://web.expasy.org/translate/>).

Conserved protein domains were searched using SMART tool [52,53] from the European Molecular Biology Laboratory (<http://smart.embl-heidelberg.de/>). Signal peptide sequences were predicted using both SignalP 4.0 Server (<http://www.cbs.dtu.dk/services/SignalP-4.0/>) and Interproscan. Transmembrane domains were predicted using the TMHMM Server version 2.0 (<http://www.cbs.dtu.dk/services/TMHMM-2.0/>), whereas glycosylphosphatidylinositol (GPI)-anchor sites were predicted using the GPI-SOM Server (<http://gpi.unibe.ch/>). Amino acid sequence alignments between all TEP proteins were generated by BioEdit software version 7.1.3.0 using ClustalW multiple alignment.

### 2.3. RNA Extraction and Quantitative RT-PCR Analysis

Tissues were collected from 9 individual *B. glabrata* snails (Brazilian strain, Recife) under binocular microscope dissection. Albumen gland, head-foot, intestine, stomach, hepatopancreas, and ovotestis were recovered. Six hemocyte samples were each recovered by pooling the hemolymph of 20 snails followed by a 10 min centrifugation at 10,000× g. Total RNA was then extracted with TRIzol reagent (Sigma Life Science, Saint Louis, Missouri, USA) according to manufacturer's instructions and subsequently reverse transcribed to first strand cDNA from oligo dT using Maxima H Minus First Strand cDNA Synthesis Kit with dsDNase (Thermo Scientific, Waltham, Massachusetts, USA) according to manufacturer's instructions.

Quantitative PCR analyses were performed with the LightCycler 480 System (Roche, Basel, Switzerland) in a 10 µL volume comprising 2 µL of cDNA diluted to 1:30 with ultrapure-water, 5 µL of No Rox SYBR Master Mix blue dTTP (Takyon Eurogentec, Liege, Belgium), 1 µL of

ultrapure-water, and 10  $\mu$ M of each primer. Each couple of primers used in this study has been tested to determine the exponential and efficiency of PCR reaction. PCR amplification efficiencies were established for each target and house-keeping gene by calibration curves using two times serial dilutions of cDNA (from 1/20 to 1/640) in triplicates. Amplification efficiencies were calculated using slope values of the log-linear portion of the calibration curves by the LightCycler 480 Software release 1.5 (Roche). As a result, only the primer couples presenting a PCR amplification efficiency of 2 were retained for the study (Table S1). None of the primer used for quantification of gene expression was designed in region submitted to alternative splicing. The cycling program is as follows: Denaturation step at 95 °C for 2 min, 45 cycles of amplification (denaturation at 95 °C for 10 s, annealing and elongation together at 60 °C for 45 s) with single fluorescence acquisition at the end of each amplification cycle. Q-PCR was ended by a melting curve step from 65 °C to 97 °C with a heating rate of 0.11 °C/s and continuous fluorescence measurement. For each reaction, the cycle threshold (Ct) was determined using the 2nd derivative method of the LightCycler 480 Software release 1.5 (Roche). PCR experiments were performed in triplicate (technical replicates) from each biological replicate. The mean value of Ct was calculated. Corrected melting curves were checked using the Tm-calling method of the LightCycler 480 Software release 1.5. The relative expression of each TEP gene was calculated with the  $\Delta\Delta$ Ct method as the efficiency of both couple of primers (target and housekeeping genes) presented the same PCR amplification efficiency. Results were normalised with respect to a housekeeping gene: The S19 ribosomal protein gene, as previously described [45].

#### 2.4. PCR Analysis of BgTEP2 Isoforms

According to BgTEP2 cDNA sequence, a couple of primer (F1: 5'-ACTACGGAGGCAGTGATGC-3'; R1: 5'-GATAGTATCTGGTACTGTTGC-3') designed in non-variable parts of the sequence and framing the highly variable region was used to detect all the different alternatively spliced isoforms of BgTEP2 from snail tissues prepared as described above. Then, 1  $\mu$ L of a 1:30 dilution of cDNA was used as template amplification with primers F1 and R1 using the GoTaq G2 HotStart enzyme (Promega, Madison, Wisconsin, USA) in a 50  $\mu$ L final reaction, according to the manufacturer's instructions. PCR reaction was performed in a Biometra TOne thermocycler (Analytik Jena AG, Jena, Thuringia, Germany), using the following cycling program: Initial denaturation 3 min at 95 °C, 45 cycles of amplification (denaturation at 95 °C for 30 s, annealing at 55 °C for 30 s, elongation at 72 °C for 20 s), final elongation of 2 min at 72 °C. Then, 20  $\mu$ L of the resulting PCR were analyzed by electrophoresis on a 2% agarose gel. Another couple of primer targeting the cDNA coding for the ribosomal protein S19 was used as a positive control of PCR amplification (S19F: 5'-TTCTGTTGCTCGCCAC-3'; S19R: 5'-CCTGTATTTGCATCCTGTT-3').

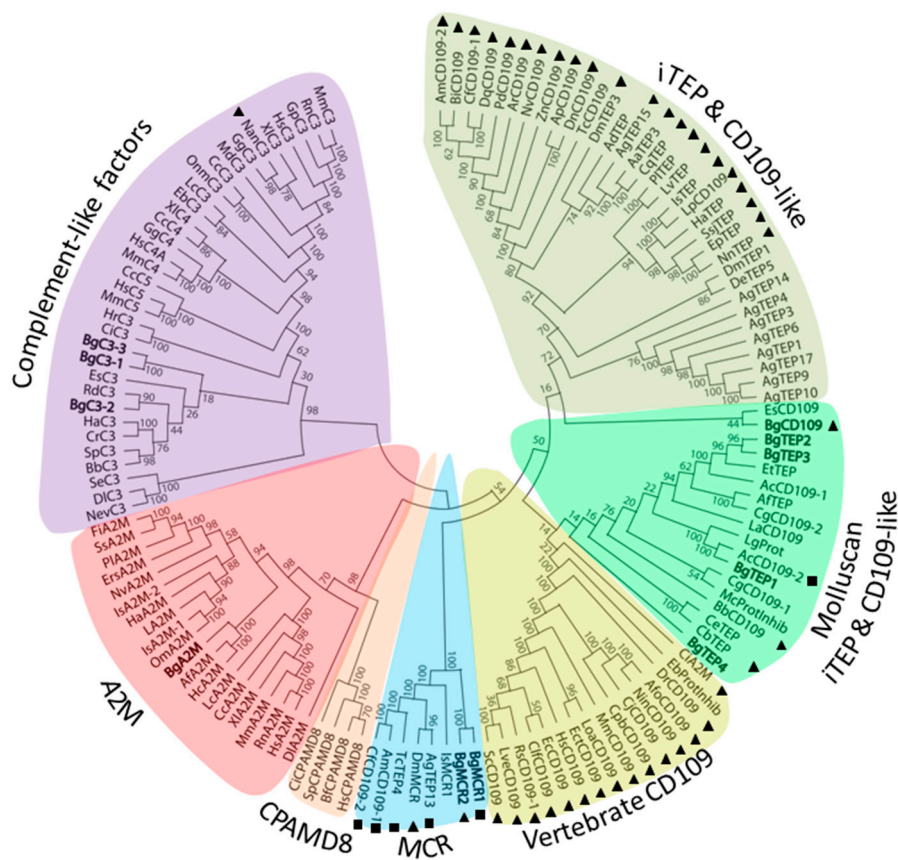
#### 2.5. Phylogenetic Analyses

Homologous amino acid sequences were identified using BLASTp searches against the GenBank non redundant database (Bethesda, MD, USA). Full length protein sequences covering a wide range of animal phyla were used for multiple sequence alignment with the Guidance 2 server (<http://guidance.tau.ac.il/ver2/>) using the MAFFT algorithm [54] (Figure S1). Phylogenetic analyses of these aligned sequences was performed by using the maximum likelihood method and Whelan And Goldman + Freq. model [55], with pairwise deletion option as gaps/missing data treatment. A bootstrap analysis of 1000 replications was carried out to assess the robustness of the tree branches. The tree with the highest log likelihood (-345,206.19) is shown. Initial tree(s) for the heuristic search were obtained automatically by applying neighbor-join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. A discrete  $\gamma$  distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 1.5454)). The rate variation model allowed for some sites to be evolutionarily invariable (+I, 0.06% sites). This analysis involved 134 amino acid sequences. There were a total of 5324 positions in the final dataset. Evolutionary analyses were conducted in MEGA X [56].

