Insect Biochemistry and Molecular Biology 103 (2018) 36–45

Contents lists available at ScienceDirect

Insect Biochemistry and Molecular Biology

journal homepage: www.elsevier.com/locate/ibmb

Tissue-specific transcription of proteases and nucleases across the accessory salivary gland, principal salivary gland and gut of *Nezara viridula*

Sijun Liu, Purushottam R. Lomate[1], Bryony C. Bonning[*,2]

Department of Entomology, Iowa State University, Ames, 50011, IA, USA

ARTICLE INFO

Keywords:
Stink bugs
Nezara viridula
Transcriptome
Proteases
Nucleases
Digestion

ABSTRACT

The phytophagous stink bug, *Nezara viridula* (L.) infests multiple plant species and impacts agricultural production worldwide. We analyzed the transcriptomes of *N. viridula* accessory salivary gland (ASG), principal salivary gland (PSG) and gut, with a focus on putative digestive proteases and nucleases that present a primary obstacle for the stability of protein- or nucleic acid-based stink bug control approaches. We performed high throughput Illumina sequencing followed by de novo transcriptome assemblies. We identified the sequences of 141 unique proteases and 134 nucleases from the *N. viridula* transcriptomes. Analysis of relative transcript abundance in conjunction with previously reported proteome data (Lomate and Bonning, 2016) supports high levels of serine protease expression in the salivary glands and high cysteine protease expression in the gut. Specifically, trypsin and chymotrypsin transcripts were abundant in the PSG, and cathepsin L-like cysteine protease transcripts were abundant in the gut. Nuclease transcript levels were generally lower than those of the proteases, the exception being abundant transcripts of ribonuclease-C20 in the PSG. The abundance of chymotrypsin, trypsin, and some carboxypeptidase transcripts suggests a significant role for the PSG in production of digestive enzymes. This result is at odds with the premise that the ASG produces watery saliva, which is high in enzymatic activity, while the PSG produces only sheath saliva. We have generated a comprehensive transcriptome sequence dataset from the digestive organs of *N. viridula*, identified major protease and nuclease genes and confirmed expression of the most abundant enzymes thereby providing greater insight into the digestive physiology of *N. viridula*.

1. Introduction

Stink bugs are serious agricultural pests that inflict widespread damage with both native and invasive species presenting particular challenges in various regions of the world (Panizzi, 2015). In Brazil alone stink bug activities result in yield losses of more than US$600 million in soybean and more than US$100 million in maize each year (CEPEA/ESALQ, 2017). These insects have the remarkable ability to feed on plant species ranging from herbs and vegetables to trees, and host plants include major crops such as soybean, cotton, corn, cereals, legumes, fruits and vegetables (Corrêa-Ferreira and Azevedo, 2002; Panizzi, 2015; Panizzi et al., 2000). The cosmopolitan southern green stink bug, *N. viridula*, is known to feed on 145 plant species from 32 families (Kiritani et al., 1965; Todd, 1989). This wide-ranging phytophagous behavior of stink bugs prevents focused management appropriate for pests restricted to a given cropping system.

The primary management strategy for stink bugs is the application of chemical insecticides, which may pose ecological risks due to their lack of specificity (O'Neal et al., 2018), and may have limited usefulness due to the evolution of resistance in the absence of an integrated pest management approach (Snodgrass and Scott, 2003; Willrich et al., 2003). Alternative approaches such as the use of *Bacillus thuringiensis* –derived pesticidal proteins (Bravo et al., 2011), and the silencing of genes essential for insect survival have been applied for suppression of some pest insects (Scott et al., 2013; Zhang et al., 2013). However, the successful application of such protein- and nucleic acid-based agents for suppression of *N. viridula* populations requires a thorough understanding of the host digestive environment, which presents a hostile environment for the stability of such protein or nucleic acid-based actives. Similar to other stink bug species, *N. viridula* employs both extra-oral and gut-based digestion with digestive enzymes produced by both the salivary glands and gut. Investigation of digestive enzyme activities

* Corresponding author. Department of Entomology and Nematology, University of Florida, PO Box 110620, Gainesville, FL, 32611, USA.
E-mail address: bbonning@ufl.edu (B.C. Bonning).
[1] Present address: Ajeet Seeds Pvt. Ltd., Maharashtra, India.
[2] Present address: Department of Entomology and Nematology, University of Florida, Gainesville, FL, 32611, USA.

https://doi.org/10.1016/j.ibmb.2018.10.003
Received 13 April 2018; Received in revised form 30 July 2018; Accepted 19 October 2018
Available online 21 October 2018
0965-1748/

in the *N. viridula* salivary gland, saliva and gut using enzyme assays and proteomics analysis highlighted the division in tissue function with salivary gland production of serine proteases active at alkaline pH, and gut production of cysteine proteases active at acidic pH (Lomate and Bonning, 2016).

To build on these prior observations, we generated a comprehensive RNA-Seq transcriptome dataset from *N. viridula* accessory salivary gland (ASG), principal salivary gland (PSG), and gut. We identified and functionally annotated a set of protease and nuclease transcript sequences. Differential transcription of these enzymes across tissues (ASG, PSG and gut), was assessed by comparison of RPKM values, and further validated by quantitative real time PCR (RT-qPCR) analysis. Protein expression from the most abundant transcripts was confirmed by mapping peptide libraries derived from salivary gland and gut proteomes against the translated protein sequences derived from the tissue-specific transcriptomes. In addition to allowing for identification of potential target enzymes for stink bug suppression, these digestive tissue-based, transcriptomic datasets provide fundamental information on the specific roles of the respective tissues in digestive enzyme transcription and the enzymatic challenges to be overcome for protein- and nucleic acid-based stink bug management approaches.

2. Material and methods

2.1. Insect rearing, tissue dissection and RNA isolation

N. viridula were reared under controlled conditions (16 h light; 8 h dark, 27–28 °C, 65% humidity) and fed a mixed diet comprised of green beans, sweet corn, carrots and raw peanuts. Adults (100) were chilled for 10 min over ice prior to dissection. The ASG, PSG and gut were isolated under a dissection microscope and directly placed into Trizol reagent (Invitrogen, Carlsbad, CA, USA).

2.2. RNA isolation and transcriptomic sequencing

Total RNA was isolated using Trizol reagent according to the manufacturer's directions. The quality and integrity of RNA samples was checked using a 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA) and by agarose gel electrophoresis. Illumina sequencing was performed for libraries for each of the three tissues. All cDNA libraries were generated according to Illumina protocols (Illumina Inc., San Diego, CA, USA). Briefly, 10 μg of total RNA from each *N. viridula* tissue (ASG, PSG and gut) was treated with DNase I (Sigma, St. Louis, USA). The mRNA was purified by using oligo (dT) magnetic beads and fragmented in the range of ~100–400 bp. The cDNA was synthetized by reverse transcriptase (Invitrogen, Carlsbad, CA, USA) using random hexamer-primers with the mRNA fragments as templates. The quality of each library was checked by 2100 Bioanalyzer. Sequencing of cDNA libraries 100 base reads was performed using Illumina HiSeq2500™ (Illumina Inc.,) using standard procedures at the Iowa State University DNA Facility.

2.3. De novo RNA-seq assembly and annotation

All raw RNA-seq sequence data were submitted to the NCBI Sequence Read Archive (SRA; accession number SRP119668). Raw sequencing reads were analyzed for quality using FastQC. Low quality bases and adaptor sequences were trimmed by using the FASTX toolkit and high quality reads were assembled into tissue-specific transcripts using the Trinity assembler (Grabherr et al., 2011). Assembled transcripts were analyzed by BLASTx against the NCBI non-redundant (nr) database. Gene ontology (GO) annotations of transcripts were attained by use of Blast2GO software (http://www.geneontology.org). Top hits from BLAST analyses were sorted to identify proteases and nucleases. Potential protein sequences of ≥100 amino acids were translated by TransDecoder software (https://transdecoder.github.io). The RNA and protein sequences of proteases and nucleases were manually checked for sequence similarity using a multiple sequence alignment tool (Clustal Omega; http://www.ebi.ac.uk/Tools/msa/clustalo/). Putative conserved domains were identified using BLASTp and full length, non-redundant protease and nuclease sequences obtained. Signal peptides were identified using SignalP (with default parameters) for the most abundant proteases (Petersen et al., 2011; http://www.cbs.dtu.dk/services/SignalP/).

To analyze transcript abundance, sequence reads were mapped to contigs, and reads per kilobase of transcript per million mapped reads (RPKM) calculated based on the numbers of mapped reads. RPKM estimates were inferred by using the "align_and _*estimate*_abundance.pl" script in the Trinity package, using default parameters. RSEM was used for expression estimations (Li and Dewey, 2011). Read alignment was conducted using Bowtie (Langmead et al., 2009). Transcripts were categorized as abundant (RPKM > 500), moderately abundant (RPKM 25–500), or low abundance (RPKM < 25).

