

Research paper

Filling in the gaps: A reevaluation of the *Lygus hesperus* peptidome using an expanded *de novo* assembled transcriptome and molecular cloningJ. Joe Hull^{a,*}, Roni J. Gross^a, Colin S. Brent^a, Andrew E. Christie^b^a Pest Management and Biocontrol Research Unit, US Arid Land Agricultural Research Center, USDA Agricultural Research Services, Maricopa, AZ 85138, USA^b Békésy Laboratory of Neurobiology, Pacific Biosciences Research Center, School of Ocean and Earth Science and Technology, University of Hawaii at Manoa, 1993 East-West Road, Honolulu, HI 96822, USA

ARTICLE INFO

Keywords:

In silico transcriptome mining
RT-PCR
Western tarnished plant bug
Arthropoda
Hexapoda
Insecta
Agatoxin-like peptide

ABSTRACT

Peptides are the largest and most diverse class of molecules modulating physiology and behavior. Previously, we predicted a peptidome for the western tarnished plant bug, *Lygus hesperus*, using transcriptomic data produced from whole individuals. A potential limitation of that analysis was the masking of underrepresented genes, in particular tissue-specific transcripts. Here, we reassessed the *L. hesperus* peptidome using a more comprehensive dataset comprised of the previous transcriptomic data as well as tissue-specific reads produced from heads and accessory glands. This augmented assembly significantly improves coverage depth providing confirmatory transcripts for essentially all of the previously identified families and new transcripts encoding a number of new peptide precursors corresponding to 14 peptide families. Several families not targeted in our initial study were identified in the expanded assembly, including agatoxin-like peptide, CNMamide, neuropeptide-like precursor 1, and periviscerokinin. To increase confidence in the *in silico* data, open reading frames of a subset of the newly identified transcripts were amplified using RT-PCR and sequence validated. Further PCR-based profiling of the putative *L. hesperus* agatoxin-like peptide precursor revealed evidence of alternative splicing with near ubiquitous expression across *L. hesperus* development, suggesting the peptide serves functional roles beyond that of a toxin. The peptides predicted here, in combination with those identified in our earlier study, expand the *L. hesperus* peptidome to 42 family members and provide an improved platform for initiating molecular and physiological investigations into peptidergic functionality in this non-model agricultural pest.

1. Introduction

While many classes of molecules have been shown to play roles in the modulation of physiological and behavioral control systems (Christie, 2011), peptides are by far the largest and most diverse of these groups (Kastin, 2006). Among the Insecta, the ~50 peptide- or neuropeptide-encoding gene families identified to date can be roughly divided into two groups consisting of basal peptides found in most genomes and a variable set of peptides that are species or order dependent (Hauser et al., 2010). Given their regulatory roles in critical biological functions, such as reproduction, development, growth, and feeding (Altstein and Nässel, 2010), these peptidergic signaling systems have been proposed as targets of next generation insect pest control strategies (Altstein, 2004; Audsley and Down, 2015; Nachman et al., 2009; Scherckenbeck and Zdobinsky, 2009; Van Hiel et al., 2010; Xie et al., 2015; Zhang et al., 2015a; 2015b).

The western tarnished plant bug, *Lygus hesperus*, is a pest species of concern in arid regions of the western and southwestern United States, where it targets an extensive range of host plants, including many economically important food, fiber, and seed crops (Naranjo et al., 2011; Ritter et al., 2010; Scott, 1977; Wheeler, 2001). Previously, we employed *in silico* transcriptome mining to predict a peptidome for *L. hesperus* (Christie et al., 2017). Approximately 120 distinct peptides were identified in GBRD01000000 (BioProject PRJNA210219; unpublished direct GenBank submission) and GBHO00000000 (BioProject PRJNA238835; Hull et al., 2014). The predicted peptidome includes isoforms of 23 generally recognized arthropod peptide families: allatostatin A (AST-A), allatostatin B (AST-B), allatostatin C (AST-C), allatotropin (ATR), bursicon, CCHamide, corazonin (CRZ), crustacean cardioactive peptide (CCAP), diuretic hormone 31 (DH31), GSE-FLamide, insulin-like peptide (ILP), ion transport peptide (ITP), myo-suppressin (MS), natalisin, neuroparsin, neuropeptide F (NPF),

* Corresponding author.