### 3. Results

#### 3.1. Phylogenetic Analysis of *B. glabrata* Thioester-Containing Proteins

Interproscan analysis of two snail transcriptomes was conducted to identify new TEP-related members and had revealed several positive transcripts that were assembled to generate a total of 11 TEP cDNA sequences (Figure S2). In order to investigate the relationship of the different snail TEP (BgTEPs) identified with various orthologous sequences recovered from other animal species, a phylogenetic analysis was performed using 123 other sequences (Table S2). Amino-acid sequences of full-length TEP superfamily proteins from both vertebrate and invertebrate phyla, including Complement-like factors, A2M, MCR, CPAMD8, iTEP, and CD109, and the 11 BgTEP sequences were used to construct a phylogenetic tree with the maximum likelihood method (Figure 1).



**Figure 1.** Phylogenetic tree of various thioester-containing protein (TEP) superfamily members. A total of 134 full-length protein sequences from diverse animal origin were used to construct the tree. A bootstrap analysis of 1000 replications was carried out on the tree inferred from the maximum likelihood method and the values are shown at each branch of the tree. BgTEPs are indicated in bold characters. Black triangles correspond to predicted glycosylphosphatidylinositol (GPI)-anchor signal, while black square corresponds to C-terminal transmembrane helix. TEP subgroups are highlighted by different colors: Purple for complement-like factors, pink for A2M, orange for CPAMD8, blue for MCR, and shades of green for iTEP and CD109-like.

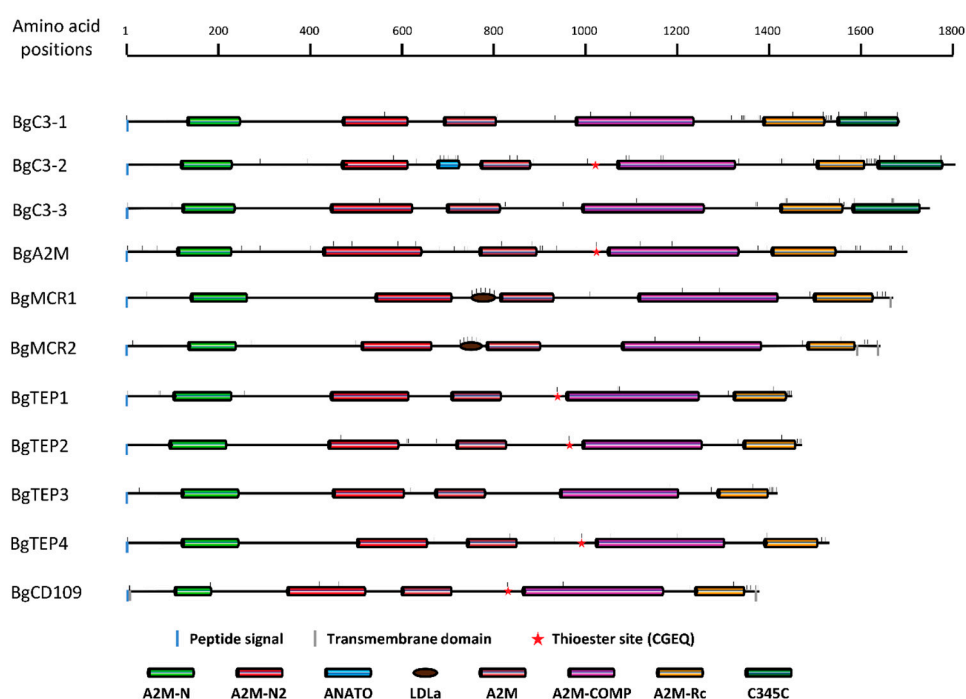
TEP molecules from *B. glabrata* belong to four clades of the TEP superfamily, and were named according to the actual nomenclature: (1) Complement-like factors (BgC3-1, BgC3-2, and BgC3-3), (2) Alpha2-macroglobulins (BgA2M), (3) Macroglobulin complement-related proteins (BgMCR1 and BgMCR2), and (4) insect TEP/CD109 molecules (BgTEP1, BgTEP2, BgTEP3, BgTEP4, and BgCD109). BgC3 molecules are close to the C3-like proteins from two other mollusk species, with BgC3-2 being more distant from BgC3-1 and BgC3-3 that clustered together. It is of note that a hypothetical

transcripts corresponding to incomplete BgC3-2 cDNA was previously identified from the *B. glabrata* genome publication [49]. BgA2M also clustered with the A2M members from other mollusks. The two BgMCR clustered together and are the first macroglobulin complement-related proteins reported from the Mollusca phylum. Between the four iTEP, only BgTEP2 and BgTEP3 cluster together, while the two others are distributed among the cluster of molluscan iTEP and CD109-like. It is noteworthy that the name of BgTEP1 was assigned to the sole BgTEP previously identified from *B. glabrata* [39,45]. The taxonomic position of BgCD109 is less clear as it is more distant from BgTEPs and clusters with the CD109 from the bobtail squid *E. scolopes* (EsCD109).

As vertebrate CD109 are characterized by the presence of GPI-anchoring site, we looked for the presence of such GPI-anchor signal prediction as well as C-terminal transmembrane helix prediction in all the TEP members used in this analysis. This confirmed the presence of GPI-anchor site in vertebrate CD109, but also in numerous invertebrate molecules either named CD109-like or iTEP, including BgCD109. Most of the GPI-anchor site predicted molecules tend to form a specific cluster sustained by a bootstrap value of 92%. Another notable result is that iTEP and CD109-like molecules from mollusks tend to form a specific cluster separated from other invertebrate iTEP and CD109. Such classification has been previously highlighted [45] but is not well supported by bootstrap values, and need to be validated by an analysis including much more TEP members from Mollusca phylum.

### 3.2. BgTEP Protein Features Analysis

To confirm taxonomic positions revealed by the phylogenetic analysis, we investigated for the presence of evolutionary conserved domains in the BgTEP amino acid sequences using SMART tool. Figure 2 shows an overview of the protein domain organization for each TEP member. All BgTEP candidates possess a signal peptide meaning that all these molecules are secreted. Several conserved domains (A2M-N, A2M-N2, A2M, A2M-complement, and A2M-Receptor binding domain) defining the A2M proteins were also recovered in BgA2M with the similar organization. In addition, A2M proteins include the presence of a highly variable bait region, which determines the specificity of these protease inhibitors [57].



**Figure 2.** Schematic representation of BgTEP protein domains organization. Colored rectangles correspond to SMART conserved domains with following accession number: N-terminal

alpha2-macroglobulin domain (A2M-N, PF01835), N-terminal  $\alpha$ -2-macroglobulin domain2 (A2M-N2, SM001359), anaphylotoxin domain (ANATO, SM000104), low-density lipoprotein receptor domain class A (LDLa, SM000192),  $\alpha$ -2-macroglobulin domain (A2M, SM001360),  $\alpha$ -2-macroglobulin complement domain (A2M-COMP, PF07678),  $\alpha$ -2-macroglobulin receptor-binding domain (A2M-Rc, SM001361), Netrin C-terminal domain (C345C, SM000643). Blue and grey vertical bar positioned under the main horizontal line respectively correspond to peptide signal and transmembrane helix predictions. Small horizontal bars positioned above the main lines correspond to cysteine positions in BgTEP protein sequences.