2.4. Mapping of proteome-derived peptides to digestive enzyme transcripts

The short peptides from the *N. viridula* gut and salivary gland proteomes (Lomate and Bonning, 2016), were mapped against the putative protein sequences translated from the transcripts of ASG, PSG and gut. The mapping was performed using Mascot and the Thermo Scientific Proteome Discoverer 2.1 software package at the Iowa State University Protein Facility. A minimum of 3 peptides matched to a protein was considered a positive result.

2.5. Quantitative real-time PCR

The relative transcription levels of eight protease and four nuclease genes were examined across the three tissues, ASG, PSG and gut. Enzyme coding sequences were first confirmed by Sanger sequencing. Primers designed for amplification of 500 bp regions using Primer3Plus software (http://primer3plus.com/cgi-bin/dev/primer3plus.cgi) are listed in Supplementary Table S1. RT-qPCR analysis was performed with three biological and three technical replicates on an iCycler™ iQ Optical system (Bio-Rad, Hercules, CA, USA). Tissues (ASG, PSG and gut) were isolated from a total of 40 N. *viridula* adults for each replicate. Total RNA was isolated using Trizol reagent. For each replicate, first strand cDNA was synthesized from 1 μg RNA using the Superscript Reverse Transcriptase III kit (Life Technologies, Carlsbad, CA, USA). Reactions of 20 μL to amplify selected protease and nuclease cDNAs by RT-qPCR were set up using iQ SYBR Green Supermix (Bio-Rad). Reactions were performed in triplicate and run in 96-well plates under the following conditions: 95 °C for 3 min, and 40 cycles of 95 °C for 10 s and 58 °C for 45 s. Melting curves (55 °C–95 °C) were derived for each reaction to ensure a single product. Relative gene expression of *N. viridula* proteases and nucleases was calculated by $2^{-\Delta\Delta\ Ct}$ method (Livak and Schmittgen, 2001). The transcript level of ribosomal protein S17 was quantified for normalization of the cDNA templates (Nakamura et al., 2016; Sparks et al., 2014).

3. Results

3.1. Assembly and functional annotation of the N. viridula tissue transcriptomes

Three separate assemblies of high quality reads obtained by deep sequencing of RNA from the ASG, PSG and gut were prepared using the Trinity Assembler. The numbers of assembled contigs, the total length of those contigs, mean transcript length, and N50 for each tissue are summarized in Table 1. Analysis using BLASTx against the NCBI nr protein database showed that 27% of PSG-, 34% of ASG- and 53% of gut-assembled sequences hit previously reported protein sequences, with significant E-values ranging from $1e^{-180}$ to 0 (Fig. 1A). The

Table 1
Summary of *N. viridula* ASG, PSG and gut transcriptome statistics.

Tissue	Total no. contigs ≥200 bp	Total length of contigs (bp)	Mean transcript length (bp)	N50
ASG	72,944	70,015,884	389	1411
PSG	86,836	53,565,662	358	845
Gut	33,841	29,993,800	443	1182

majority (65–82%) of the top BLAST hits for sequences from all three tissues were to the recently annotated genes of the brown marmorated stink bug, *Halyomorpha halys* (Fig. 1B). In addition to insect proteins, 5% of the ASG top hits were to proteins derived from microsporidia, *Nosema* spp., which are pathogens of stink bugs (Hajek et al., 2017). Approximately 4% of the top hits for gut transcripts were to sequences derived from *Pantoea* spp., bacterial symbionts that reside in stink bug guts (Prado and Almeida, 2009). About 1% of the gut transcripts hit *E. coli* genes (Fig. 1B).

The Gene Ontology (GO) annotation grouping of transcripts into different categories is shown in Fig. 1C. Within the 'biological process' GO category some 4000 transcript sequences for each tissue were assigned to 'cellular process' and about 4500 to 'metabolic process'. Transcripts assigned to the 'cell' (~2000) and 'organelle' (~1500) subcategories were the most abundant among the 'cellular component' GO category. Within the 'molecular function' category, more than 5000 and 3500 transcript sequences were associated with binding and catalytic activity functions, respectively for each of the three tissue transcriptomes (Fig. 1C).

3.2. Confirmation of protein expression by reference to gut and salivary gland proteomes

To analyze contigs that potentially encoded proteins ≥100 amino acids (aa), putative full-length or partial protein sequences were translated (CAP3 contig sets). Analysis of potential protein sequences translated from the three contig sets (one per tissue) is summarized in Table 2. The number of contigs that potentially encoded protein sequences of ≥100 aa ranged from 12,056 (gut) to 16,664 (ASG) (Table 2). The gut had the highest percentage (38%) of transcripts that potentially encoded proteins ≥ 100 aa, while the PSG had the lowest with only 16.67% (Table 2). These translated protein sequence sets were used as targets for mapping of peptides that resulted from a previous proteomics analysis of *N. viridula* salivary gland and gut (Lomate and Bonning, 2016). The mapping results are summarized in Table 3. All of the protease and nuclease sequences identified from the tissue-specific proteomes, were represented in the transcriptomes in the current analysis. The numbers of peptides that mapped to translated transcripts were comparable across the three tissues. Although the fewest contigs were obtained from the gut transcriptome (Table 1), the translated gut transcriptome had the highest number and percentage of hits from gut proteome-derived peptides (Table 3). Overall, some 25% of the gut transcripts predicted to express proteins ≥100 aa were

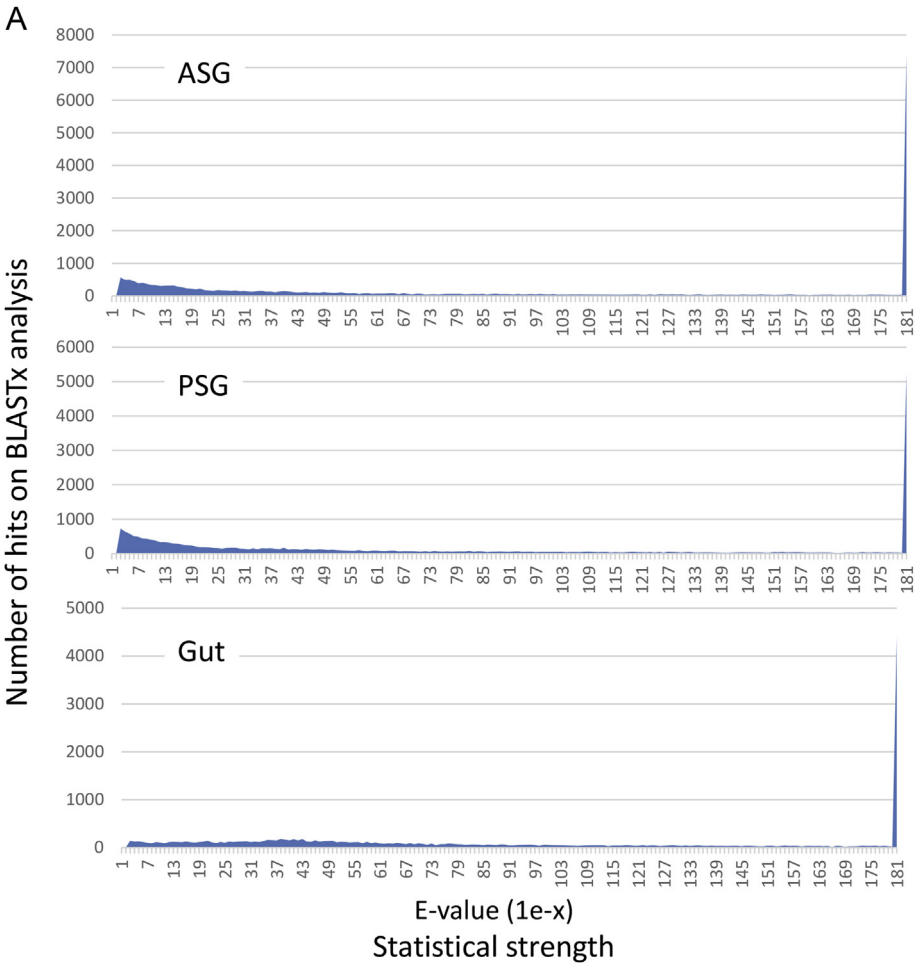

Fig. 1. Summary of BLAST statistics and gene ontology terms of *N. viridula* ASG, PSG and gut transcriptomes. **A:** E-value distribution for BLASTx matches for *N. viridula* ASG, PSG and gut transcriptomes. The majority of hits with a high level of significance were sequences derived from *H. halys*. **B:** Species distribution of the best BLASTx hits in the nr database for ASG, PSG and gut transcripts. The proportion of hit sequences (%) derived from specific species are shown for species with the highest hit numbers. **C:** Gene ontology (GO) terms assigned to gene sequences involved in biological process, cellular component, and molecular function.

