In-depth characterization of trypsin-like serine peptidases in the midgut of the sugar fed Culex quinquefasciatus

André Borges-Veloso¹, Leonardo Saboia-Vahia², Geovane Dias-Lopes¹, Gilberto B. Domont³, Constança Britto¹, Patrícia Cuervo²* and José B. De Jesus¹,4*

Abstract

Background: Culex quinquefasciatus is a hematophagous insect from the Culicidae family that feeds on the blood of humans, dogs, birds and livestock. This species transmits a wide variety of pathogens between humans and animals. The midgut environment is the first location of pathogen-vector interactions for blood-feeding mosquitoes and the expression of specific peptidases in the early stages of feeding could influence the outcome of the infection. Trypsin-like serine peptidases belong to a multi-gene family that can be expressed in different isoforms under distinct physiological conditions. However, the confident assignment of the trypsin genes that are expressed under each condition is still a challenge due to the large number of trypsin-coding genes in the Culicidae family and most likely because they are low abundance proteins.

Methods: We used zymography for the biochemical characterization of the peptidase profile of the midgut from C. quinquefasciatus females fed on sugar. Protein samples were also submitted to SDS-PAGE followed by liquid chromatography–tandem mass spectrometry (LC–MS/MS) analysis for peptidase identification. The peptidases sequences were analyzed with bioinformatics tools to assess their distinct features.

Results: Zymography revealed that trypsin-like serine peptidases were responsible for the proteolytic activity in the midgut of females fed on sugar diet. After denaturation in SDS-PAGE, eight trypsin-like serine peptidases were identified by LC-MS/MS. These peptidases have structural features typical of invertebrate digestive trypsin peptidases but exhibited singularities at the protein sequence level such as: the presence of different amino acids at the autocatalytic motif and substrate binding regions as well as different number of disulfide bounds. Data mining revealed a group of trypsin-like serine peptidases that are specific to C. quinquefasciatus when compared to the culicids genomes sequenced so far.

Conclusion: We demonstrated that proteomics approaches combined with bioinformatics tools and zymographic analysis can lead to the functional annotation of trypsin-like serine peptidases coding genes and aid in the understanding of the complexity of peptidase expression in mosquitoes.

Keywords: Culex quinquefasciatus, Trypsin-like serine peptidases, Zymography, Mass spectrometry
Background

The mosquito *Culex quinquefasciatus* is widespread in tropical and subtropical regions of the world, and it is adapted to urban/peri-urban areas. Despite having anthropophilic and endophilic habits, adult females exhibit high plasticity in their feeding behavior that characterizes this species as an opportunistic insect that feeds on the blood of humans, dogs, birds and livestock. This feature makes this species important in the zoonotic transmission of a wide variety of pathogens between humans and animals [1, 2]. *C. quinquefasciatus* is implicated in the dissemination of several arboviruses such as West Nile virus, St. Louis encephalitis virus, and Venezuelan equine encephalitis virus, and it has also been implicated in the transmission of protozoan parasites such as *Plasmodium relictum*. In addition, this species plays an important role as a vector of helminths such as the causative agent of lymphatic filariasis, *Wuchereria bancrofti*, and the dog heartworm, *Dirofilaria immitis* [3–8].

In mosquitoes, the main proteolytic enzymes responsible for food digestion are trypsin- and chymotrypsin-like serine peptidases as well as carboxy and amino-exopeptidases [9–12]. Trypsin-like peptidases (EC 3.4.21.4) belong to serine peptidases family S1 characterized by the His, Asp, and Ser amino acids residues within the catalytic triad [13]. It has been observed that trypsin-like serine peptidases are a multi-gene family that can be expressed as different isoforms under distinct physiological conditions [14–16]. Whereas the expression of some trypsin genes is constitutive, the expression of other trypsin genes is induced by the blood meal; hence, the expression pattern of trypsin-coding genes is biphasic [10, 16–22]. However, the confident assignment of the trypsin genes that are expressed under each condition is still a challenge due to the large number of trypsin-coding genes in the Culicidae family [7, 14, 23]. For example, 380 serine peptidase genes were reported in the genome of *Aedes aegypti* [14], but only six trypsin-like enzymes have been characterized at the protein level in the midgut tissue [12, 17, 24].

In addition to their role in food digestion, trypsin-like serine peptidases have been described as key mediators of pathogen-vector interaction. Among several midgut trypsin isoforms in *Ae. aegypti*, only one could limit Dengue virus-2 (DENV-2) infectivity [25]. Although the proteolytic environment of the midgut lumen could lead to pathogen degradation and consequently limit infectivity, arboviruses from different families such as DENV-2 (Flaviviridae), La Crosse virus (Bunyaviridae) and blue tongue virus (Reoviridae) use vector midgut peptidases for the proteolytic processing of virion surface proteins, increasing viral binding to midgut cells [26–30]. In addition, *Ae. aegypti* secreted trypsin peptidases activate a *Plasmodium gallinaceum* chinidase that is essential for peritrophic matrix evasion [31, 32]. Thus, not only the time course and the quantity of peptidase expression in the initial time of feeding does influence the infection, replication and dissemination of pathogens, but the quality of these peptidases could also be important for this interaction.

In the midgut of *C. quinquefasciatus*, trypsin-like serine peptidases have been detected after blood feeding [33]. However, the “peptidase status” of the midgut when the blood arrives into the lumen corresponds to that set by sugar feeding. Nevertheless, the expression of peptidase genes in the midgut of mosquitoes fed on sugar as well as the identity of the peptidases expressed (if any) when mosquitoes are feeding on sugar remains elusive, most likely because of the abundance of these enzymes is not enough to detect them [15, 24].

Herein, we focused on the characterization and identification of trypsin-like serine peptidases constitutively expressed in the midgut of females of *C. quinquefasciatus* that were fed only sugar. We used zymography for the biochemical characterization of the enzymes and SDS-PAGE followed by liquid chromatography–tandem mass spectrometry (LC–MS/MS) analysis for protein identification. Eight trypsin-like serine peptidases were identified by MS/MS and their molecular features were analyzed by bioinformatic tools.

Methods

Chemicals

All reagents were purchased from Sigma (St. Louis, MO, USA) or Merck (São Paulo, SP, Brazil). MilliQ-purified water (Millipore Corp., Bedford, MA, USA) was used to prepare all of the solutions.

Insects

Experiments were carried out using 5-day-old *C. quinquefasciatus* female adults (Colônia strain) from a closed colony reared in the Laboratório de Fisiologia e Controle de Artrópodes Vetores - Instituto Oswaldo Cruz, FIOCRUZ, Rio de Janeiro. Larvae of *C. quinquefasciatus* were reared in plastic basins (33 × 24 × 8 cm) containing 1 L of dechlorinated water and 1 g of cat food (FriskiesW, Purina, Camaquã/RS). Larvae were kept in a biological oxygen demand incubator (BOD) at 25 ± 1 °C, with a relative humidity of 60 ± 10 % and a light:dark photoperiod of 14:10 h. The adult mosquitoes were maintained on a 10 % sucrose diet.