The other TEP superfamily members are defined by the presence of conserved domains similar to the A2M, but also possess additional features or domains. Indeed, TEP complement factors possess an extra C-terminal C345C domain characterized by four conserved cysteine residues that are likely to form internal disulphide bonds [58]; this domain was recovered in our three BgC3 (See Figure 2 and Figure S3). On the other hand, the anaphylotoxin domain and the arginine-rich region, two other specific features of C3-like molecules, were only recovered for BgC3-2 and not for BgC3-1 and BgC3-3 (Figure S3). The anaphylotoxin domain is characterized by a stretch of six cysteines, and the arginine-rich region is a classic posttranslational proteolytic cleavage site where C3 molecules could be cut into  $\alpha$ - and  $\beta$ -chain by respective serine protease [59,60]. Furthermore, the canonical CGEQ thioester site is mutated in BgC3-1 and BgC3-3 protein sequences, which clearly indicates that these two members cannot bind target molecules via a functional thioester bond.

MCR molecules contain an additional low-density lipoprotein receptor domain class A (LDLa domain), in place of (i) the anaphylotoxin domain harbored by the complement factors or (ii) the A2M bait region [15]. This LDLa domain constituted by a cysteine-rich repeat typical from the N-terminal part of LDL receptor involved in lipoproteins binding is present in each BgMCR molecule. The presence of one or two C-terminal transmembrane helices were predicted respectively in BgMCR1 and BgMCR2, which is also consistent with MCR in other organisms [15]. Like all MCR orthologs, the canonical thioester site is mutated in BgMCR sequences. Finally, they also present a C-terminal stretch of cysteine residues, four of them being extremely conserved between all known MCR (See Figure 2 and Figure S3).

The four BgTEPs and the BgCD109 do not contain additional domain when compared with BgA2M. Nevertheless, they are shorter in length by 200 to 350 amino acids and present a C-terminal stretch of cysteines, a common signature of iTep (Figure 2). Only BgTEP3 possesses a point mutation in the thioester site, changing the key cysteine into a serine residue (Figure S3). In addition, N- and C-terminal transmembrane domains were also predicted for BgCD109 molecule, which is consistent with the predicted GPI-anchor site, as CD109 are known to be cell surface-membrane bound molecules [25,61].

Another important feature is the presence of a catalytic residue located approximately 110 amino acids downstream of the thioester site in the primary sequence, but very close in the tridimensional structure of TEP molecules [3]. The nature of this catalytic residue partly determines binding specificity of the thioester site [62,63]. This catalytic amino acid was identified as a serine residue for BgA2M like in other mollusk A2M [26] and most other protostomia invertebrate A2M, while asparagine residue is the most commonly encountered in vertebrates [64]. Histidine, which is the typical catalytic residue of complement factors, was only found in BgC3-2 sequence, but also in BgTEP2, BgTEP4, and BgCD109, while BgTEP1 harbors an aspartic acid (Figure S3). It is noteworthy that aspartic acid is a well-known alternative catalytic residue also found in human C4a molecule [62].

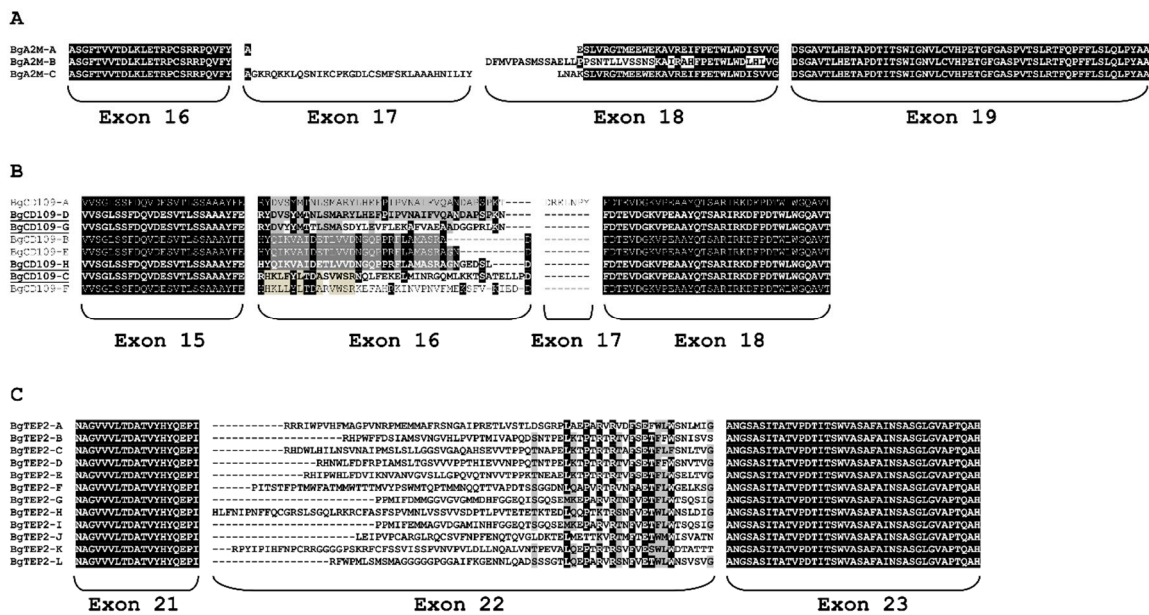
### 3.3. Organization and Structure of BgTEP Genes

BgTEPs are large proteins ranging from 1389 to 1804 amino acid lengths (Figure 2). In order to determine their genomic organizations, BgTEP cDNA sequences obtained from two snail transcriptomes [50,51] were mapped to the genome BglaB1assembly [49] using the Blast tool of the VectorBase website. BgTEP genes are composed of various exon numbers ranging from 35 to 43 (Table S3). All reconstructed cDNA were retrieved in the recently published genome of *B. glabrata*.



However, 9 out of 11 corresponding genes were distributed upon several different scaffolds in the BglaB1 genome assembly (from 2 to 11 different scaffolds). Only BgC3-3 and BgA2M genes were each assembled on a single scaffold. Moreover, several genes comprised missense or badly positioned exons, for example: BgC3-1 (exons 27-29), BgC3-3 (exon 33), BgA2M (exon 1), BgMCR1 (exons 23-26), BgMCR2 (exon 34), BgTEP1 (exon 31), BgTEP4 (exons 7, 16, 17), and BgCD109 (exons 10, 28, 29) (Table S3). BgA2M was the only gene not completely determined as the exon(s) coding for signal peptide and 5'UTR were not recovered in the snail genome assembly. Sequences of exon-intron boundaries were conserved along all sequences (Table S4) as already described for BgTEP1 [45]. Importantly, none of the BgTEP protein was correctly predicted from BglaB1 genome assembly (Figure S4).

Interestingly, alternative splicing was identified from transcriptomic data for the three genes encoding BgA2M, BgTEP2, and BgCD109 (Tables S3 and S4). The alternative exons encode the highly variable region corresponding to the bait region of BgA2M protein and to the proteolytic cleavage site for BgTEP2 and BgCD109 proteins. Thus, we identified three variants of BgA2M protein of which two were also predicted from BglaB1 assembly (Figure 3A). Regarding BgCD109, a total of eight protein variants were predicted from BglaB1 genome assembly, while transcripts corresponding to only four of these eight variants were recovered from Bg transcriptomes (Figure 3B). Particular attention was paid to BgTEP2 gene for which we identified 12 possible variants from *B. glabrata* transcriptomes (Figure 3C, Table S3), of which only one was predicted from BglaB1 assembly. The 12 possible alternative exons number 22 are positioned tandemly along 30kb length of the genome (Tables S3 and S4).