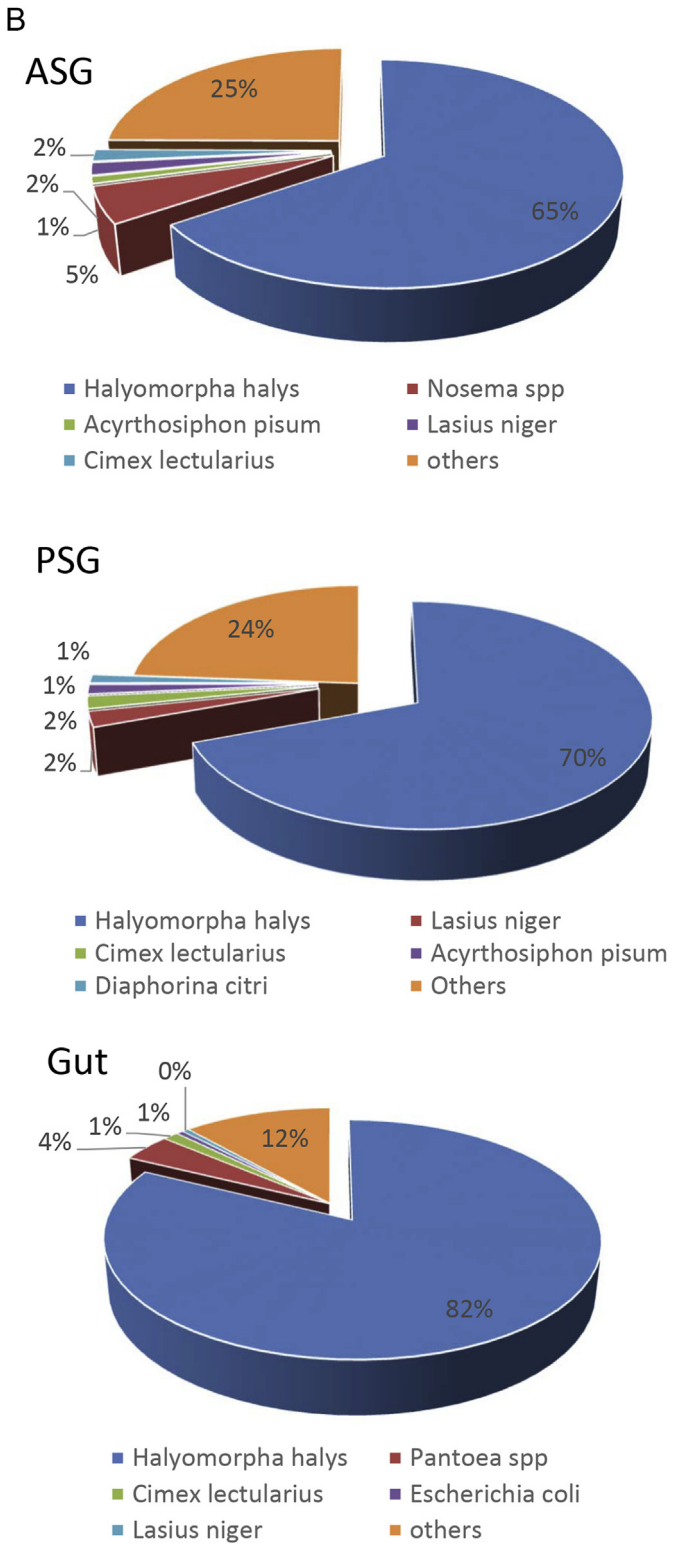

Fig. 1. (*continued*)

detected at the protein level, with 17% and 21% in the ASG and PSG, respectively (Table 3).

3.3. Comparison of transcript and protein abundance

Protein production is regulated at the levels of both transcription and of translation, with high transcript numbers typically correlating with high abundance proteins. To examine *N. viridula* transcript abundance, we mapped the Illumina sequencing reads to the assembled contigs and calculated RPKM. The RPKM values for contigs encoding putative proteins (≥100 aa) were compared to the numbers of translated transcripts mapped against proteome-derived peptides to determine whether high RPKM values correlated with highly expressed proteins identified in the proteome. The relationship between transcript abundance and proportion of proteins identified in the proteome is shown in Fig. 2. The numbers of transcript-translated proteins detected by peptide mapping increased with increasing transcript RPKM. Less than 5% of the transcripts with fewer than 5 RPKM were detected in protein profiles. Around half of the transcripts with RPKM of > 500 encoded proteins that were detected in the proteome, and up to 80% of transcripts with RPKM > 1000 encoded proteins that were detected.

3.4. Identification of protease sequences

Protease- and peptidase-encoding transcripts identified by BLASTp were selected for further analysis. A total of 141 unique protease and peptidase transcripts were identified, which hit 111 putative protease and peptidase proteins. The majority of the top hit proteases and peptidases were derived from *H. halys* with a few apparently derived from pathogens and endosymbionts based on top hit and codon preference. The majority of the enzyme sequences appear to be full-length (Supplementary Fig. S1), which will be clarified once the *N. viridula* genome sequence becomes available. The number of *N. viridula* proteases with homology to the major protease groups is listed in Table 4. Of the major enzyme groups, 44 peptidase transcripts (21 aminopeptidases, 19 carboxypeptidase and 4 dipeptidases) were identified along with 80 transcripts of cathepsin-, chymotrypsin- and trypsin-like proteases (Table 4). About 37.6% (53) of the proteins translated from these protease and peptidase-like transcripts were identified in the proteomic profiles of *N. viridula* (Table 4, Supplementary Table S2). Of the 51 proteases detected in the proteome, signal peptides were identified in 33. Twelve lacked signal peptides and five were indeterminate due to incomplete 5' sequence data (Supplementary Table S2).

To assess the relative abundance of the protease and peptidase transcripts, transcript RPKM was compared between tissues (Fig. 3, Supplementary Table S2). Although more than 100 protease and peptidase-like proteins were identified from the transcripts, relatively few were expressed at very high levels (> 500 RPKM). Proteins encoded by the most abundant transcripts were detected in the proteome (Supplementary Table S2). Expression of the different enzymes varied across ASG, PSG and gut. Among the aminopeptidases, only aminopeptidase-2 (aminopeptidase N-like) had > 1000 RPKM in the gut transcriptome, and moderate transcript levels (RPKM > 25) in PSG. Five other aminopeptidases (aminopeptidase-11, 12, 13, 15, 20 and 21) were moderately expressed (RPKM > 25) in the ASG, with aminopeptidases 20 and 21 also moderately expressed in the PSG and gut.

The most abundant carboxypeptidase transcripts (carboxypeptidase-1 and -2; both carboxypeptidase B-like) were observed in PSG, with all but three of the 18 enzymes showing moderate transcript levels in one or more of the three tissues. Cathepsins are essential digestive enzymes as reflected by identification of 42 transcripts encoding cathepsin-like proteases from the *N. viridula* transcriptomes. All of the high abundance transcripts encoding cathepsin-like proteases were observed in the gut. Similarly, most of the cathepsin proteins detected in the proteome were identified from the gut. Transcripts encoding a single cathepsin B-like protein (cathepsin-5) and two cathepsin D-like proteins (cathepsin-8 and 9) were found in all tissues.