Midgut dissection

The mosquitoes were anesthetized on ice and decapitated. Dissection was performed in cold PBS buffer, pH 7.4 (150 mM NaCl, 10 mM Na₂HPO₄). The thorax of each decapitated mosquito was immobilized with forceps (#5) and the gut, Malpighian tubules and gonads
were dissected by gently pulling at the eighth abdominal segment region with another pair of forceps. The Malpighian tubules, hindgut and gonads were cut away, and the midguts were delicately washed twice with PBS buffer and transferred to a microcentrifuge tube containing the specific lysis buffers for proteome or zymography analysis. In addition, optical differential interference contrast microscopy (DIC) was used to record images from different midgut dissected samples obtained for both zymographic analysis and mass spectrometry. Such images were made with the main objective to verify the quality of the midgut dissections, *i.e.* to verify if the cuts were done consistently in the same regions of the gut and also to rule out the possibility of contaminations with metamorphosis remaining tissues into the midgut lumen.

**Zymography assays**

A pool of 20 midguts were lysed with a VWR\textsuperscript* disposable pellet mixer and cordless motor, and homogenized in a plastic eppendorff microtube containing a lysis buffer with 10\% glycerol, 0.6 \% Triton X-100, 100 mM Tris–HCl pH 6.8 and 150 mM NaCl. The homogenate was centrifuged at 14,000 g at 4 °C for 15 min, and the supernatant was collected. The protein concentration of the resulting extracts was determined using the Pierce 660 nm Protein assay (Thermo Scientific). For protein separation, 10 \mu g of protein were loaded in 10 % polyacrylamide gels copolymerized with 0.1 % porcine gelatin as the substrate. Electrophoresis was performed at 4 °C at a constant voltage of 110 V. Peptidase activity was detected as previously reported with few modifications [34]. Briefly, 10 \mu g of protein from the midgut were diluted in 100 mM sodium acetate (at pH 7.5 or 5.5), 100 mM Tris–HCl (pH 7.5 or 10.0) with or without inhibitor addition. The reactions were initiated as described previously [34]. The reactions were carried out in triplicate. The effects of pH and peptidase inhibitors on the proteolytic activities of midgut homogenates were also evaluated *in vitro* using the fluorogenic substrate 7-amido-4-methylcoumarin hydrochloride (Z-Phe-Arg-AMC). For each assay, 100 \mu M of substrate were used. The reactions were initiated as described previously [34]. Briefly, 10 \mu g of protein from the midgut were diluted in 100 mM sodium acetate (at pH 3.5 or 5.5), 100 mM Tris–HCl (pH 7.5 or 10.0) and the peptidases were resolved as described above.

The results were derived from three independent experiments carried out in triplicate.

**In vitro enzyme assays**

The effects of pH and peptidase inhibitors on the proteolytic activities of midgut homogenates were also evaluated *in vitro* using the fluorogenic substrate 7-amido-4-methylcoumarin hydrochloride (Z-Phe-Arg-AMC). For each assay, 100 \mu M of substrate were used. The reactions were carried out in triplicate. The effects of pH and peptidase inhibitors on the proteolytic activities of midgut homogenates were also evaluated *in vitro* using the fluorogenic substrate 7-amido-4-methylcoumarin hydrochloride (Z-Phe-Arg-AMC). For each assay, 100 \mu M of substrate were used. The reactions were initiated as described previously [34]. Briefly, 10 \mu g of protein from the midgut were diluted in 100 mM sodium acetate (at pH 3.5 or 5.5), 100 mM Tris–HCl (pH 7.5 or 10.0) and the peptidases were resolved as described above.

The results were derived from three independent experiments carried out in triplicate.

**Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE), protein digestion and peptide extraction**

Fifty pooled midguts were directly lysed in Laemmli sample buffer containing 4 \% SDS, 20 \% glycerol, 10 \% 2-mercaptoethanol, 0.004 \% bromophenol blue and 0.125 M Tris–HCl, pH approx. 6.8. Lysis was performed by mechanical homogenization using a plastic pestle. The lysate was centrifuged twice at 14,000 g at 4 °C and the proteins in the resulting supernatant were collected. The protein concentration was determined using the Pierce 660 nm Protein assay (Thermo Scientific). Then, the samples were heated for 5 min in a boiling water bath and separated by 12 \% SDS-PAGE, 30 \% acrylamide, 0.8 \% bis-acrylamide. Proteins were stained using Coomassie Brilliant Blue and photographed. Three gels from three independent midgut suspensions were performed. Proteins were enzymatically digested following procedures previously described [35] with some modifications. Briefly, fine slices from each protein lane were manually excised and de-stained three times in 400 \mu L of 50 \% acetonitrile, 25 mM NH$_4$HCO$_3$ pH 8.0 for 15 min. Proteins were subsequently reduced and alkylated using 65 mM dithiothreitol (DTT) and 200 mM iodoacetamide, respectively. Gel slices were washed with 100 mM NH$_4$HCO$_3$ followed by dehydration with acetonitrile. Slices were rehydrated with a solution of 20 ng/\mu L of sequencing grade
modified porcine trypsin (Promega, USA) in 50 mM NH₄HCO₃ and incubated overnight at 37 °C. Peptides were extracted using 0.1 % formic acid in 50 % v/v acetonitrile, desalted and concentrated with Poros oligo R3 C18 resin (Applied Biosystems, USA). The eluted peptides were loaded in a nano-high performance liquid chromatography (nanoHPLC) in-line with a hybrid linear trap quadrupole (LTQ) Orbitrap mass spectrometer.

**Mass spectrometry analysis**

For each sample 4 μL of peptides solution (0.1 % formic acid) were applied to an EASY II-nanoHPLC system (Thermo Fisher Scientific) coupled online to an electrospray (ESI)-LTQ-Orbitrap Velos mass spectrometer (Thermo Fisher Scientific). Peptides were eluted through a trap column (150 μm × 2 cm) packed in-house with C-18 ReproSil 5 μm resin (Dr. Maisch) and an analytical column (100 μm × 15 cm) packed in-house with C-18 ReproSil 3 μm resin (Dr. Maisch) using a mobile phase A of 0.1 % (v/v) formic acid in water and a mobile phase B 0.1 % (v/v) formic acid in acetonitrile. Gradient conditions were as follows: 5 to 40 % B in 180 min. Mass spectra were acquired in the positive mode using a data-dependent automatic (DDA) survey MS scan and tandem mass spectra (MS/MS) acquisition. Each DDA consisted of a survey scan in a 300 – 2000 m/z range and resolution 60000 with a target value of 1 × 10^6 ions. Each survey scan was followed by the MS/MS of the 10 most intense ions in the LTQ using collision-induced dissociation (CID). Ions previously fragmented were dynamically excluded for 60 s.