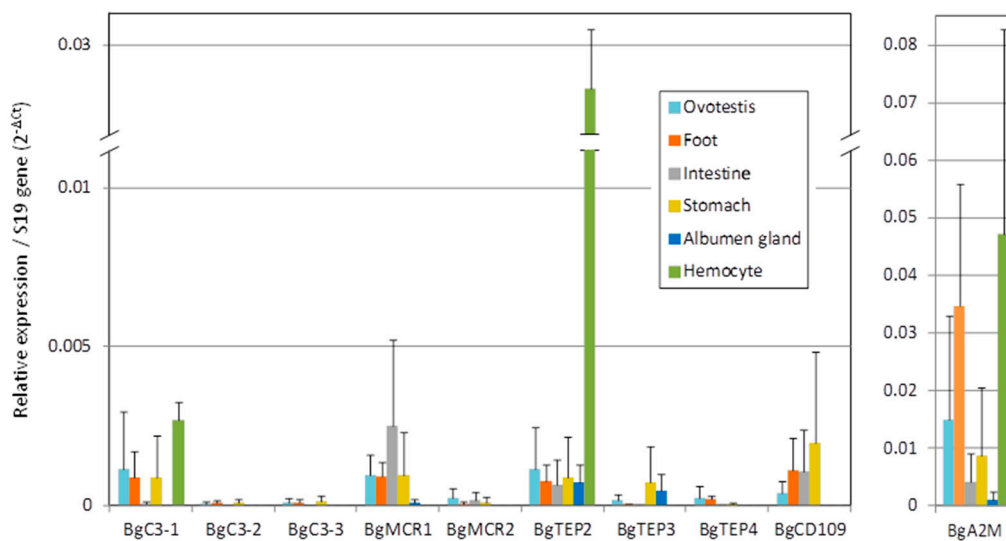


**Figure 3.** ClustalW alignment of alternative splice variants for three BgTEP proteins. Amino acids encoded by alternative exons corresponding to the highly variable region of some BgTEP were aligned, together with amino acids from border conserved exons, using clustalW. Identical amino acids are highlighted in black, similar ones in grey. (A) Alignment of BgA2M variants showing alternative splicing in exons 17 and 18. (B) Alignment of BgCD109 splice variants. Underlined variant names correspond to variants recovered from *Biomphalaria glabrata* transcriptomes. The other variants were only predicted from BglaB1 genome assembly. (C) Alignment of the 12 BgTEP2 splice variants.

### 3.4. BgTEPs Genes Expression Pattern in Snail Tissues

Tissue specific expression of BgTEP by quantitative RT-PCR was conducted using head-foot, albumen gland, ovotestis, intestine, stomach, and hemocytes of *B. glabrata* (Figure 4). BgTEP1 was not included in the present analysis as already published [45]. We observed a large variability

between the different *B. glabrata* TEP members in both tissue distribution and expression levels. At the tissue distribution level, the expression of all BgTEPs was detected in ovotestis, head-foot, and stomach, while it is more variable for other tissues. Six members belonging to the four TEP clades (BgC3-1, BgMCR1, BgMCR2, BgTEP2, BgCD109, and BgA2M) are also expressed in intestine, whereas the expression of only four members (BgMCR1, BgTEP2, BgTEP3, and BgA2M) was detected in albumen gland, and finally, three members (BgC3-1, BgTEP2, and BgA2M) were exclusively detected in hemocytes. At the expression level, BgA2M is 10 to 50 times more expressed than all other members, depending on the tissue considered, with the exception of BgTEP2 in hemocytes, which is expressed at the same level than BgA2M. The expression of BgC3-1, BgMCR1, BgTEP2, and BgCD109 was easily detected, while BgC3-2, BgC3-3, BgMCR2, BgTEP3, and BgTEP4 are more weakly expressed in most of the tissues tested.



**Figure 4.** Relative expression of BgTEP genes in snail tissues. Transcription profile of BgTEPs in different tissues derived from nine individual snails. Six snail tissues were dissected: Ovotestis, head-foot, intestine, stomach, albumen gland. Six pools of hemocytes were each collected from 20 snails. Quantitative RT-PCR was performed on mRNA.  $C_t$  values of transcripts were normalized to the transcript level of the reference S19 ribosomal protein gene. Error bars represent the SD of the  $\Delta C_t$  mean values obtained for each tissue. BgA2M is presented apart as transcripts level reaches a different scale.

We paid a particular attention to the splice variants of BgTEP2. We investigated the presence of putative spliced variants in different snail tissues by end point PCR amplification (Figure S5). In most tissues, several bands in the expected size range of 296–371 bp, were recovered, indicating that several isoforms of BgTEP2 are expressed in these tissues. In addition, we obtained variable band patterns between snail tissues suggesting that different BgTEP2 splice variants might be expressed depending of the tissue considered (Figure S5).

#### 4. Discussion

In the present study, we identified 10 new TEP members containing A2M domain and/or thioester domain in addition to the previously identified BgTEP1 [45]. These new TEP members were named according to their classification provided by a phylogenetic analysis based on their full-length amino acid sequences. Three C3-like complement factors, one A2M, two MCR, three iTEP, and one CD109 were identified (Figure 1). Such diversity of *B. glabrata* TEP members is unexpected, as solely few members of the TEP family have been previously reported in Mollusca, and especially for Gastropoda. The class of iTEP/CD109 is the most represented in *B. glabrata* with a total of four iTEP (including the previously described BgTEP1) and one CD109.

All BgTEPs were further characterized by the presence of specific domains, like A2M domains for all members, C345C for C3-like complement factors, or LDLa for MCR. BgTEP1-4 present the C-terminal cysteine stretch that is a characteristic of iTep [39], the MCR possess a C-terminal transmembrane domain and a mutated thioester site, and the CD109 show a GPI anchor-signal (Figure 2). The analysis of BgTEP gene structure revealed a number of exons from 35 to 43. Such structure is consistent with the TEP gene family yet described for bivalve [28] or for human [61]. However, these BgTEPs are rigorously different from *A. gambiae* or *D. melanogaster* TEPs, which are composed of 11 to 15 exons [65,66]. Such results suggest a different evolutionary history between mollusks and insects TEPs, as previously raised by other studies [13,67]. However, relationships between the two clusters formed by on the one hand vertebrate CD109 and iTep/CD109 from mollusks and on the other hand iTep/CD109 from other invertebrates are not clear as the clustering is not sufficiently sustained by bootstrap values (Figure 1). Additional iTep/CD109 sequences from *Mollusca* phylum are needed to answer this question.

Herein, the presence of MCR molecules was reported for the first time from a mollusk species, whereas this subclass of TEP was previously demonstrated solely in few arthropod species. In addition, the presence of MCR genes in mollusk genomes may be much more common than previously thought. Indeed, in a recent study on the Yesso scallop *Patinopecten yessoensis*, five TEP members were identified from this bivalve species [68], among which a member was classified as a CD109 whereas it is more likely a MCR. Indeed, this molecule assigned as CD109 because of the presence of a C-terminal transmembrane domain, also possesses an LDLa domain and a mutated thioester site, which are typical features of MCR family members [15]. In the same way, two CD109-like molecules have been previously reported from the *Littorina littorea* transcriptome [29]. Nevertheless, the LICD109-2 molecule from this gastropod also bears all the features of MCR, including LDLa domain, mutated thioester site, and C-terminal transmembrane domain.

The canonical CGEQ thioester site is mutated in 5 out of 11 BgTEP proteins: BgC3-1, BgC3-3, BgMCR1, BgMCR2, and BgTEP3. BgTEP3 harbors a single nucleotide mutation in this site leading to the replacement of the cysteine residue to a serine residue, but the Gln-moiety of the thioester is preserved. Such replacement of the cysteine residue by a serine was also reported for tsetse fly TEP2, an iTep of which the expression is particularly up-regulated in the salivary gland following *Trypanosoma brucei* infection [69]. The replacement of cysteine residue in the thioester site was also evidenced for chicken ovostatin [70], without alteration of its proteinase-inhibitor activity [71]. The absence of thioester site in the BgMCRs and two of the three BgC3 suggests that these molecules cannot covalently bind to microbe surface, but that does not mean that they are not functional. Indeed, these molecules possess several other protein–protein interacting domains, and many other TEP proteins without thioester site were shown to exert important immunological functions. For example, MCR from *A. aegypti* was shown to be implicated in the regulation of Flavivirus infection, by binding to a scavenger receptor that also interacts with viral particles, leading to the regulation of antimicrobial peptide with anti-flavivirus activity [24]. MCR from *D. melanogaster* was reported to specifically bind the *Candida albicans* surface, and to subsequently promote its phagocytosis [15]. Otherwise, some TEP molecules with canonical thioester site were shown to be able to bind pathogens in a thioester-dependent or independent manner, like the mosquito AgTEP1 towards bacteria [72]. Moreover, it was recently evidenced that some invertebrate TEPs also play roles in other functions than immunity. *Drosophila* MCR was found to be implicated in development [73], nitric oxide regulation, and metabolic processes [74,75], while AgTEP1 was shown to increase male fertility [76].