For the chymotrypsin group, four of the six chymotrypsin-like transcripts were highly abundant in PSG with three of these identified from the salivary gland proteome. The number of chymotrypsins was relatively low in both the transcriptome and proteome datasets. A total of 35 trypsin-like transcripts were identified only six of which were abundant. Similar to the chymotrypsins, the high abundance trypsin transcripts were detected in the PSG transcriptome, with relatively low

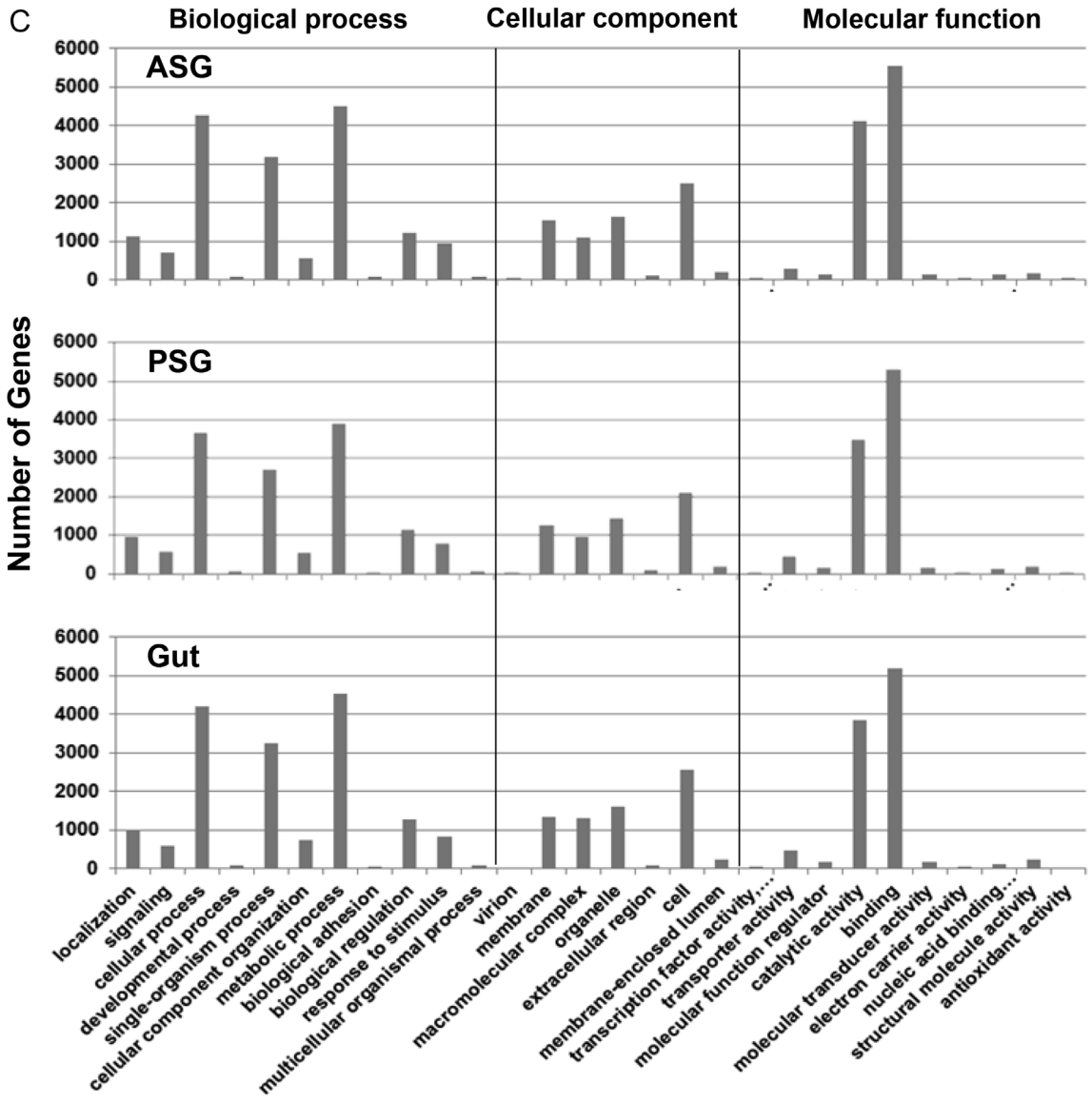

Fig. 1. (*continued*)

Table 2
Summary of BLAST hit and protein translation from ASG, PSG and gut transcripts of *N. viridula*.

Tissue	Number of contigs -					No. of translated putative protein sequences (≥100 aa)[b]
	Used for BLAST search[a]	With a hit	Proportion with a hit (%)	Encoding putative proteins ≥100aa	Proportion encoding putative proteins ≥100aa (%)	
ASG	67,348	24,588	36.51	16,664	24.74	23,546
PSG	80,537	22,768	28.27	13,424	16.67	17,245
Gut	31,714	17,865	56.33	12,056	38.01	14,776

[a] Trinity-assembled contigs were merged with CAP3.
[b] Protein sequences were translated by TransDecoder. More than one peptide sequence may be translated from a single contig.

RPKM in ASG and gut. Six trypsin-like proteins were detected in the proteome, one in the gut and five in the salivary gland (Supplementary Data 1).

3.5. Identification of *N. viridula* nucleases

Analysis of hundreds of contigs that hit nuclease-like genes from the transcriptomes of ASG, PSG and gut tissues by BLASTp annotation resulted in the identification of 133 contigs of nuclease-like transcripts. These included contigs of 34 endonucleases, 23 exonucleases, 18 nucleosidases, 11 endoribonucleases 15 exoribonucleases, and 33 ribonucleases (Supplementary Table S3). The contigs hit 90 unique nuclease-like genes the majority of which were derived from *H. halys*. Compared to proteases and peptidase transcripts, fewer full-length or near full-length nuclease transcripts were assembled. The relative abundance of the nuclease transcripts (RPKM) across tissues is shown in Fig. 4. The overall abundance of nuclease transcripts was lower than that of protease transcripts with only 15 nucleases found in the

Table 3
Summary of proteome-generated peptides mapped to translated transcripts of the ASG, PSG and gut. Each set of translated protein sequences derived from the ASG, PSG or gut transcriptome was used as a target for mapping of peptides derived from the salivary gland (ASG, PSG and salivary duct) or gut (midgut and hindgut) proteome.

Tissue	No. of transcripts mapped to proteome peptides of		Number of unique translated transcript hits	% of translated transcripts mapped to proteome (≥100 aa)
	Salivary gland	Gut		
ASG	2590	1680	2881	17.29
PSG	2685	2167	2894	21.56
Gut	2764	1036	3090	25.63

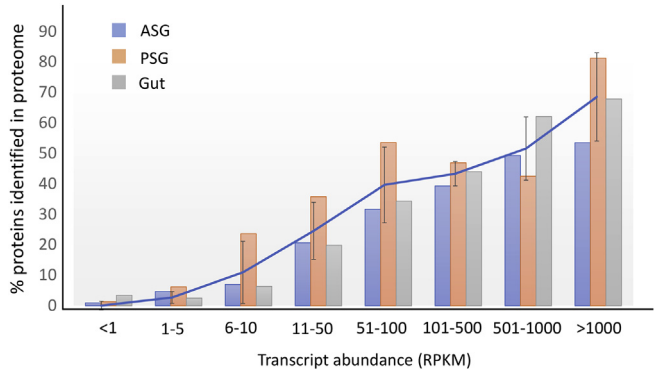

Fig. 2. Proteins with highly abundant transcripts were more likely to be detected in the proteome. Translated protein sequence sets derived from the ASG, PSG and gut were used as targets for mapping of peptides that resulted from a proteomic analysis of the *N. viridula* gut and salivary gland reported previously (Lomate and Bonning, 2016). The proportion of proteins detected from the proteome increased with increasing RPKM. The line shows the average RPKM of ASG, PSG and gut transcripts combined. Error bars represent the standard error of the mean.

Table 4
Number of proteases or protease fragments identified from *N. viridula* tissue specific transcriptomes and proteomes.

Enzyme	Number of enzymes identified from	
	Transcriptome	Proteome
Aminopeptidase	21	7
Carboxypeptidase	19	9
Cathepsin/cysteine proteinase	42	20
Chymotrypsin	6	3
Trypsin/serine protease	32	6
Dipeptidase	4	4
Aspartic proteinase	8	2
Other	5	2
Total	136	53

proteomic datasets. Of these, most were found in the salivary gland (Supplementary Table S3). Although 34 endonuclease-related contigs were assembled, no endonucleases were identified in the protein profiles.

A single ribonuclease O-like transcript (ribonuclease-C20), a widespread RNase T2 family of secreted RNases, had high transcript abundance (RPKM of ~4320) in the PSG. Clear trends with specific proteolytic enzyme types predominating in specific tissues (Fig. 3) were not apparent for the nucleases (Fig. 4, Supplementary Table S3).

3.6. RT-qPCR-based validation of relative N. viridula protease and nuclease transcript levels

The relative abundance of enzyme transcripts is reflected by their respective RPKM values (Supplementary Tables S2 and S3). Furthermore, the majority of protease and nuclease-like proteins with high RPKM values were identified from the *N. viridula* gut and salivary gland proteomes. To further compare relative transcript abundance, we selected 12 transcripts for analysis by RT-qPCR with gene-specific primers. The RT-qPCR data for the selected protease and nuclease transcripts displayed differential abundance across tissues that correlated well with RPKM data when RPKM > 1000 (Fig. 5A). For example, transcripts of trypsin-22 (RPKM 5455) and trypsin-23 (RPKM 1955) were most abundant in PSG and transcripts of cathepsin-26 (cathepsin L; RPKM 26,272) were most abundant in gut, with negligible transcript numbers in the other two tissues (Fig. 5A).