**Database searching**

Mass spectra were searched against a customized non-redundant database including sequences of all Culicidae species available at UniRef100 (101,993 sequences, downloaded May 2015, http://uniprot.org) using the Mascot MS/MS ion search engine (Matrix Science, Oxford, UK, version 2.4.1). The search parameters in the Mascot server were as follows: lack of taxonomic restrictions; one tryptic missed cleavage; carbamidomethylation of cysteine residues as a fixed modification and oxidation of methionine and acetylation as variable modifications; 10 ppm mass tolerance for the MS mode and 0.5 Da tolerance for its corresponding MS/MS fragments. Scaffold (version 4.3.0, Proteome Software Inc., Portland) was used to validate MS/MS peptide and protein identifications. Peptide identifications were accepted at 95.0 % probability by the Peptide Prophet algorithm [36] using the Scaffold delta mass correction. Protein identifications were accepted at 95.0 % probability and if they were supported by two or more independent pieces of evidence (e.g., identification of a peptide with different charge states, a modified and a non-modified version of the same peptide, or two different peptides). Protein probabilities were assigned by the Protein Prophet algorithm [37].

To confirm peptidase identifications, mass spectra were also analyzed using the ProLuCID 1.3 engine at the PatternLab platform [38] against the same customized database. Searches were performed with one missed cleavage, with carbamidomethylation of cysteine residues as a fixed modification, methionine oxidation as a variable modification and mass tolerances of 40 ppm and 0.5 Da for precursor and fragment ions, respectively. The validity of the peptide sequence matches (PSMs) was assessed using the Search Engine Processor (SEPro) at the PatternLab platform [39].

**Multiple sequence alignment and bioinformatics analysis**

The complete amino acid sequences of the peptidases identified by mass spectrometry were fully retrieved from the VectorBase database (http://biomart.vectorbase.org) [40]. Multiple sequence alignments were performed using CLUSTAL Omega [41]. FASTA sequences of all trypsin identified by mass spectrometry were compared against well annotated sequences of bovine chymotrypsinogen (CTRA_BOVIN), bovine trypsinogen (TRY1_BOVIN), Ae. aegypti trypsin 3A1 (TRY3_AEDAE) and An. gambiae trypsin-6 (TRY6_ANOGA). The amino acid sequence of each identified trypsin was scanned for various domains and motifs. The residues at the active site (His, Asp, Ser), the signal peptide, the conserved cysteine residues of disulfide bounds and the protein size of precursor and mature forms of peptidases were detected using the PROSCAN function of the PROSITE suite (http://prosite.expasy.org) [42]. The signal peptide was also predicted by SignalP 4.0 (http://www.cbs.dtu.dk/services/SignalP) [43]. To predict N-glycosylation and O-glycosylation sites, amino acid sequences were analyzed using the NetNGlyc 1.0 Server (http://www.cbs.dtu.dk/services/NetNGlyc) [44] and NetOGlyc 4.0 Server (http://www.cbs.dtu.dk/services/NetOGlyc) [45], respectively. To identify species-specific trypsin we used the Skyline software (http://proteome.gs.washington.edu/software/skyline) [46] to search against the same database of Culicidae sequences used for proteomic analysis.

**Results and discussion**

Zymographic assays revealed a complex serine peptidase profile in the midgut of *C. quinquefasciatus* females composed of at least eleven bands of proteolytic activity (Fig. 1). Among these bands, 3 migrated at 28 to36 kDa, which is the expected molecular mass for monomeric trypsin [12, 17, 20, 47]. In addition, trypsin activities at high molecular mass regions such as 55, 80 and 130 kDa as well as at low molecular mass regions below 20 kDa were observed (Fig. 1). These activities could be due to
(i) sample preparation, *i.e.*, protein samples are not boiled in the presence of SDS and β-mercaptoethanol, therefore peptidases are not completely denatured or reduced, enabling protein aggregation and/or oligomerization that slows the electrophoretic migration; and (ii) interaction of peptidases with the substrate could also account for the slow migration [48, 49]. Despite such factors that could impede the regular migration of the peptidases, we cannot rule out the possibility that some peptidases could be extensively processed at the post-translational level, increasing their apparent molecular mass in the gel and allowing their association with other proteins in the midgut extract [49–51]. Despite these possibilities, zymographic analysis is a highly reproducible method for the study of the proteolytic profiles in different Culicidae species, suggesting that such high molecular mass enzymes are common findings and that they are not experimental artifacts [34, 52, 53]. Similarly, other authors have observed such results in other insects [54].

To further characterize the profile of proteolytic activities in the midgut of *C. quinquefasciatus* females, we performed a time-course analysis of peptidase activities over a range of 2-12 h. Although proteolytic bands began to be visualized from 6 h of incubation, the complete profile was detected after 12 h (Fig. 1). These results differ from the proteolytic activities in larval stages of *C. quinquefasciatus* [34] where activities were detected at 2 h of incubation. Such difference may be due to the fact that the larval midgut exhibits high peptidase activities that are more easily detected due to the high and constant feeding activity of larvae. Conversely, there is little peptidase activity in midgut of a sugar-fed *Culex* adult female, possibly because the insect does not need it. Such results indicate that different life stages of *C. quinquefasciatus* exhibit stage-specific proteolytic profiles, which may be related to qualitative and quantitative differential expression of peptidases according to the feeding behavior.