The level of BgTEP genes expression from *B. glabrata* tissues was investigated by RT-qPCR in order to investigate a tissue-specific expression. BgA2M and BgTEP2 were recovered from all tested tissues, with BgA2M being the more expressed in all tissues (Figure 4). We found that BgMCR1, BgA2M, BgTEP2, and BgTEP3 were expressed in the albumen gland, an accessory gland of the genital tract also known to produce immune relevant molecules, like antimicrobial glycoprotein LBP/BPI, or C-type lectin [77]. Otherwise, expression of BgC3-1, BgA2M, and BgTEP2 in hemocytes suggests a potential role in immune functions in addition to BgTEP1 known to be expressed by a

specific subtype of hemocytes [45]. The probable immune function of BgTEP2 is supported by its expression in hemocytes and by its identification in a proteomic screen searching for the *B. glabrata* plasma proteins able to bind surface of *S. mansoni* sporocyst [44]. Indeed, in this study, Wu and collaborators identified peptides corresponding to BgTEP1 but also to another candidate named CD109 antigen-like (see Table 2 from [44]). We confirmed that the peptide sequences of this “CD109 antigen-like” perfectly matched to BgTEP2 protein sequence described herein. This suggests that BgTEP2 is able to directly or indirectly bind *S. mansoni* tegument, as shown for BgTEP1 [45].

Finally, we paid particular attention to BgA2M, BgCD109, and BgTEP2 that present alternative splicing in their highly variable central region. We found from *B. glabrata* transcriptomes three possible forms of BgA2M (BgA2M-A to BgA2M-C) coded by either one or two exons (Figure 3A), while only two of these three variants were predicted from BglaB1 genome assembly (Figure S4). Such alternative splicing was also reported for the A2M of the shrimp *Fenneropenaeus chinensis* [64] or in insects, and may serve to extend the repertoire of inhibited proteases [21].

Concerning BgCD109, eight isoforms were predicted from BglaB1 genome assembly (BgCD109-A to BgCD109-H), while only four were recovered from *B. glabrata* transcriptomes (BgCD109-C, -D, -G, and -H) (Figure 3B). This might, of course, be due to the fact that the unrecovered transcript forms are not expressed in the available transcriptomes used herein. However, incorrect prediction of some variants from BglaB1 genome assembly has also to be considered as (i) BgCD109 isoforms A and D have the same exon 16 (same genome position), but differ by the presence of an additional exon 17 only predicted for isoform A that was never reported to be expressed and probably does not exist, and (ii) BgCD109 isoforms B, E, and H are predicted from the transcription of the same exon 16, which starts at the same position but ends at different positions for the three isoforms, that is highly improbable and explains why only one of these isoforms (BgCD109-H) was recovered from Bg transcriptomes. Unlike other TEP members, CD109 is a GPI-linked glycoprotein, which clearly indicates different function than other TEP. CD109 was originally found on endothelial cells, platelets, and activated T-cell of human [78], it was shown to interfere with TGF- $\beta$  signaling, to be highly expressed in some cancer diseases, and to play role in bone metabolism [79–82]. The role of CD109 in invertebrates remains seldom investigated, nonetheless for juvenile Hawaiian bobtail squid *E. scolopes*, a down-regulation of CD109 (EsCD109) expression in the light organ of the squid after colonization by its bioluminescent symbiont *Vibrio fischeri* was shown. The authors suggested that the presence of the symbiont modulated the squid’s immune system, including EsCD109, to permit symbiont growth and colonization of the light organ [25]. In the present study, the only data available about BgCD109 is that it is not expressed by naïve snail hemocytes (Figure 4). However, it is to note that for bobtail squid too, EsCD109 transcript was not detected in hemocyte transcriptome [83], while the EsCD109 protein was detected by Western blot in the animal blood [25], suggesting that circulating CD109 might be produced by other immune tissues than hemocytes and secreted in the hemolymph.

The more intriguing molecule is BgTEP2 for which 12 different variants were recovered from *B. glabrata* transcriptomes. After mapping on the genome, the variants were found to be the result of alternative splicing from 12 mutually exclusive exons tandemly positioned along 30 kb in the BglaB1 genome (BgTEP2-A to BgTEP2-L) (Table S4). The amino acid sequences encoded by these 12 exons are variables in length and composition. However, some amino acids are extremely conserved in the C-terminal part of the sequences encoded by these exons (Figure 3C) suggesting a key role in the structure-function of the molecule. To our knowledge, such degree of diversity is not common. Moreover, BgTEP2, the most expressed member in hemocyte, might display a central role in the innate immune function of the snail as it was shown able to bind *S. mansoni* larva surface [44].

The highly variable region of BgTEP2 and BgCD109 corresponds to the bait region of A2M, which is known to be structurally exposed and sensitive to protease cleavage [57]. Such cleavages induce conformational changes of the protein and activation of the internal thioester, leading both to fine interactions with targeted proteins, and to the exposition of receptor binding site allowing fixation to receptor-bearing cells followed by clearance of the targeted protein [84–86]. Such diversity in the sequences of these highly variable regions might produce different avidities/affinities to

various ligands. Alternative splicing in this bait-like region was previously highlighted in several invertebrate species. The A2M-2 of the shrimp *F. chinensis* harbors six variants in the same part of the sequence that are activated differently toward *Vibrio* challenges [64]. Moreover, five isoforms of *Drosophila* TEP2 were transcribed by alternative splicing of the bait region encoding exon 5 [14]. Finally, seven isoforms of *Chlamys farreri* TEP (CfTEP) were reported, being produced by alternative splicing of six mutually exclusive exons. Only one isoform is expressed in the gonads, which differs between male and female, and the other isoforms were induced following immune challenges, in a pathogen-dependent specific pattern [87]. In *B. glabrata*, several isoforms of BgTEP2 display a restricted expression pattern in naive snail tissues suggesting a broad repertoire of recognition molecules. Their patterns of expression following specific immune challenges will deserve further investigations.

*D. melanogaster* possesses six TEP differing in their preference between opsonizing different pathogens [15]. The discovery of 11 TEP proteins in the genome of *B. glabrata* and the multiple alternatives splicing for three of them will generate a high level of protein diversity in this family, increasing in the same time the affinity/avidity of such immune receptors towards potential pathogens. Moreover, biochemical study has shown that BgTEP1 is likely to form homo-tetramers [46], that might also contribute in increasing the binding capacities of these BgTEP molecules by heteromeric recognition complex formation. To conclude, this paper revealed a diversity of TEP family member higher than expected in a gastropod species and could potentially enhance the snail immune repertoire.

**Supplementary Materials:** The following are available online at [www.mdpi.com/2073-4425/11/1/69/s1](http://www.mdpi.com/2073-4425/11/1/69/s1), Figure S1: Alignment file of TEP protein sequences used for the phylogenetic analysis; Figure S2: cDNA sequences of the 11 TEP of *Biomphalaria glabrata*; Figure S3: ClustalW alignment of BgTEP protein sequences with related sequences from other organisms; Figure S4: ClustalW alignment of each BgTEP protein sequence with corresponding protein sequences predicted from BglaB1 genome assembly; Figure S5: BgTEP2 alternative splicing in snail tissues, Table S1: Table list of primers used to measure gene expression by quantitative PCR; Table S2: Table list of protein sequences used to construct phylogenetic tree; Table S3: Genome organization of BgTEP genes; Table S4: Genome organization of BgTEP2 splice variants; File S1: Supplementary materials caption.