However, relative transcript abundance as determined by RT-qPCR did not directly correlate with RPKM data for RPKM < 500. For example, while RT-qPCR indicated higher abundance of cathepsin-5 (a cathepsin B gene) in ASG and PSG compared to the gut, the RPKM was higher for ASG (47) than for the other two tissues (PSG 14, gut 11). Similarly, cathepsin-8 (a cathepsin D gene) had ~2-fold difference in RPKM values across the tissues with the highest (96) in the gut. Similar observations were made for exopeptidase (carboxypeptidase-4), aminopeptidase-11 (aminopeptidase N) and aminopeptidase-20 (a leucyl aminopeptidase). Relatively low abundance transcripts (RPKM) were detected across all three tissues for these proteases (Fig. 5A).

Similar results were observed for nuclease transcripts with RT-qPCR data reflecting RPKM data when RPKM > 500 (Supplementary Table S3). Ribonuclease-C20 in the PSG was the only nuclease transcript with an RPKM > 50 (RPKM 4321; Fig. 5B). Although this relatively high transcript abundance in PSG was reflected by both RT-qPCR and RPKM data, RPKM data suggested that transcript levels in ASG and gut were comparably low. Of the 4 nuclease genes examined (Fig. 5B), full length transcripts were assembled for exoribonuclease-C15 (RNA exonuclease 4) and nucleotidase-C14 (cytosolic purine 5′-nucleotodiase). Transcripts for these nucleases were detected at low levels across all three tissues.

4. Discussion

The motivation behind the work reported here was delineation of the specific enzymes contributing to protease and nuclease activity in the salivary gland and gut of *N. viridula*, building on previously reported enzyme activity and proteomic data for this species (Lomate and Bonning, 2016). In addition, we sought to clarify the potential role of the ASG in digestive processes through production of digestive enzyme transcripts. We sequenced the *N. viridula* ASG, PSG and gut transcriptomes, identified and functionally annotated the key protease and nuclease transcripts (Supplementary Tables S2 and S3). Analysis of transcript abundance with reference to the salivary gland and gut proteomes indicated that the more abundant the transcript, the more likely the detection of the corresponding protein in the proteome. The transcriptome data for the ASG, PSG and gut largely support previous findings related to differential expression of proteases between the salivary and gut tissues of *N. viridula*. Comparison of RT-qPCR data with RPKM values suggests that only RPKM values > 500 provide a reliable indication of relative transcript abundance.

Although numerous transcripts were identified from the transcriptomes of ASG, PSG and gut with homology to known sequences, less than 40% of the transcripts from ASG and PSG were annotated, compared to 56% of those from the gut, possibly due to the lack of genomic information for *N. viridula*. Only 17–38% of transcripts from ASG, PSG and gut encoded proteins > 100 aa. Nevertheless, the unique transcripts from different tissues reported here represent a significant proportion of the functional genes from this insect pest.

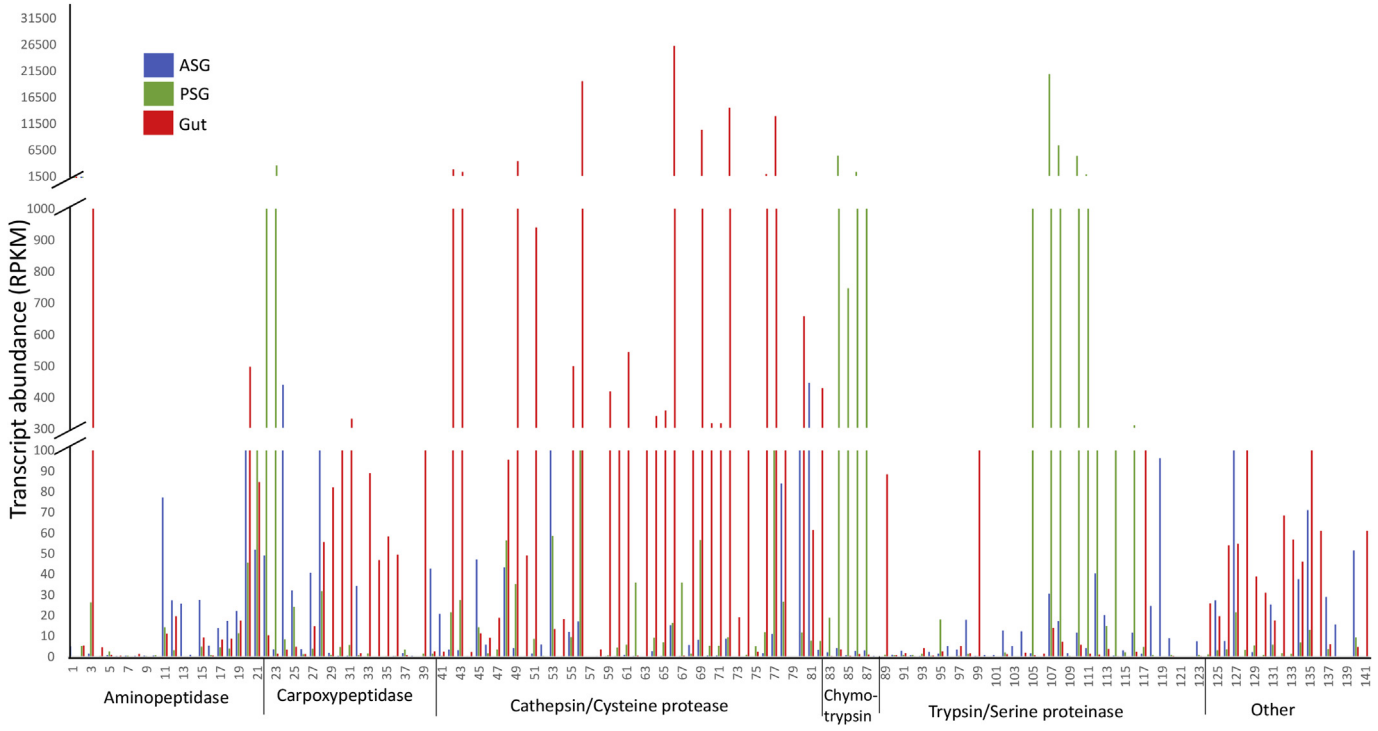

Fig. 3. Relative abundance of protease transcripts identified from the *N. viridula* ASG, PSG and gut transcriptomes. The predominance of cathepsin/cysteine proteases in the gut, and chymotrypsin and trypsin-like proteases in the PSG is evident. Protease numbers shown on the x-axis are cross-referenced in Supplementary Table S2.

Stink bug genomic resources have largely been focused on *H. halys* with examination of whole insects from different developmental stages, with several important protease and nuclease transcript sequences identified (Ioannidis et al., 2014; Sparks et al., 2014). The majority of the putative digestive enzyme genes reported here were predicted from annotation of the *H. halys* genome. Identification of 141 unique