The proteolytic activities were evaluated for pH dependence and sensitivity to inhibitors. Although weak peptidase activities are observed at acid pH, the activities increased at alkaline pH between 7.5 and 10.0 (Fig. 2). We observed high proteolytic activities at pH 10, but several bands overlapped, which impeded an accurate analysis of the proteolytic profile by zymography (Fig. 2a). For this reason, all subsequent assays were conducted at pH 7.5. Although at pH 10 we could hardly detect bands different from those observed at pH 7.5, we cannot rule out the possibility that other peptidase activities could be present at pH 10. In addition, the effect of pH on peptidase activities was also analyzed using a fluorogenic substrate (Fig. 2b). This assay corroborated the results observed by zymography and allowed a quantitative analysis of proteolytic activities at the distinct pH. In agreement with the results obtained previously for other Diptera, trypsin-like serine peptidases of *C. quinquefasciatus* are highly active at alkaline pH [34, 52, 53, 55]. PMSF, a specific inhibitor of serine peptidases, revealed that the profile of active peptidases expressed in the midgut of females fed a sugar diet is due to serine peptidases (Fig. 3). To determine whether such activities are specifically due to trypsin- or chymotrypsin-like serine peptidases, specific inhibitors TLCK and TPCK, respectively, were used. All activity bands were strongly inhibited by TLCK, indicating that the serine peptidases detected here belong to the trypsin-like family. In addition, *in vitro* assays confirmed the results obtained in the zymographic analysis. Proteolytic activities were inhibited by PMSF and TLCK but not by E-64 or TPCK, inhibitors of cysteine peptidases or chymotrypsin-like serine peptidases, respectively. That means that the proteolytic profile detected under the conditions here analyzed is due to trypsin-like serine peptidases. These results agree with previous reports on the expression of trypsin-like serine peptidases in the midgut of other Culicidae [12, 14, 17, 24].
The *C. quinquefasciatus* genome codes for 403 putative trypsin-like serine peptidase genes [14], but it is unknown which of them are expressed in the midgut tissue. Here, we identified seven trypsin-like serine peptidases constitutively expressed in the midgut of females fed a sugar diet using two independent search engines: Mascot (followed by Scaffold validation) and ProLuCID (Table 1). In addition, one trypsin-like serine peptidase was exclusively identified by MASCOT (followed by Scaffold validation) based on one peptide and one spectrum evidence (Table 1, B0WW44, gray filled). Interestingly, the SDS-PAGE bands where peptidases were identified by MS/MS coincide with the zymographic regions where peptidase activities were observed (Fig. 4). Although most of the peptidases were identified in electrophoretic bands migrating between 25 to 40 kDa (Fig. 4), the Trypsin5 and Trypsin7 were the only ones found in the high molecular mass region (Fig. 4). Conspicuously, these enzymes exhibited predicted sites for N-Glycosylation. Particularly, Trypsin5 also present predicted O-Glycosylation sites and transmembrane regions (Table 3). Such features could in fact alter the migration pattern of the mature protein. Nevertheless, as sample preparation for each electrophoresis is different, a comparison of peptidase mobility is difficult, but, in any case, both methodologies serve for mapping the identified peptidases. All identified trypsin proteins matched with *C. quinquefasciatus* protein sequences. The alignment of the full sequence of the peptidases identified by mass spectrometry showed several structural features typical of invertebrate digestive trypsin peptidases: (i) the conserved histidine, aspartic acid and serine residues forming the catalytic triad; (ii) six cysteine residues at conserved positions involved in the forming of disulfide bonds; (iii) the signal peptide sequence; (iv) the putative autocatalytic activation motifs immediately after an arginine or lysine residue (R/K-IVGG); (v) the motifs characteristic of active peptidases LTHAAC, DIAL, and GDSGGP (Fig. 5, Table 2) [56]. Interestingly, some trypsin peptidases identified here have distinct features. For example, we observed that the autocatalytic motif of Trypsin 4 has a His residue instead of R/K residues, which could suggest that this enzyme has a specific signal for activation. In addition, the activation motifs in Trypsin 5, IIGG, and cationic trypsin, VVGG, differ by one amino acid residue from the classical motif sequence (IVGG) [57, 58].

An important difference between vertebrate and invertebrate trypsin is the number and location of disulfide bonds. Vertebrate trypsins commonly have six disulfide bonds. Fig. 2 Effect of pH on the proteolytic activities of midgut extracts from *C. quinquefasciatus* females fed on sugar. a. The pH influence was evaluated by incubation of protein extracts at 37 °C for 12 h in 0.1 M sodium acetate buffer pH 3.5, 5.5 or 0.1 M Tris–HCl buffer pH 7.5, 10.0. The numbers on the left indicate the molecular mass of standards utilized in the gel (kDa). b. In-solution assays were performed using the fluorogenic substrate Z-Phe-Arg-AMC in 0.1 M sodium acetate buffer pH 3.5, 5.5 or 0.1 M Tris–HCl buffer pH 7.5 or 10.0.
bonds, whereas, in general, trypsins from insects and crustaceans have only three disulfide bonds at conserved positions, close to the active site [56, 59]. The alignment of trypsin peptidases identified here shows that five of them, Trypsin1, Trypsin2, Trypsin4, Serine protease SP24D and Serine protease ½ have three disulfide bonds while the Trypsin5 and Trypsin7 have four disulfide bonds and the Cationic trypsin only has one. Although the number of disulfide bonds is different in the trypsins identified here, the role of the disulfide bonds is crucial for the tridimensional structure of the enzymes and, consequently, for their activity [56, 60].

Trypsin-like serine peptidases identified here were further analyzed regarding their predicted cellular location, presence of transmembrane helices, and glycosylation motifs using bioinformatics tools (Table 3). Analysis of the prediction of cellular location using the Target P server indicates that all trypsin peptidases are secreted enzymes, which is a typical feature of the digestive enzymes found in the midgut lumen [14, 15, 17]. However, the prediction of transmembrane helices, using the TMHMM server, revealed that Trypsin4 and Trypsin5 have one transmembrane domain (Table 3) suggesting that these enzymes could be targeted to the midgut membrane [55, 61]. Although glycosylation is not a common post-translational modification in trypsin, some glycosylation motifs have been observed in invertebrate trypsin [15]. Five of the eight trypsin isoforms identified here have predicted sites for O- and N-glycosylation (Table 3, Fig. 5). In agreement with this observation, trypsins peptidases from An. gambiae could be glycosylated and that such modification might be required for the association of peptidases with peritrophins in the peritrophic membrane [51].