**Author Contributions:** R.G. conceived and supervised the work, he also performed bioinformatic, phylogenetic, and gene expression analysis, and wrote the original draft preparation. D.D. helped in conceptualization, performed spliced variant analysis, and help in reviewing the article. B.G. helped in validation of methodological choice and participate in reviewing the paper. D.L. prepared cDNA from snail tissues for gene expression analysis. R.P. and M.L. performed Q-PCR analysis. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding and the APC was funded by ANR (Agence Nationale de la Recherche) in France (project ANR-19-CE11-0016-01). This study is set within the framework of the "Laboratoires d'Excellences (LABEX)" TULIP (ANR-10-LABX-41).

**Conflicts of Interest:** The authors declare that they have no competing interests for this research.

## References

1. Shokal, U.; Eleftherianos, I. Evolution and Function of Thioester-Containing Proteins and the Complement System in the Innate Immune Response. *Front. Immunol.* **2017**, *8*, 759.
2. Theopold, U.; Schmid, M. Thioester-containing proteins: At the crossroads of immune effector mechanisms. *Virulence* **2017**, *8*, 1468–1470.
3. Dodds, M.W.; Law, A.S.K. The phylogeny and evolution of the thioester bond-containing proteins C3, C4 and  $\alpha$ 2-macroglobulin. *Immunol. Rev.* **1998**, *166*, 15–26.
4. Barrett, A.J.; Starkey, P.M. The interaction of alpha 2-macroglobulin with proteinases. Characteristics and specificity of the reaction, and a hypothesis concerning its molecular mechanism. *Biochem. J.* **1973**, *133*, 709–724.
5. Armstrong, P.B. Proteases and protease inhibitors: A balance of activities in host-pathogen interaction. *Immunobiology* **2006**, *211*, 263–281.

6. Armstrong, P.B.; Quigley, J.P. Alpha2-macroglobulin: An evolutionarily conserved arm of the innate immune system. *Dev. Comp. Immunol.* **1999**, *23*, 375–390.
7. Garcia-Ferrer, I.; Marrero, A.; Gomis-Ruth, F.X.; Goulas, T. alpha2-Macroglobulins: Structure and Function. *Sub-Cell. Biochem.* **2017**, *83*, 149–183.
8. Rehman, A.A.; Ahsan, H.; Khan, F.H. alpha-2-Macroglobulin: A physiological guardian. *J. Cell. Physiol.* **2013**, *228*, 1665–1675.
9. Muller-Eberhard, H.J. Complement. *Annu. Rev. Biochem.* **1975**, *44*, 697–724.
10. Ricklin, D.; Reis, E.S.; Mastellos, D.C.; Gros, P.; Lambris, J.D. Complement component C3—The "Swiss Army Knife" of innate immunity and host defense. *Immunol. Rev.* **2016**, *274*, 33–58.
11. Ricklin, D.; Hajishengallis, G.; Yang, K.; Lambris, J.D. Complement: A key system for immune surveillance and homeostasis. *Nat. Immunol.* **2010**, *11*, 785–797.
12. Nonaka, M. The complement C3 protein family in invertebrates. *Invertebr. Surviv. J.* **2011**, *8*, 21–32.
13. Bou Aoun, R.; Hetru, C.; Troxler, L.; Doucet, D.; Ferrandon, D.; Matt, N. Analysis of Thioester-Containing Proteins during the Innate Immune Response of *Drosophila melanogaster*. *J. Innate Immun.* **2010**, *3*, 52–64.
14. Lagueux, M.; Perrodou, E.; Levashina, E.A.; Capovilla, M.; Hoffmann, J.A. Constitutive expression of a complement-like protein in toll and JAK gain-of-function mutants of *Drosophila*. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 11427–11432.
15. Stroschein-Stevenson, S.L.; Foley, E.; O'Farrell, P.H.; Johnson, A.D. Identification of *Drosophila* Gene Products Required for Phagocytosis of *Candida albicans*. *PLoS Biol.* **2006**, *4*, e4.
16. Baxter, R.H.G.; Chang, C.-I.; Chelliah, Y.; Levashina, E.A.; Deisenhofer, J. Structural basis for conserved complement factor-like function in the antimalarial protein TEP1. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 11615–11620.
17. Buresova, V.; Hajdusek, O.; Franta, Z.; Loosova, G.; Grunclova, L.; Levashina, E.A.; Kopáček, P. Functional Genomics of Tick Thioester-Containing Proteins Reveal the Ancient Origin of the Complement System. *J. Innate Immun.* **2011**, *3*, 623–630.
18. Castillo, J.C.; Creasy, T.; Kumari, P.; Shetty, A.; Shokal, U.; Tallon, L.J.; Eleftherianos, I. *Drosophila* anti-nematode and antibacterial immune regulators revealed by RNA-Seq. *BMC Genom.* **2015**, *16*, 519.
19. Shokal, U.; Kopydlowski, H.; Eleftherianos, I. The distinct function of Tep2 and Tep6 in the immune defense of *Drosophila melanogaster* against the pathogen *Photorhabdus*. *Virulence* **2017**, *8*, 1668–1682.
20. Dostálová, A.; Rommelaere, S.; Poidevin, M.; Lemaitre, B. Thioester-containing proteins regulate the Toll pathway and play a role in *Drosophila* defence against microbial pathogens and parasitoid wasps. *BMC Biol.* **2017**, *15*, 79.
21. Blandin, S.; Shiao, S.H.; Moita, L.F.; Janse, C.J.; Waters, A.P.; Kafatos, F.C.; Levashina, E.A. Complement-like protein TEP1 is a determinant of vectorial capacity in the malaria vector *Anopheles gambiae*. *Cell* **2004**, *116*, 661–670.
22. Blandin, S.A.; Marois, E.; Levashina, E.A. Antimalarial Responses in *Anopheles gambiae*: From a Complement-like Protein to a Complement-like Pathway. *Cell Host Microbe* **2008**, *3*, 364–374.
23. Cheng, G.; Liu, L.; Wang, P.; Zhang, Y.; Zhao, Y.O.; Colpitts, T.M.; Feitosa, F.; Anderson, J.F.; Fikrig, E. An in vivo transfection approach elucidates a role for *Aedes aegypti* thioester-containing proteins in flaviviral infection. *PLoS ONE* **2011**, *6*, e22786.
24. Xiao, X.; Liu, Y.; Zhang, X.; Wang, J.; Li, Z.; Pang, X.; Wang, P.; Cheng, G. Complement-related proteins control the flavivirus infection of *Aedes aegypti* by inducing antimicrobial peptides. *PLoS Pathog.* **2014**, *10*, e1004027.
25. Yazzie, N.; Salazar, K.A.; Castillo, M.G. Identification, molecular characterization, and gene expression analysis of a CD109 molecule in the Hawaiian bobtail squid *Euprymna scolopes*. *Fish Shellfish Immunol.* **2015**, *44*, 342–355.
26. Ma, H.; Mai, K.; Xu, W.; Liufu, Z. Molecular cloning of alpha2- macroglobulin in sea scallop *Chlamys farreri* (Bivalvia, Mollusca). *Fish Shellfish Immunol.* **2005**, *18*, 345–349.
27. Xue, Z.; Wang, L.; Liu, Z.; Wang, W.; Liu, C.; Song, X.; Wang, L.; Song, L. The fragmentation mechanism and immune-protective effect of CfTEP in the scallop *Chlamys farreri*. *Dev. Comp. Immunol.* **2017**, *76*, 220–228.
28. Zhang, H.; Song, L.; Li, C.; Zhao, J.; Wang, H.; Gao, Q.; Xu, W. Molecular cloning and characterization of a thioester-containing protein from Zhikong scallop *Chlamys farreri*. *Mol. Immunol.* **2007**, *44*, 3492–3500.