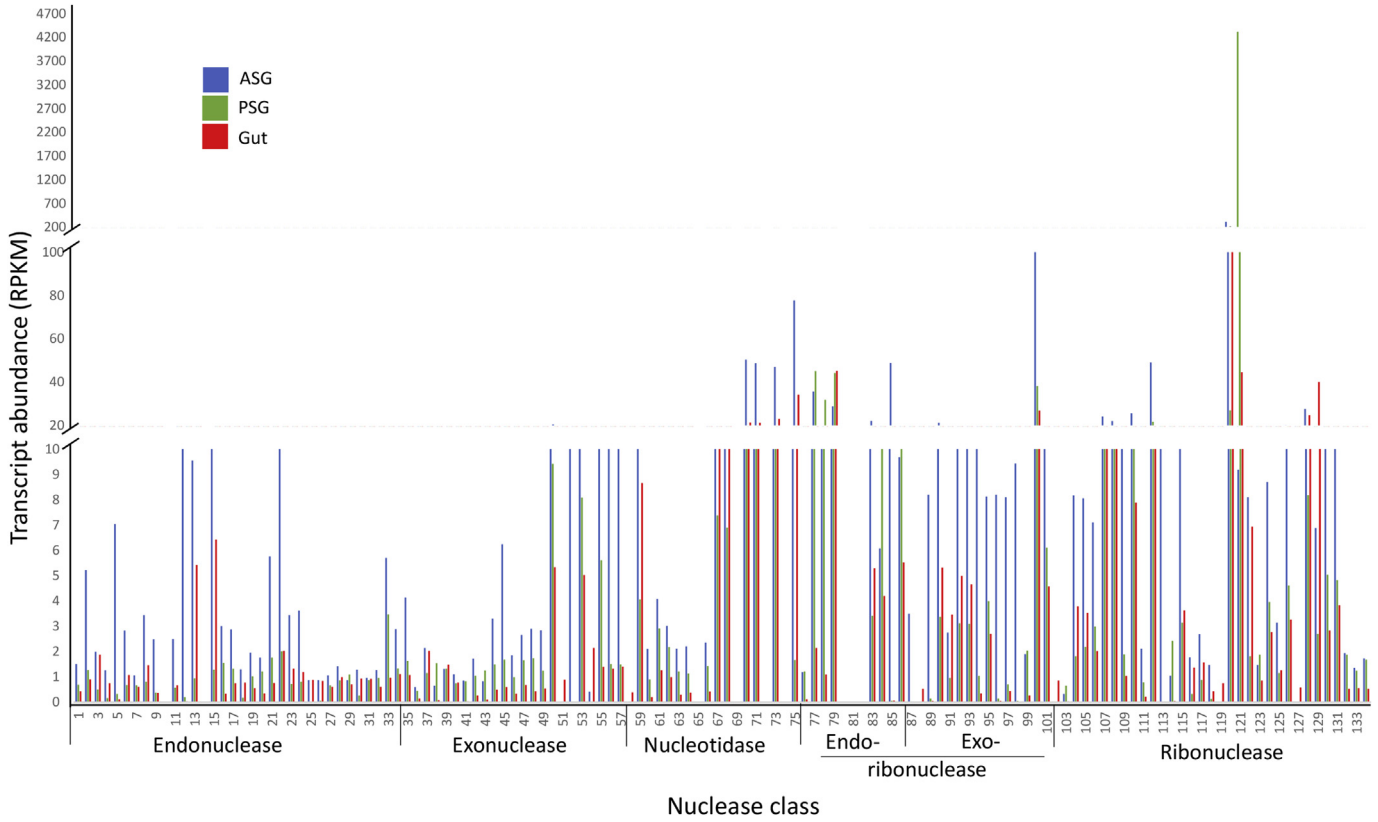

Fig. 4. Relative abundance of nuclease transcripts identified from ASG, PSG and gut of *N. viridula.* Nuclease numbers shown on the x-axis are cross-referenced in Supplementary Table S3.

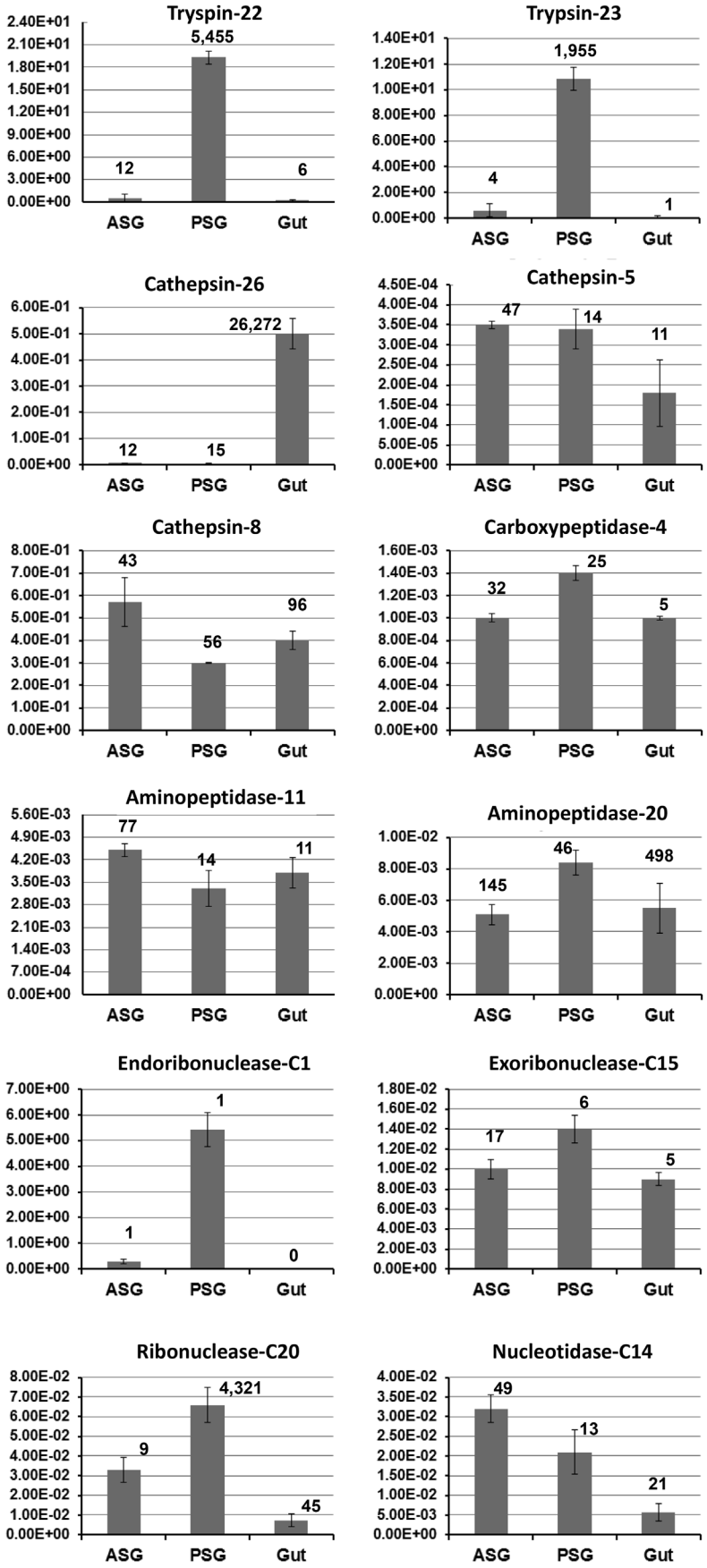

Fig. 5. Transcript validation for selected transcripts from *N. viridula* ASG, PSG and gut and comparison with RPKM values. RT-qPCR assessment of relative abundance of **A:** eight protease transcripts and **B:** four nuclease transcripts. Three biological replicates with three technical replicates were performed. Bars denote the mean of sample-specific $2^{-\Delta Ct}$ values, and error bars represent the standard error of the mean. Numbers by each bar indicate RPKM. Trypsin transcripts were predominantly expressed in PSG while cathepsin-26 transcripts were abundant in the gut.

protease transcripts likely reflects the highly evolved digestive capabilities of *N. viridula*. This number is comparable to the numbers of unique proteases identified in other polyphagous species including cotton bollworm, *Helicoverpa armigera* (100; (Pearce et al., 2017)) and fall armyworm, *Spodoptera frugiperda* (86–112; (Gouin et al., 2017)), and more than the number of digestive proteases in the oligophagous tobacco hornworm, *Manduca sexta* (68; (Cao et al., 2015)). The predominant digestive enzymes of Hemiptera are typically cathepsins or cysteine proteases (Terra and Ferreira, 1994), with high expression of cathepsins in the aphid gut for example (Liu et al., 2012; Rispe et al., 2008). Our results highlight the use by *N. viridula* of multiple proteases with representation from each class of proteases, with differential transcription across the ASG, PSG and gut. These proteases include aminopeptidases, carboxypeptidases, cathepsins/cysteine proteases, trypsins, chymotrypsins and other serine proteases, although the transcription of more than half of the enzymes was at relatively low levels. Cathepsins were highly expressed in the gut, while trypsins and chymotrypsins were highly expressed in the PSG. No highly abundant transcripts for proteolytic enzymes were detected in the ASG transcriptome. These results were largely consistent with protease activities reported for the gut, salivary gland and saliva, and were further confirmed by the presence of the corresponding proteases in the *N. viridula* salivary gland and gut proteomes (Lomate and Bonning, 2016). The proportion of enzymes identified that function in digestion remains to be determined as neither the enzyme activity assays, nor the proteome and transcriptome analyses account for the subcellular localization of these proteins. As indicated by Fig. 1C, many are expected to function within the cell.

A total of 134 unique nuclease transcripts with putative conserved domains for DNase, RNase or dsRNase activities were identified. Interestingly, high transcript abundance was detected for a single ribonuclease (T2 like nuclease), with nuclease transcripts at lower abundance (most < 10 RPKM) compared to that of the proteases. Incomplete assembly of the majority of nuclease genes may have further exacerbated the relatively low RKPM. Hence, RPKM values may not accurately reflect nuclease transcript abundance, and RT-qPCR data may offer greater reliability. Only three nuclease proteins (nucleotidase-C14, nucleotidase-C16 and ribonuclease-C20) were identified from the peptide profiles of *N. viridula* gut, while 14 nucleases were found in the salivary gland peptide database. These data for differential transcription of proteases and nucleases, indicate that the gut and PSG play the most important roles in protein digestion, while the PSG and ASG are the most important for RNA and DNA degradation. This results align well with overall protease activity and nuclease activity profiles reported previously for the *N. viridula* salivary gland and gut (Lomate and Bonning, 2016). It is likely that following gene duplication some of the genes may have acquired functions other than that of digestive protease (Li et al., 2005), as noted previously for a venom protease in aphid soldiers (Kutsukake et al., 2004).