Using VectorBase we analyzed the structure of the genes encoding the trypsin-like serine peptidases identified here. We observed that the exon number of the trypsin coding genes varies from one to three. With the exception of the Serine protease ½ that has an intron with 298 nucleotides, the other intron sequences are shorter than those observed in trypsin genes of vertebrates, varying between 25 and 71 nucleotides. Our analyses show that the intron exon structure is not conserved between all trypsin identified here, suggesting that several events of intron loss and gain have occurred in this species, which is in agreement with previous...
| Identified Proteins | Accession Number | Molecular Weight | Mascot exclusive peptides | Mascot total spectra | Mascot Coverage % | Peptide sequence identified by MASCOT | Mascot Ion score | Peptide sequences identified by ProLuCID | ProLuCID unique peptides | ProLuCID total spectra | ProLuCID Coverage % |
|---------------------|------------------|------------------|--------------------------|---------------------|-------------------|----------------------------------------|----------------|------------------------------------------|---------------------|---------------------|-------------------|
| Trypsin 4 OS = Culex quinquefasciatus GN = CpipJ_CPIJ017414 | B0XCW2_CULQU | 28 kDa | 3 | 13 | 18 | (R)VGSSYDYQGGTVIDVA GMTIHPR(Y) (K)DFDFALLRL(K) (K)GCACPDDYGYVADVEK(A) | 35.21 | (R)VGSSYDYQGGTVIDVA GMTIHPR(Y) (K)DFDFALLRL(K) (K)GCACPDDYGYVADVEK(A) | 6 | 29 | 28 |
| Trypsin 7 OS = Culex quinquefasciatus GN = CpipJ_CPIJ017964 | B0XES8_CULQU | 27 kDa | 4 | 14 | 18 | (R)GGQLIAVTR(K) (R)DYALLNLAK(S) (R)AVDVPIADHDR(C) (K)DACLGDSGGPLTCG K(V) | 53.31 | (R)GGQLIAVTR(K) (R)DYALLNLAK(S) (R)AVDVPIADHDR(C) (K)DACLGDSGGPLTCG K(V) | 5 | 25 | 18 |
| Trypsin-5 OS = Culex quinquefasciatus GN = CpipJ_CPIJ015103 | B0X667_CULQU | 30 kDa | 3 | 7 | 16 | (K)IIGGFPAAEQGDTLHQSIR(F) (K)GCGLAAAPGYIYSDVAYYR(G) (R)WGWDSCLAGK(C) | 35.64 | (K)IIGGFPAAEQGDTLHQSIR(F) (K)GCGLAAAPGYIYSDVAYYR(G) (R)WGWDSCLAGK(C) | 3 | 14 | 16 |
| Trypsin-1 OS = Culex quinquefasciatus GN = CpipJ_CPIJ007079 | B0WIS4_CULQU | 29 kDa | 3 | 6 | 18 | (R)IVGGFIADAPHQVSL QSR(G) (K)HASGGSVSIKR(K) (R)AAAYVPAPYNNQNCNSA YAR(Y) | 51.51 | (R)IVGGFIADAPHQVSL QSR(G) (K)HASGGSVSIKR(K) (R)AAAYVPAPYNNQNCNSA YAR(Y) | 6 | 10 | 34 |
| Table 1 | Trypsin-like serine peptidases identified by mass spectrometry in the midgut of Culex quinquefasciatus females fed on sugar (Continued) |
|---|---|
| Trypsin 2 OS = Culex quinquefasciatus GN = CpipJ_CPU005273 | |
| B0WE94_CULQU 28 kDa 2 2 12 | (K)DACQGDSGGGPLVADGKL 33.42 (R)NTIDYDYSLELKL(S) |
| | (R)GSICGGSILSPK(W) |
| | (K)WILTAHCTDGASVSNLR(I) |
| | 5 13 30 |
| Serine protease 1/2 OS = Culex quinquefasciatus GN = CpipJ_CPU003826 PE = 3 SV = 1 | |
| B0W959_CULQU 30 kDa 2 3 18 | (R)LEFGHAVQPVDLVR(D) 19.14 (R)LEFGHAVQPVDLVR(D) |
| | (R)DEPADESQSLVSGWG 27.7 |
| | (R)DEPADESQSLVSGWDTR(S) |
| | (R)WLTAHCTENTDAGISVR(V) |
| | (R)GLVLVPLVNR(E) |
| | (K)LGMPVTESMICAGFAK(E) |
| | 2 3 10 |
| Serine protease SP24D OS = Culex quinquefasciatus GN = CpipJ_CPU015368 | |
| B0X870_CULQU 27 kDa 1 1 10 | (R)TGETFVDNQATVSFG 35.91 (R)TGETFVDNQATVSFG |
| | (R)Gr(T) |
| | (R)TVDGGPVSEPTK(N) 35.12 |
| | (R)TVDGGPVSEPTK(N) |
| | 1 3 10 |
| Cationic trypsin OS = Culex quinquefasciatus GN = CpipJ_CPU011378 | |
| B0WW44_CULQU 26 kDa 1 1 8 | (K)LGSEIEYDELSQIALYEGDLPK(D) 34.98 |
| | (K)LGSEIEYDELSQIALYEGDLPK(D) |
| | (R)WHPOQYAEGRNLADIAVIRV 32.92 |
observations in other species [56, 62–64] (Table 3). In addition, the number of paralogues of each peptidase identified here was verified (Table 3). According to this analysis, the trypsin peptidases identified here have between 13 and 38 paralogues. The Trypsin1, 2, 4, 7, SP24D and Cationic trypsin are paralogues among them, suggesting that these peptidases were originated by gene duplication [14, 23, 65, 66]. In addition, the database mining shows that trypsin coding genes are generally clustered. For example, according to VectorBase, Trypsin1 is clustered with five other trypsin genes. It was suggested that the ancestors of dipterans had only one trypsinogen gene and that extra copies were gained by gene duplication [67]. In Culicidae, many trypsin-like serine peptidase coding genes are clustered in tandem arrays in different chromosomes, indicating that tandem duplication plays an important role in the expansion of this gene family [14, 20, 22, 54]. The expression of trypsin coding genes in this mosquito represents a substantial challenge for the assignment of putative functions, for determining their precise localization and mechanisms of regulation of expression. In fact, the understanding of the peptidase tissue expression patterns may be useful for the assignment of the putative function of such peptidases [64]. Thus, the use of techniques for the identification of active and tissue-specific peptidases in the midgut, as performed here, contributes for such function assignment.

The identification of active trypsin peptidases in C. quinquefasciatus females fed on sugar is in agreement with previous reports of our group that showed that Ae. albopictus females fed on sugar express active forms of trypsin [53]. The presence of active trypsin peptidases in sugar fed females of other mosquito species has been reported [20, 68, 69]. Several hematophagous diptera express a series of constitutive and blood meal-induced trypsin genes in the gut [16, 20, 22, 54]. The expression
of peptidases in the midgut of sugar-fed females may represent the induction of enzymes that was involved in the digestion of the larval/pupal meconium, or still induced by commensal bacteria into the midgut lumen [70]. In addition, because nondiaposing anautogenous mosquitoes need to feed on blood to complete the gonotrophic cycle, it is reasonable that they prepare their midgut tissue for blood digestion prior to blood feeding, so it is not surprising that after five days of adult emergence they express active trypsin peptidases. In fact, trypsin-coding genes were downregulated in anautogenous diapause-destined females. At the end of the diapause period (2–3 months at 18 °C), the expression of digestive peptidases increases, preparing the females for blood meal uptake [71]. Thus, the constitutive expression of trypsin peptidases could guarantee an efficient midgut metamorphosis and digestion of the blood meal, probably by zymogen activation, leading to improved biological fitness [55–57].