29. Gorbushin, A.M. Immune repertoire in the transcriptome of *Littorina littorea* reveals new trends in lophotrochozoan proto-complement evolution. *Dev. Comp. Immunol.* **2018**, *84*, 250–263.
30. Mitta, G.; Gourbal, B.; Grunau, C.; Knight, M.; Bridger, J.M.; Theron, A. The Compatibility Between *Biomphalaria glabrata* Snails and *Schistosoma mansoni*: An Increasingly Complex Puzzle. *Adv. Parasitol.* **2016**, *97*, 111–145.
31. Pila, E.A.; Li, H.; Hambrook, J.R.; Wu, X.; Hanington, P.C. Schistosomiasis from a Snail's Perspective: Advances in Snail Immunity. *Trends Parasitol.* **2017**, *33*, 845–857.
32. Portet, A.; Pinaud, S.; Tetreau, G.; Galinier, R.; Cosseau, C.; Duval, D.; Grunau, C.; Mitta, G.; Gourbal, B. Integrated multi-omic analyses in *Biomphalaria-Schistosoma* dialogue reveal the immunobiological significance of FREP-SmPoMuc interaction. *Dev. Comp. Immunol.* **2017**, *75*, 16–27.
33. Bouchut, A.; Roger, E.; Gourbal, B.; Grunau, C.; Coustau, C.; Mitta, G. The compatibility polymorphism in invertebrate host/trematodes interactions: Research of molecular determinants. *Parasite* **2008**, *15*, 304–309.
34. Mitta, G.; Adema, C.M.; Gourbal, B.; Loker, E.S.; Theron, A. Compatibility polymorphism in snail/schistosome interactions: From field to theory to molecular mechanisms. *Dev. Comp. Immunol.* **2012**, *37*, 1–8.
35. Theron, A.; Coustau, C. Are *Biomphalaria* snails resistant to *Schistosoma mansoni*? *J. Helminthol.* **2005**, *79*, 187–191.
36. Morand, S.; Manning, S.D.; Woolhouse, M.E. Parasite-host coevolution and geographic patterns of parasite infectivity and host susceptibility. *Proceedings* **1996**, *263*, 119–128.
37. Combes, C. Selective pressure in host-parasite systems. *J. Soc. Biol.* **2000**, *194*, 19–23.
38. Galinier, R.; Portela, J.; Moné, Y.; Allienne, J.F.; Henri, H.; Delbecq, S.; Mitta, G.; Gourbal, B.; Duval, D. Biomphalysin, a New  $\beta$  Pore-forming Toxin Involved in *Biomphalaria glabrata* Immune Defense against *Schistosoma mansoni*. *PLoS Pathog.* **2013**, *9*, e1003216.
39. Moné, Y.; Gourbal, B.; Duval, D.; du Pasquier, L.; Kieffer-Jaquinod, S.; Mitta, G. A large repertoire of parasite epitopes matched by a large repertoire of host immune receptors in an invertebrate host/parasite model. *PLoS Negl. Trop. Dis.* **2010**, *4*, e813.
40. Adema, C.M.; Hertel, L.A.; Miller, R.D.; Loker, E.S. A family of fibrinogen-related proteins that precipitates parasite-derived molecules is produced by an invertebrate after infection. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 8691–8696.
41. Roger, E.; Grunau, C.; Pierce, R.J.; Hirai, H.; Gourbal, B.; Galinier, R.; Emans, R.; Cesari, I.M.; Cosseau, C.; Mitta, G. Controlled Chaos of Polymorphic Mucins in a Metazoan Parasite (*Schistosoma mansoni*) Interacting with Its Invertebrate Host (*Biomphalaria glabrata*). *PLoS Negl. Trop. Dis.* **2008**, *2*, e330.
42. Galinier, R.; Roger, E.; Mone, Y.; Duval, D.; Portet, A.; Pinaud, S.; Chaparro, C.; Grunau, C.; Genthon, C.; Dubois, E.; et al. A multistrain approach to studying the mechanisms underlying compatibility in the interaction between *Biomphalaria glabrata* and *Schistosoma mansoni*. *PLoS Negl. Trop. Dis.* **2017**, *11*, e0005398.
43. Mone, Y.; Ribou, A.C.; Cosseau, C.; Duval, D.; Theron, A.; Mitta, G.; Gourbal, B. An example of molecular co-evolution: Reactive oxygen species (ROS) and ROS scavenger levels in *Schistosoma mansoni/Biomphalaria glabrata* interactions. *Int. J. Parasitol.* **2011**, *41*, 721–730.
44. Wu, X.J.; Dinguirard, N.; Sabat, G.; Lui, H.D.; Gonzalez, L.; Gehring, M.; Bickham-Wright, U.; Yoshino, T.P. Proteomic analysis of *Biomphalaria glabrata* plasma proteins with binding affinity to those expressed by early developing larval *Schistosoma mansoni*. *PLoS Pathog.* **2017**, *13*, e1006081.
45. Portet, A.; Galinier, R.; Pinaud, S.; Portela, J.; Nowacki, F.; Gourbal, B.; Duval, D. BgTEP: An Antiprotease Involved in Innate Immune Sensing in *Biomphalaria glabrata*. *Front. Immunol.* **2018**, *9*, 1206.
46. Bender, R.C.; Bayne, C.J. Purification and characterization of a tetrameric alpha-macroglobulin proteinase inhibitor from the gastropod mollusc *Biomphalaria glabrata*. *Biochem. J.* **1996**, *316 Pt 3*, 893–900.
47. Bender, R.C.; Fryer, S.E.; Bayne, C.J. Proteinase inhibitory activity in the plasma of a mollusc: Evidence for the presence of alpha-macroglobulin in *Biomphalaria glabrata*. *Comp. Biochem. Physiol. B Comp. Biochem.* **1992**, *102*, 821–824.
48. Fryer, S.E.; Bender, R.C.; Bayne, C.J. Inhibition of cysteine proteinase from *Schistosoma mansoni* larvae by alpha-macroglobulin from the plasma of *Biomphalaria glabrata*. *J. Parasitol.* **1996**, *82*, 343–347.
49. Adema, C.M.; Hillier, L.W.; Jones, C.S.; Loker, E.S.; Knight, M.; Minx, P.; Oliveira, G.; Raghavan, N.; Shedlock, A.; do Amaral, L.R.; et al. Whole genome analysis of a schistosomiasis-transmitting freshwater snail. *Nat. Commun.* **2017**, *8*, 15451.