The combined use of transcript abundance (RPKM) and proteomics analysis proved to be an efficient approach to investigate the profiles of the most abundant enzymes. Validation of relative transcript abundance between tissues by RT-qPCR could only be used to confirm differential expression of proteases that had significantly different RPKM i.e. with > 100-fold difference. For the proteases and nucleases that did not exhibit differential transcription across tissues with RPKM of < 500, RT-qPCR results did not consistently agree with the relative RPKM values highlighting the sensitivity limitations of the respective methods.

Transcriptome data reflect our previous enzymatic activity data from *N. viridula* salivary gland and gut (Lomate and Bonning, 2016). Overall, data showed the expression of distinct protease genes in different tissues and support our observation that *N. viridula* uses separate pool of enzymes for gut-based and extra-oral digestion. Genome- and transcriptome-wide analysis of the serine protease family in the brown planthopper *Nilaparvata lugens* (Hemiptera: Delphacidae) resulted in identification of 65 serine protease genes with intact catalytic triads (22 trypsin-like), and indication of frequent gene duplication (Bao et al., 2014). Transcripts for 10 of the active trypsins were detected exclusively in the gut consistent with a role in digestion. In contrast to the high expression of several trypsin-like enzymes in *N. viridula* reported here, trypsin transcript abundance was relatively low in the salivary gland of *N. lugens* (Bao et al., 2014). We identified 26 cathepsin-L-like genes in *N. viridula* (Supplementary Table S2). Based on semi-quantitative PCR, 33 of 40 *H. halys* cathepsin-L genes analyzed exhibited gut-specific expression (Bansal and Michel, 2018) reflecting the importance of this enzyme in proteolytic digestion and defense against protease inhibitors. Importantly, insects regulate expression of cathepsins during adaptation to different host plants (Zhu-Salzman and Zeng, 2015).

The two lobes of the stink bug salivary gland are believed to produce watery saliva (ASG) and sheath saliva (PSG) (Kumar and Sahayaraj, 2012; Miles, 1972; Will et al., 2012). Watery saliva is involved with digestion of plant food and contains essential digestive enzymes. Sheath saliva forms the salivary sheath, a hard, straw-like structure that forms a seal around the stylets during feeding. We characterized the transcriptomes of ASG and PSG separately although dissection of the two separate lobes is not trivial, to address their respective roles in digestive enzyme production. Data presented in Fig. 3 indicate that the ASG play a greater role than the PSG in the overall transcription of protease enzymes. Some of the enzymes are secreted by the PSG as zymogens requiring subsequent proteolytic activation. Trypsins are known to play a significant role in the downstream activation and regulation of other proteases in insects (Barillas-Mury et al., 1995; Lehane et al., 1996). In addition to providing efficient protein degradation, the trypsin complex also contributes to an adaptive advantage for insects feeding on plants such as soybean that produce trypsin inhibitors in defense against herbivores (Bown et al., 1997; Mishra et al., 2015; Zhu-Salzman and Zeng, 2015). The impact of host plant on the digestive enzymes of *N. viridula* is a question that is now ripe for investigation.

Transcripts derived from *Nosema* sp. were abundant in the ASG (5%, Fig. 1B) with few reads in the PSG and gut transcriptomes, while transcripts from *Pantoea* spp. were abundant in the gut (4%). *Pantoea* spp. are gut symbionts of stink bugs (Taylor et al., 2014), while the microsporidia, *Nosema* spp., are parasitic (Hajek et al., 2017). Interestingly, several of the *Nosema* and *Pantoea* transcripts encoded putative proteases and nucleases. A putative subtilisin-like protease (other protease-6; Supplementary Table 2) is also likely to be derived from *Nosema* (Bott et al., 1988). Whether proteases and nucleases expressed by these symbiotic microorganisms function in digestion remains to be determined. Both *Nosema* and *Pantoea* transcripts were similarly detected in the whole insect transcriptomes of *H. halys* (Sparks et al., 2014), consistent with the importance of symbionts in the overall survival and success of *H. halys* in the field (Taylor et al., 2014) and in the distribution of *Nosema* spp. across stink bug species.

In conclusion, we have identified the transcripts for 269 proteases and nucleases from the *N. viridula* ASG, PSG and gut transcriptomes. We provide a comprehensive analysis of the relative transcript abundance in the three digestive tissues based on both transcriptomic and proteomic datasets. The differential relative abundance of putative digestive enzyme transcripts highlights the use of distinct digestive enzymes by gut and salivary gland toward the highly evolved and unique digestive physiology of this insect. To be effective, protein- or nucleotide-based bioactives must withstand enzymatic exposure during both extra-oral and gut-based digestion for stink bug suppression.

Author contributions statement

S.L. conducted bioinformatics analyses. P.R.L. carried out tissue dissections and conducted RT-qPCR. B.C.B. conceived the study and contributed to design of experiments. All authors contributed to the writing and review of the manuscript.

Additional information

Competing financial interests

The authors declare no competing financial interests.

Acknowledgements

The authors thank Dr. Jeffrey A. Davis, Louisiana State University, USA for provision of *N. viridula* adults. This work was supported by the National Science Foundation I/UCRC, the Center for Arthropod Management Technologies (grant numbers IIP-1338775 and IIP-1821914), and by industry partners.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ibmb.2018.10.003.