Despite molecular approaches that have allowed the identification of trypsin coding genes, the confirmation of the presence of these enzymes at protein level under different physiological conditions has not been reached, most likely because they are low abundance proteins [15, 24]. With the aim of analyzing whether we can develop selected reaction monitoring (SRM) experiments for detection of specific C. quinquefasciatus trypsin peptidases, we used the SKYLINE software for determining the occurrence of proteotypic peptides in these enzymes. SRM is a powerful method for monitoring target peptides within a complex protein sample and is particularly useful for hypothesis driven proteomics [72, 73]. Despite the presence of conserved motifs in the trypsin peptidases, the SKYLINE output shows that five out of the eight trypsin peptidases identified by mass spectrometry in our study have differences in amino acid sequences that allow the detection of unique peptides (Table 4). Remarkably, these unique peptides were the same identified by mass spectrometry in our study. The methodology used here for identifying proteotypic peptides can be used for developing SRM mass spectrometry assays for finding different trypsin peptidases in specific tissues or under specific stimulus. Noticeably, when we used the SKYLINE considering not only the sequences of C. quinquefasciatus trypsin peptidases but also the sequences of those peptidases from other species with known genome sequences, such proteotypic peptides are both peptidase-specific and species-specific. This result is not conclusive because those genomes are not well annotated.

### Table 2

| Uniprot accession number | Protein name      | Active site triad position | Cysteine pair residues | Residues conferring substrate specificity | Protein size (aa) | Activation site | Conserved regions |
|--------------------------|-------------------|---------------------------|-----------------------|------------------------------------------|------------------|----------------|-----------------|
| B0WIS4                   | Trypsin 1         | His88, Asp133, Ser229     | 73-89, 198-214, 225-249 | Asp223, Gly246,Gly256                     | 274              | 226            | YR^IVGG        |
| B0WE94                   | Trypsin 2         | His75, Asp120, Ser216     | 60-76, 183-200, 212-236 | Asp210, Gly234,Gly244                     | 261              | 226            | GK^IVGG        |
| B0XCV2                   | Trypsin 4         | His70, Asp116, Ser213     | 151-219, 181-198, 209-233 | Asp207, Gly230,Gly240                     | 258              | 233            | FH^IVNG        |
| B0X667                   | Trypsin 5         | His73, Asp127, Ser221     | 58-74, 160-227, 192-208, 217-241 | Asp215, Gly238,Gly248                   | 293              | 268            | PK^IIGG        |
| B0XES8                   | Trypsin 7         | His67, Asp112, Ser207     | 52-68, 146-213, 176-192, 203-227 | Asp201, Gly224,Gly234                   | 252              | 229            | SR^IVNG        |
| B0X870                   | SP24D             | His63, Asp109, Ser195     | 48-64, 172-181, 191-216 | ?                                        | 240              | 217            | RR^IFGG        |
| B0WW59                   | Serine protease 1/2 | His92, Asp135, Ser233    | 77-93, 202-217, 229-259 | ?                                        | 283              | 235            | SR^IVNG        |
| B0WW44                   | Cationic trypsin  | His69, Asp114, Ser202     | 54-70                  | ?                                        | 244              | 216            | GR^VGGG        |

Δ Extracted after CLUSTAL Omega alignment analysis
(aa) total number of amino acid residues
? = other residues different than DGG
^ = Indicates the cleavage site for zymogen activation
| Uniprot accession number | Protein name | Target P prediction ① | Signal P prediction ② | TMHMM prediction ③ | N-Glycosylation prediction ④ | O-Glycosylation prediction ⑤ | Exon number ⑥ | Paralogues number ⑥ | Supercontig ⑥ |
|--------------------------|--------------|------------------------|-----------------------|------------------------|-----------------------------|-----------------------------|----------------|----------------------|----------------|
| B0WIS4                   | Trypsin 1    | S (0.901)              | 23^24 (0.761)         | No                     | 158-NETV (0.7243)           | 36-T (0.62795)/40-S (0.689668) | 1              | 36                   | 3.14            |
| B0WE94                   | Trypsin 2    | S (0.910)              | 18^19 (0.818)         | No                     | No                          | No                          | 2              | 36                   | 3.94            |
| B0XCW2                   | Trypsin 4    | S (0.973)              | inside: 1-6/Tmhelix: 7-26/outside: 27-258 | 27-NGTQ (0.8040) | No                          | No                          | 2              | 36                   | 3.91            |
| B0X667                   | Trypsin 5    | S (0.952)              | inside: 291-293/Tmhelix: 268-290/outside: 1-267 | 65-NRTV (0.6702)/183-NVT (0.8306) | 151-T (0.653105)/159-S (0.523482) | 2              | 36                   | 3.11            |
| B0XE85                   | Trypsin 7    | S (0.832)              | 21^22 (0.665)         | No                     | 106-NVT (0.6360)           | No                          | 3              | 36                   | 3.66            |
| B0X870                   | SP24D        | S (0.891)              | 20^21 (0.801)         | No                     | 69-NGSV (0.6998)/75-NLSV (0.6183) | No                          | 2              | 24                   | 3.54            |
| B0W959                   | Serine protease 1/2 | S (0.960) | 26^27 (0.750)         | No                     | 44-S (0.785129)            | No                          | 2              | 38                   | 3.33            |
| B0WW44                   | Cationic trypsin | S (0.926) | 20^21 (0.794)         | No                     | No                          | No                          | 3              | 24                   | 3.33            |

① TargetP 1.1 Server. Prediction of the subcellular location of trypsin. S = secreted. The number into the parenthesis indicates the probability
② SignalP 4.0 Server. Prediction of presence and location of signal peptide cleavage sites in the trypsin sequences. The number indicates the number of amino acid residues involved in the cleavage. The number into the parenthesis indicates the probability
③ TMHMM 2.0 Server. Prediction of transmembrane helices in proteins. Tmhelix: transmembrane helix
④ NetNglyc 1.0 Server. Prediction of N-Glycosylation sites based on the presence of Asn-Xaa-Ser/Thr motifs. The number into the parenthesis indicates the probability
⑤ NetOglyc 4.0 Server. Prediction of mucin type GalNAc O-glycosylation sites
⑥ According to VectorBase database
and in addition, the genomes of other related species have not yet been sequenced. However, the possibility to identify species-specific proteotypic peptides from trypsin peptidases is very interesting and should be followed.