50. Dheilly, N.M.; Duval, D.; Mouahid, G.; Emans, R.; Allienne, J.-F.; Galinier, R.; Genthon, C.; Dubois, E.; Du Pasquier, L.; Adema, C.M.; et al. A family of variable immunoglobulin and lectin domain containing molecules in the snail *Biomphalaria glabrata*. *Dev. Comp. Immunol.* **2015**, *48*, 234–243.
51. Pinaud, S.; Portela, J.; Duval, D.; Nowacki, F.C.; Olive, M.A.; Allienne, J.F.; Galinier, R.; Dheilly, N.M.; Kieffer-Jaquinod, S.; Mitta, G.; et al. A Shift from Cellular to Humoral Responses Contributes to Innate Immune Memory in the Vector Snail *Biomphalaria glabrata*. *PLoS Pathog.* **2016**, *12*, e1005361.
52. Letunic, I.; Bork, P. 20 years of the SMART protein domain annotation resource. *Nucleic Acids Res.* **2018**, *46*, D493–D496.
53. Letunic, I.; Doerks, T.; Bork, P. SMART: Recent updates, new developments and status in 2015. *Nucleic Acids Res.* **2015**, *43*, D257–D260.
54. Penn, O.; Privman, E.; Ashkenazy, H.; Landan, G.; Graur, D.; Pupko, T. GUIDANCE: A web server for assessing alignment confidence scores. *Nucleic Acids Res.* **2010**, *38*, W23–W28.
55. 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.
56. Kumar, S.; Stecher, G.; Li, M.; Niyaz, C.; Tamura, K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol. Biol. Evol.* **2018**, *35*, 1547–1549.
57. Sottrup-Jensen, L.; Sand, O.; Kristensen, L.; Fey, G.H. The alpha-macroglobulin bait region. Sequence diversity and localization of cleavage sites for proteinases in five mammalian alpha-macroglobulins. *J. Biol. Chem.* **1989**, *264*, 15781–15789.
58. Banyai, L.; Patthy, L. The NTR module: Domains of netrins, secreted frizzled related proteins, and type I procollagen C-proteinase enhancer protein are homologous with tissue inhibitors of metalloproteases. *Protein Sci.* **1999**, *8*, 1636–1642.
59. de Bruijn, M.H.; Fey, G.H. Human complement component C3: cDNA coding sequence and derived primary structure. *Proc. Natl. Acad. Sci. USA* **1985**, *82*, 708–712.
60. Gerard, N.P.; Lively, M.O.; Gerard, C. Amino acid sequence of guinea pig C3a anaphylatoxin. *Protein Seq. Data Anal.* **1988**, *1*, 473–478.
61. Prosper, J.Y.A. Characterization of CD109. In *Graduate Department of Medical Biophysics*; University of Toronto: Toronto, YTO, Canada, 2011.
62. Law, S.K.; Dodds, A.W. The internal thioester and the covalent binding properties of the complement proteins C3 and C4. *Protein Sci.* **1997**, *6*, 263–274.
63. Gadjeva, M.; Dodds, A.W.; Taniguchi-Sidle, A.; Willis, A.C.; Isenman, D.E.; Law, S.K. The covalent binding reaction of complement component C3. *J. Immunol.* **1998**, *161*, 985–990.
64. Ma, H.; Wang, B.; Zhang, J.; Li, F.; Xiang, J. Multiple forms of alpha-2 macroglobulin in shrimp *Fenneropenaeus chinensis* and their transcriptional response to WSSV or *Vibrio* pathogen infection. *Dev. Comp. Immunol.* **2010**, *34*, 677–684.
65. Adams, M.D.; Celniker, S.E.; Holt, R.A.; Evans, C.A.; Gocayne, J.D.; Amanatides, P.G.; Scherer, S.E.; Li, P.W.; Hoskins, R.A.; Galle, R.F.; et al. The genome sequence of *Drosophila melanogaster*. *Science* **2000**, *287*, 2185–2195.
66. Holt, R.A.; Subramanian, G.M.; Halpern, A.; Sutton, G.G.; Charlab, R.; Nuskern, D.R.; Wincker, P.; Clark, A.G.; Ribeiro, J.M.; Wides, R.; et al. The genome sequence of the malaria mosquito *Anopheles gambiae*. *Science* **2002**, *298*, 129–149.
67. Falade, M.O.; Otariqho, B. Comparative Functional Study of Thioester-containing Related Proteins in the Recently Sequenced Genome of *Biomphalaria glabrata*. *Iran. J. Parasitol.* **2018**, *13*, 79–88.
68. Liao, H.; Wang, J.; Xun, X.; Zhao, L.; Yang, Z.; Zhu, X.; Xing, Q.; Huang, X.; Bao, Z. Identification and characterization of TEP family genes in Yesso scallop (*Patinopecten yessoensis*) and their diverse expression patterns in response to bacterial infection. *Fish Shellfish Immunol.* **2018**, *79*, 327–339.
69. Matetovici, I.; Van Den Abbeele, J. Thioester-containing proteins in the tsetse fly (*Glossina*) and their response to trypanosome infection. *Insect Mol. Biol.* **2018**, *27*, 414–428.
70. Nielsen, K.L.; Sottrup-Jensen, L.; Nagase, H.; Thogersen, H.C.; Etzerodt, M. Amino acid sequence of hen ovomacroglobulin (ovostatin) deduced from cloned cDNA. *DNA Seq.* **1994**, *5*, 111–119.
71. Nagase, H.; Harris, E.D., Jr. Ovostatin: A novel proteinase inhibitor from chicken egg white. II. Mechanism of inhibition studied with collagenase and thermolysin. *J. Biol. Chem.* **1983**, *258*, 7490–7498.



72. Levashina, E.A.; Moita, L.F.; Blandin, S.; Vriend, G.; Lagueux, M.; Kafatos, F.C. Conserved Role of a Complement-like Protein in Phagocytosis Revealed by dsRNA Knockout in Cultured Cells of the Mosquito, *Anopheles gambiae*. *Cell* **2001**, *104*, 709–718.
73. Batz, T.; Forster, D.; Luschnig, S. The transmembrane protein Macroglobulin complement-related is essential for septate junction formation and epithelial barrier function in *Drosophila*. *Development* **2014**, *141*, 899–908.
74. Lin, L.; Rodrigues, F.; Kary, C.; Contet, A.; Logan, M.; Baxter, R.H.G.; Wood, W.; Baehrecke, E.H. Complement-Related Regulates Autophagy in Neighboring Cells. *Cell* **2017**, *170*, 158–171.
75. Shokal, U.; Kopydlowski, H.; Harsh, S.; Eleftherianos, I. Thioester-Containing Proteins 2 and 4 Affect the Metabolic Activity and Inflammation Response in *Drosophila*. *Infect. Immun.* **2018**, *86*, e00810-17.
76. Pompon, J.; Levashina, E.A. A New Role of the Mosquito Complement-like Cascade in Male Fertility in *Anopheles gambiae*. *PLoS Biol.* **2015**, *13*, e1002255.
77. Guillou, F.; Mitta, G.; Galinier, R.; Coustau, C. Identification and expression of gene transcripts generated during an anti-parasitic response in *Biomphalaria glabrata*. *Dev. Comp. Immunol.* **2007**, *31*, 657–671.
78. Lin, M.; Sutherland, D.R.; Horsfall, W.; Totty, N.; Yeo, E.; Nayar, R.; Wu, X.F.; Schuh, A.C. Cell surface antigen CD109 is a novel member of the alpha(2) macroglobulin/C3, C4, C5 family of thioester-containing proteins. *Blood* **2002**, *99*, 1683–1691.
79. Finnson, K.W.; Tam, B.Y.; Liu, K.; Marcoux, A.; Lepage, P.; Roy, S.; Bizet, A.A.; Philip, A. Identification of CD109 as part of the TGF-beta receptor system in human keratinocytes. *FASEB J.* **2006**, *20*, 1525–1527.
80. Hagiwara, S.; Murakumo, Y.; Sato, T.; Shigetomi, T.; Mitsudo, K.; Tohnai, I.; Ueda, M.; Takahashi, M. Up-regulation of CD109 expression is associated with carcinogenesis of the squamous epithelium of the oral cavity. *Cancer Sci.* **2008**, *99*, 1916–1923.
81. Mii, S.; Hoshino, A.; Enomoto, A.; Murakumo, Y.; Ito, M.; Yamaguchi, A.; Takahashi, M. CD109 deficiency induces osteopenia with an osteoporosis-like phenotype in vivo. *Genes Cells* **2018**, *23*, 590–598.
82. Zhou, S.; da Silva, S.D.; Siegel, P.M.; Philip, A. CD109 acts as a gatekeeper of the epithelial trait by suppressing epithelial to mesenchymal transition in squamous cell carcinoma cells in vitro. *Sci. Rep.* **2019**, *9*, 16317.
83. Collins, A.J.; Schleicher, T.R.; Rader, B.A.; Nyholm, S.V. Understanding the role of host hemocytes in a squid/vibrio symbiosis using transcriptomics and proteomics. *Front. Immunol.* **2012**, *3*, 91.
84. van Lookeren Campagne, M.; Wiesmann, C.; Brown, E.J. Macrophage complement receptors and pathogen clearance. *Cell. Microbiol.* **2007**, *9*, 2095–2102.
85. Alcorlo, M.; Lopez-Perrote, A.; Delgado, S.; Yebenes, H.; Subias, M.; Rodriguez-Gallego, C.; Rodriguez de Cordoba, S.; Llorca, O. Structural insights on complement activation. *FEBS J.* **2015**, *282*, 3883–3891.
86. Williams, M.; Contet, A.; Hou, C.D.; Levashina, E.A.; Baxter, R.H.G. *Anopheles gambiae* TEP1 forms a complex with the coiled-coil domain of LRIM1/APL1C following a conformational change in the thioester domain. *PLoS ONE* **2019**, *14*, e0218203.
87. Zhang, H.; Wang, L.; Song, L.; Zhao, J.; Qiu, L.; Gao, Y.; Song, X.; Li, L.; Zhang, Y.; Zhang, L. The genomic structure, alternative splicing and immune response of *Chlamys farreri* thioester-containing protein. *Dev. Comp. Immunol.* **2009**, *33*, 1070–1076.