References

Bansal, R., Michel, A., 2018. Expansion of cytochrome P450 and cathepsin genes in the generalist herbivore brown marmorated stink bug. BMC Genomics 19, 60.
Bao, Y.Y., Qin, X., Yu, B., Chen, L.B., Wang, Z.C., Zhang, C.X., 2014. Genomic insights into the serine protease gene family and expression profile analysis in the planthopper, *Nilaparvata lugens*. BMC Genomics 15, 507.
Barillas-Mury, C.V., Noriega, F.G., Wells, M.A., 1995. Early trypsin activity is part of the signal transduction system that activates transcription of the late trypsin gene in the midgut of the mosquito, *Aedes aegypti*. Insect Biochem. Mol. Biol. 25, 241–246.
Bott, R., Ultsch, M., Kossiakoff, A., Graycar, T., Katz, B., Power, S., 1988. The three-dimensional structure of *Bacillus amyloliquefaciens* subtilisin at 1.8 A and an analysis of the structural consequences of peroxide inactivation. J. Biol. Chem. 263, 7895–7906.
Bown, D.P., Wilkinson, H.S., Gatehouse, J.A., 1997. Differentially regulated inhibitor-sensitive and insensitive protease genes from the phytophagous insect pest, *Helicoverpa armigera*, are members of complex multigene families. Insect Biochem. Mol. Biol. 27, 625–638.
Bravo, A., Likitvivatanavong, S., Gill, S.S., Soberon, M., 2011. *Bacillus thuringiensis*: a story of a successful bioinsecticide. Insect Biochem. Mol. Biol. 41, 423–431.
Cao, X., He, Y., Hu, Y., Zhang, X., Wang, Y., Zou, Z., Chen, Y., Blissard, G.W., Kanost, M.R., Jiang, H., 2015. Sequence conservation, phylogenetic relationships, and expression profiles of nondigestive serine proteases and serine protease homologs in *Manduca sexta*. Insect Biochem. Mol. Biol. 62, 51–63.
CEPEA/ESALQ, A., 2017. Impacto econômico de pragas agrícolas no Brasil.
Corrêa-Ferreira, B.S., Azevedo, J.D., 2002. Soybean seed damage by different species of stink bugs. Agric. For. Entomol. 4, 145–150.
Gouin, A., Bretaudeau, A., Nam, K., Gimenez, S., Aury, J.M., Duvic, B., Hilliou, F., Durand, N., Montagne, N., Darboux, I., Kuwar, S., Chertemps, T., Siaussat, D., Bretschneider, A., Mone, Y., Ahn, S.J., Hanniger, S., Grenet, A.G., Neunemann, D., Maumus, F., Luyten, I., Labadie, K., Xu, W., Koutroumpa, F., Escoubas, J.M., Llopis, A., Maibeche-Coisne, M., Salasc, F., Tomar, A., Anderson, A.R., Khan, S.A., Dumas, P., Orsucci, M., Guy, J., Belser, C., Alberti, A., Noel, B., Couloux, A., Mercier, J., Nidelet, S., Dubois, E., Liu, N.Y., Boulogne, I., Mirabeau, O., Le Goff, G., Gordon, K., Oakeshott, J., Consoli, F.L., Volkoff, A.N., Fescemyer, H.W., Marden, J.H., Luthe, D.S., Herrero, S., Heckel, D.G., Wincker, P., Kergoat, G.J., Amselem, J., Quesneville, H., Groot, A.T., Jacquin-Joly, E., Negre, N., Lemaitre, C., Legeai, F., d'Alencon, E., Fournier, P., 2017. Two genomes of highly polyphagous lepidopteran pests (*Spodoptera frugiperda*, Noctuidae) with different host-plant ranges. Sci. Rep. 7, 11816.
Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B.W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., Regev, A., 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. 29, 644–652.
Hajek, A.E., Solter, L.F., Maddox, J.V., Huang, W.F., Estep, A.S., Krawczyk, G., Weber, D.C., Hoelmer, K.A., Sanscrainte, N.D., Becnel, J.J., 2017, 65(3):315-330. *Nosema maddoxi* sp. nov. (Microsporidia, Nosematidae), a widespread pathogen of the green stink bug *Chinavia hilaris* (Say) and the Brown marmorated stink bug *Halyomorpha halys* (Stal). J. Eukaryot. Microbiol. 65 (3), 315–330.
Ioannidis, P., Lu, Y., Kumar, N., Creasy, T., Daugherty, S., Chibucos, M.C., Orvis, J., Shetty, A., Ott, S., Flowers, M., Sengamalay, N., Tallon, L.J., Pick, L., Dunning Hotopp, J.C., 2014. Rapid transcriptome sequencing of an invasive pest, the brown marmorated stink bug *Halyomorpha halys*. BMC Genomics 15, 738.
Kiritani, K., Hokyo, H., Kimura, K., Nakasuji, F., 1965. Imaginal dispersal of southern green stink bug, *Nezara viridula* (L.) in relation to feeding and oviposition. J. Appl. Entomol. Zool. 9, 291–297.
Kumar, S.M., Sahayaraj, K., 2012. Gross morphology and histology of head and salivary apparatus of the predatory bug, *Rhynocoris marginatus*. J. Insect Sci. 12, 19.
Kutsukake, M., Shibao, H., Nikoh, N., Morioka, M., Tamura, T., Hoshino, T., Ohgiya, S., Fukatsu, T., 2004. Venomous protease of aphid soldier for colony defense. Proc. Natl. Acad. Sci. U. S. A. 101, 11338–11343.
Langmead, B., Trapnell, C., Pop, M., Salzberg, S.L., 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 10, R25.
Lehane, M.J., Müller, H.M., Crisanti, A., 1996. Mechanisms controlling the synthesis and secretion of digestive enzymes in insects. In: Lehane, M.J., Billingsley, P.F. (Eds.), Biology of the Insect Midgut. Chapman & Hall, London, pp. 195–205.
Li, B., Dewey, C.N., 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinf. 12, 323.
Li, W.H., Yang, J., Gu, X., 2005. Expression divergence between duplicate genes. Trends Genet. 21, 602–607.
Liu, S., Chougule, N.P., Vijayendran, D., Bonning, B.C., 2012. Deep sequencing of the transcriptomes of soybean aphid and associated endosymbionts. PLoS One 7, e45161.
Livak, K.J., Schmittgen, T.D., 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods 25, 402–408.
Lomate, P.R., Bonning, B.C., 2016. Distinct properties of proteases and nucleases in the gut, salivary gland and saliva of southern green stink bug, *Nezara viridula*. Sci. Rep. 6, 27587.
Miles, P.W., 1972. The saliva of Hemiptera. Adv. Insect Physiol. 9, 183–255.
Mishra, M., Lomate, P.R., Joshi, R.S., Punekar, S.A., Gupta, V.S., Giri, A.P., 2015. Ecological turmoil in evolutionary dynamics of plant-insect interactions: defense to offence. Planta 242, 761–771.
Nakamura, A.M., Chahad-Ehlers, S., Lima, A.L., Taniguti, C.H., Sobrinho Jr., I., Torres, F.R., de Brito, R.A., 2016. Reference genes for accessing differential expression among developmental stages and analysis of differential expression of OBP genes in *Anastrepha obliqua*. Sci. Rep. 6, 17480.
O'Neal, S.T., Anderson, R.D., Wu-Smart, J.Y., 2018. Interactions between pesticides and pathogen susceptibility in honey bees. Curr. Opin. Insect Sci. 26, 57–62.
Panizzi, A.R., 2015. Growing problem with stink bugs (Hemiptera: Heteroptera: Pentatomidae): species invasive to the U.S. and potential neotropical invaders. Am. Entomol. 61, 223–233.
Panizzi, A.R., McPherson, J.E., James, D.G., Javahery, M., McPherson, R.M., 2000. Economic importance of stink bugs (Pentatomidae). In: Schaefer, C.W., Panizzi, A.R. (Eds.), Heteroptera of Economic Importance. CRC Press, Boca Raton, FL, pp. 421–474.
Pearce, S.L., Clarke, D.F., East, P.D., Elfekih, S., Gordon, K.H.J., Jermiin, L.S., McGaughran, A., Oakeshott, J.G., Papanikolaou, A., Perera, O.P., Rane, R.V., Richards, S., Tay, W.T., Walsh, T.K., Anderson, A., Anderson, C.J., Asgari, S., Board, P.G., Bretschneider, A., Campbell, P.M., Chertemps, T., Christeller, J.T., Coppin, C.W., Downes, S.J., Duan, G., Farnsworth, C.A., Good, R.T., Han, L.B., Han, Y.C., Hatje, K., Horne, I., Huang, Y.P., Hughes, D.S.T., Jacquin-Joly, E., James, W., Jhangiani, S., Kollmar, M., Kuwar, S.S., Li, S., Liu, N.Y., Maibeche, M.T., Miller, J.R., Montagne, N., Perry, T., Qu, J., Song, S.V., Sutton, G.G., Vogel, H., Walenz, B.P., Xu, W., Zhang, H.J., Zou, Z., Batterham, P., Edwards, O.R., Feyereisen, R., Gibbs, R.A., Heckel, D.G., McGrath, A., Robin, C., Scherer, S.E., Worley, K.C., Wu, Y.D., 2017. Genomic innovations, transcriptional plasticity and gene loss underlying the evolution and divergence of two highly polyphagous and invasive *Helicoverpa* pest species. BMC Biol. 15, 63.
Petersen, T.N., Brunak, S., von Heigne, G., Nielsen, H., 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat. Methods 8, 785–786.
Prado, S.S., Almeida, R.P., 2009. Phylogenetic placement of pentatomid stink bug gut symbionts. Curr. Microbiol. 58, 64–69.
Rispe, C., Kutsukake, M., Doublet, V., Hudaverdian, S., Legeai, F., Simon, J.C., Tagu, D., Fukatsu, T., 2008. Large gene family expansion and variable selective pressures for cathepsin B in aphids. Mol. Biol. Evol. 25, 5–17.
Scott, J.G., Michel, K., Bartholomay, L.C., Siegfried, B.D., Hunter, W.B., Smagghe, G., Zhu, K.Y., Douglas, A.E., 2013. Towards the elements of successful insect RNAi. J. Insect Physiol. 59, 1212–1221.
Snodgrass, G.L., Scott, W.P., 2003. Effect of ULV malathion use in boll weevil (Coleoptera: Curculionidae) eradication on resistance in the tarnished plant bug (Heteroptera: Miridae). J. Econ. Entomol. 96, 902–908.
Sparks, M.E., Shelby, K.S., Kuhar, D., Gundersen-Rindal, D.E., 2014. Transcriptome of the invasive brown marmorated stink bug, *Halyomorpha halys* (Stal) (Heteroptera: Pentatomidae). PLoS One 9, e111646.
Taylor, C.M., Coffey, P.L., DeLay, B.D., Dively, G.P., 2014. The importance of gut symbionts in the development of the brown marmorated stink bug, *Halyomorpha halys* (Stal). PLoS One 9, e90312.
Terra, W.R., Ferreira, C., 1994. Insect digestive enzymes-properties, compartmentalization and function. Comp. Biochem. Physiol., B 109, 1–62.
Todd, J.W., 1989. Ecology and behavior of *Nezara viridula*. Annu. Rev. Entomol. 34, 273–292.
Will, T., Steckbauer, K., Hardt, M., van Bel, A.J., 2012. Aphid gel saliva: sheath structure, protein composition and secretory dependence on stylet-tip milieu. PLoS One 7, e46903.
Willrich, M.M., Leonard, B.R., Cook, C.R., 2003. Laboratory and field evaluation of insecticide toxicity to stink bugs (Heteroptera: Pentatomidae). J. Cotton Sci. 7, 156–163.
Zhang, H., Li, H.C., Miao, X.X., 2013. Feasibility, limitation and possible solutions of RNAi-based technology for insect pest control. Insect Sci. 20, 15–30.
Zhu-Salzman, K., Zeng, R., 2015. Insect response to plant defensive protease inhibitors. Annu. Rev. Entomol. 60, 233–252.