**Conclusion**

The coupling of zymography, proteomic approaches and bioinformatic analyses, as performed here, shows to be a powerful approach in exploring the presence of active peptidases, which helps in the identification of genes that are in fact expressed at the protein level in a specific tissue. In this work, we identified eight different trypsin-like serine peptidases that have singularities at their gene organization level and at the protein sequence level. We identified and characterized trypsin peptidases that are expressed in the midgut of *C. quinquefasciatus*. The bioinformatics analysis conducted here allowed us to suggest that such trypsin peptidases could have primarily digestive functions. Importantly, we identified proteotypic peptide sequences that could be used in the future to

### Table 4 Proteotypic peptides from trypsins identified by MS/MS. Proteotypic peptides were predicted using SKYLINE software

| Uniprot accession number | Protein name     | Peptides identified by MS/MS | Species-specific peptide | Trypsin unique peptide | Other proteins with the same peptide                     |
|--------------------------|------------------|------------------------------|--------------------------|------------------------|----------------------------------------------------------|
| B0W544                   | Trypsin 1        | R.AAYVPAYNQNQCNSA YARY       | Yes                      | No                     | Q1KW6 (Trypsin-like fragment) - C. quinquefasciatus       |
|                          |                  | R.IVGGFEISIADAPHQVSL QSRG   | Yes                      | Yes                    |                                                          |
|                          |                  | RGSHICGGGISIPKWK            | Yes                      | Yes                    |                                                          |
|                          |                  | RNTIDYDYSLELLEKS            | Yes                      | Yes                    |                                                          |
|                          |                  | KWILTAHCTDGSYSNLRI           | Yes                      | Yes                    |                                                          |
|                          |                  | KHASGGVISIKR                |                          |                        | Not predicted by Skyline                                  |
| B0WE94                   | Trypsin 2        | R.DEPADSEQSLVSGWG DTRS      | No                       | No                     | Q962G7 / Q56GM3 (Trypsin) - Culex pipiens                |
|                          |                  | R.LEFGHAVQPVLVRD            | No                       | No                     | Q962G7 / Q56GM3 (Trypsin) - Culex pipiens                |
|                          |                  | R.GVLVPVLNRE                | No                       | No                     | Q962G7 / Q56GM3 (Trypsin) - Culex pipiens                |
| B0XCW2                   | Trypsin 4        | K.GCAOPDYGGYVYAD VEKA       | Yes                      | Yes                    |                                                          |
|                          |                  | K.DDFDALLRL                | Yes                      | Yes                    |                                                          |
|                          |                  | R.VGSSYDYQQGTVIDV AGM TIHPRY| -                        | -                      | Not predicted by Skyline                                  |
| B0X667                   | Trypsin 5        | K.IIGGFPAEQQGTLHQSIRF      | Yes                      | Yes                    |                                                          |
|                          |                  | K.GCGLAAYPGYSDVYYRG       | Yes                      | Yes                    |                                                          |
|                          |                  | RGWIDSCLAGKC               | Yes                      | Yes                    |                                                          |
| B0X870                   | Trypsin 7        | K.DACLGDSGGPTICS GKV       | Yes                      | Yes                    |                                                          |
|                          |                  | R.DYALLNLAKS               | Yes                      | Yes                    |                                                          |
|                          |                  | R.AVDVPIADHDRC             | Yes                      | Yes                    |                                                          |
|                          |                  | R.GGQAVTRK                 | -                        | -                      | Not predicted by Skyline                                  |
| B0W959                   | Serine protease 1/2 | R.TGETFVDNQATVSG FGR      | Yes                      | No                     | C. quinquefasciatus - Q23731 (Serine protease)          |
|                          |                  | R.TVDGGPVSPTKN             | Yes                      | No                     | C. quinquefasciatus - Q23731 (Serine protease)          |
| B0WW44                   | Cationic trypsin  | R.IVHPQYAEGNLAN DIAVIRV   | Yes                      | Yes                    |                                                          |
directly identify trypsin peptidases in complex tissue-specific protein extracts of C. quinquefasciatus. This work represents the first step in the identification, at the protein level, of peptidases expressed in the C. quinquefasciatus midgut and in understanding their role in the complexphysiological processes in such tissue.

Competing interests
The authors declare that they have no competing interests.

Authors’ contributions
ABV, JBJ, and PC designed the study. ABV, LSV, GBD, and GDL performed the experimental work. ABV, LSV, GDL, PC, and JBJ analyzed the data and prepared the manuscript with the critical inputs of CB and GBD. All authors read and approved the final manuscript.

Acknowledgements
This work was supported by FAPEMIG (J.B. Edital Universal Process No. APQ-01070-12), CNPq (J.B. PQ Process No. 308679/2012-1, FAPERJ (C.B. Process E-26/110.594/2012 and CNE-E-26/102.775/2012), FOIOCRUZ-IOC and CAPES. We thank Dr. José Bento (Laboratório de Fisiologia e Controle de Vetores of the Instituto Oswaldo Cruz) for kindly providing the insects.

Author details
1Laboratório de Biologia Molecular e Doenças Endêmicas, Instituto Oswaldo Cruz, FOIOCRUZ, Rio de Janeiro, RJ, Brazil. 2Laboratório de Pesquisa em Leishmaniose, Instituto Oswaldo Cruz, FOIOCRUZ, Av. Brasil 4365, Manguinhos, Pav. Leónidas Deane, Sã o Paulo, CEP 21040-360 Rio de Janeiro, RJ, Brazil. 3Unidade de Proteômica, Laboratório de Química de Proteínas, Instituto de Química, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil. 4Departamento de Medicina, Faculdade de Medicina, Universidade Federal de São João del Rei, São João del Rei, MG, Brasil.

Received: 25 February 2015 Accepted: 3 July 2015
Published online: 16 July 2015

References
1. Farajollahi A, Fonseca DM, Kramer LD, Kilpatrick AM. Bird biting mosquitoes and human disease: a review of the role of Culicoides complex mosquitoes in epidemiology. Infect Genet Evol. 2011;11(7):1577–85.
2. Simpson JE, Hurtado PJ, Medlock J, Molaei G, Andreadis TG, Galvani AP, et al. Vector-host preference studies for transmission of multi-host pathogens: West Nile virus as a model system. Proc Biol Sci. 2012;279(1730):925–33.
3. Kilpatrick AM, Randolph SE. Drivers, dynamics, and control of emerging vector-borne zoonotic diseases. Lancet. 2012;380(9857):1946–55.
4. Weaver SC, Reisen WK. Present and Future Arboviral Threats. Antiviral Res. 2010;85(2):328–45.
5. Richards SL, Lord CC, Pesko K, Tabachnick WJ. Environmental and biological factors influence Culx pipiens quinquefasciatus Say (Diptera: Culicidae) vector competence for West Nile virus. Am J Trop Med Hyg. 2010;83:126–34.
6. Reddy BN, Labbé P, Corbel V. Culx genome is not just another genome for comparative genomics. Parasit Vectors. 2012;5:63.
7. Arensburger P, Megy K, Waterhouse RM, Abrudan J, Amuedo P, Antelo B, et al. Sequencing of Culx quinquefasciatus establishes a platform for mosquito comparative genomics. Science. 2010;330(6000):86–8.
8. Barbolomew LC, Waterhouse RM, Mayhew GF, Campbell CL, Michel K, Zou Z, et al. Pathogenomics of Culx quinquefasciatus and Meta-Analysis of infection responses to diverse pathogens. Science. 2010;330(6000):88–90.
9. Kalhok SE, Tabak LM, Prosser DE, Brook W, Downe AE, White BN. Isolation sequencing and characterization of two cDNA clones coding for trypsin like enzymes from the midgut of Aedes aegypti. Infect Mol Biol. 1993;2(2):71–9.
10. Barillas-Mury C, Wells MA. Cloning and sequencing of the blood meal-induced late trypsin gene from the mosquito Aedes aegypti and characterization of the upstream regulatory region. Insect Mol Biol. 1993;2(1):7–12.
11. Noriega FG, Edgar KA, Bechet R, Wells MA. Midgut exopeptidases activities in the Aedes aegypti are induced by blood feeding. J Insect Physiol. 2002;48(2):205–12.
35. González-Caballero N, Rodríguez-Vega A, Días-Lopes G, Valenzuela JG, Ribeiro JM, Carvalho PC, et al. Expression of the mevalonate pathway enzymes in the Lutzomyia longipalpis (Diptera: Psychodidae) sex pheromone gland demonstrated by an integrated proteomic approach. J Proteomics. 2014;96:117–32.

36. Keller A, Nevszhski AI, Kolker E, Aebersold R. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. Anal Chem. 2002;74(20):5383–92.

37. Nevszhski AI, Keller A, Kolker E, Aebersold R. A statistical model for identifying proteins by tandem mass spectrometry. Anal Chem. 2003;75(17):4646–58.

38. Carvalho PC, Fisher JS, Xu T, Coccov D, Baubuena TS, Valentie RH, et al. Search engine processor: Filtering and organizing peptide spectrum matches. Proteomics. 2012;12(7):944–9.

39. Megy K, Emrich SJ, Lawson D, Campbell D, Dialynas E, Hughes DS, et al. VectorBase: improvements to a bioinformatics resource for invertebrate vector genomics. Nucleic Acids Res. 2012;40:729–34.

40. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol. 2011;7:539.

41. Sigrist CJ, de Castro E, Cerutti L, Cuche BA, Hulo N, Li W, et al. New and continuing development at PROSITE. Nucleic Acids Res. 2013;41:D344–7.

42. Petersen TN, Brunak S, von Heijne G, Nielsen H. SignalP 4.0 discriminating signal peptides from transmembrane regions. Nat Methods. 2011;10(1):785–6.

43. Blom N, Sicheritz-Pontén T, Gupta R, Gammeltoft S, Brunak S. Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. Proteomics. 2004;4(6):1633–49.

44. Steenfort C, Valikruzhev SY, Joshi HJ, Kong Y, Vester-Christensen NB, Schjoldager KT, et al. Precision mapping of the human O-GalNAc glycoproteome through SimpleCell technology. EMBO J. 2013;32(10):1478–88.

45. MacLean B, Tomazela DM, Shulman N, Chambers M, Finney GL, Frewen B, et al. Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. Bioinformatics. 2010;26(7):966–8.

46. Soares TS, Watanabe RM, Lemos FS, Tanaka AS. Molecular characterization of genes encoding trypsin-like enzymes from Aedes aegypti larvae and identification of digestive enzymes. Gene. 2011;489(2):70–9.

47. Vandooren J, Geurts N, Martens E, Van den Steen PE, Opdenakker G. Meconial peritrophic membranes and the fate of midgut bacteria during Meconial peritrophic membranes and the fate of midgut bacteria during. J Proteomics. 2013;10(3):211–20.

48. D’Avila-Levy CM, Santos ALS, Cuervo P, De Jesus JB, Branquinha MH. Applications of Zymography (Substrate-SDS-PAGE) for Peptidase Screening in a Post-Genomic Era. In: Magdelin S, editor. Gel Electrophoresis - Advanced Techniques. China: InTech; 2012. p. 265–88.

49. Nauen R, Sorge D, Sterner A, Borovsky D. TMOF-like factor controls the biosynthesis of serine peptides in the larval gut of Helothris virescens. Arch Insect Biochem Physiol. 2001;47:169–80.

50. Dinglasan RR, Devenport M, Florens L, Johnson JR, McHugh CA, Donnelly-Borges-Veloso et al. Parasites & Vectors 2014;7:253.

51. Dinglasan RR, Devenport M, Florens L, Johnson JR, McHugh CA, Donnelly-Borges-Veloso et al. Parasites & Vectors 2014;7:253.

52. Mesquita-Rodrigues C, Saboia-Vahia L, Domont GB, Britto C, Cuervo P, et al. Expression of active trypsin-like serine peptides in the midgut of sugar-feeding female Anopheles gambiae. Parasit Vectors. 2015;8:296.

53. Muhlia-Almazán A, Sánchez-Paz A, García-Carreño FL. Invertebrate trypsins: a review. J Comp Physiol B. 2008;178(6):655–72.

54. Lopes AR, Juliano MA, Marana SR, Juliano LA, Terra WR. Substrate specificity of insect trypsins and the role of their substrates in catalysis. Insect Biochem Mol Biol. 2006;36(2):130–40.

55. Caip CS, Choo QL, Swift GH, Quinto C, MacDonald RJ, Rutter WJ. Structure of two related rat pancreatic trypsin genes. J Biol Chem. 1984;259(22):14255–64.

56. Klein B, Sello S, Van Wormhoudt A. Genomic organisation and polymorphism of a crustacean trypsin multi-gene family. Gene. 1998;216(1):123–9.

57. Bao YY, Qin X, Yu B, Chen LB, Wang ZC, Zhang CX. Genomic insights into the serine protease gene family and expression profile analysis in the planthopper, Nilaparvata lugens. BMC Genomics. 2014;15:507.

58. Lehane SM, Assinder SJ, Lehane MJ. Cloning, sequencing, temporal expression and tissue-specificity of two serine proteases from the midgut of the blood-feeding fly Stomoxys calcitrans. Eur J Biochem. 1998;254(2):290–6.

59. Neurath H. Evolution of proteolytic enzymes. Science. 1984;224(4647):350–7.

60. Kénesi E, Katona G, Sallágy L. Structural and evolutionary consequences of unpaired cysteines in trypsinogen. Biochem Biophys Res Commun. 2003;309(4):749–54.

61. MacLean B, Tomazela DM, Shulman N, Chambers M, Finney GL, Frewen B, et al. Expediting the development of targeted SRM assays: using data from shotgun proteomics to automate method development. J Proteome Res. 2009;8(6):2733–9.

62. Borges-Veloso et al. Parasites & Vectors (2015) 8:373