E-mail address: joe.hull@usda.gov (J.J. Hull).<https://doi.org/10.1016/j.ygcen.2020.113708>

Received 30 September 2020; Received in revised form 22 December 2020; Accepted 29 December 2020

Available online 1 January 2021

0016-6480/Published by Elsevier Inc.

orcokinin, orcomyotropin, pyrokinin (PK), short neuropeptide F (sNPF), SIFamide, sulfakinin and tachykinin-related peptide (TRP), as well as a large suite of precursor-related peptides (PRPs) that may or may not be bioactive in their own right. While the initial dataset for native *L. hesperus* peptides has been a valuable foundational resource for this species and related hemipteran pests, the peptidome's broader applicability is limited by the absence of numerous peptide groups typically present across the Arthropoda (Christie et al., 2010). For example, no transcripts encoding putative proctolin precursors were identified (Christie et al., 2017) despite its presence in numerous arthropods and hemipteran pests (Konopińska and Rosiński, 1999; Isaac et al., 2004; Lavore et al., 2018; Wang et al., 2018; Li et al., 2020). While it is theoretically possible that these "missing" peptide groups have been lost in *L. hesperus*, their broad phylogenetic conservation suggests that they were simply overlooked during data mining.

Subsequent to our initial report of the *L. hesperus* peptidome (Christie et al., 2017), a new transcriptome, GDHC01000000 (BioProject No. PRJNA284294; Tassone et al., 2016), with greater depth of coverage was generated by combining data used to produce the earlier assemblies with new tissue-specific reads. The assembly includes reads from head (brain) and male reproductive tissue, both of which are known to be rich sources of biologically active peptides (e.g., Bao et al., 2015; Christie, 2016). Here, we reinvestigated the *L. hesperus* peptidome using the expanded assembly. In addition to reidentifying the previously reported peptide precursors, we found precursor transcripts looked for but not found in the initial assembly and identified several peptide groups not targeted in the earlier study. To confirm the validity of the *in silico* predictions, a subset of the newly identified transcripts were validated by RT-PCR. The peptides predicted from the transcripts identified here significantly expands the extant *L. hesperus* peptidome, which now potentially consists of 386 peptides representing 42 families. Furthermore, we provide evidence for alternative splicing and near ubiquitous expression of agatoxin-like peptides (ALPs), a recently recognized group of highly conserved, cysteine-rich peptides of unknown biological function in arthropods (Sturm et al., 2016; Veenstra, 2016a).

2. Materials and methods

2.1. *In silico* transcriptome mining

Searches of the expanded *L. hesperus* transcriptome, GDHC01000000, were conducted using a well-established protocol (e.g., Christie, 2008a; 2008b; 2015; Christie and Chi, 2015; Christie et al., 2017). In brief, the tBLASTn (National Center for Biotechnology Information, Bethesda, MD; <http://blast.ncbi.nlm.nih.gov/Blast.cgi>) online database was set to Transcriptome Shotgun Assembly (TSA) and restricted to data from BioProject No. PRJNA284294 (Tassone et al., 2016). Known arthropod peptide precursors, many previously identified from *L. hesperus* itself (Christie et al., 2017), were used as the query sequences for the BLAST searches. The complete list of peptide families searched for in this study, the specific queries used for the BLAST searches, and the accession numbers of the *L. hesperus* transcripts identified as encoding putative homologs are provided in Supplemental Table 1. The peptide families used here as search queries were also used to re-mine the previous assemblies (i.e., BioProjects PRJNA210219 and PRJNA238835). Given potential confounding issues that may arise from combining multiple, overlapping datasets, as was done here to construct the expanded *L. hesperus* transcriptome, we also mined head specific datasets (SRA accession #s SRX1072689, SRX1155625, SRX1155629) that were bundled with BioProject PRJNA284294. Database searches were performed using tBLASTn with an *e* value set to $1e^{-1}$ and all hits re-evaluated by BLASTx using an Insecta (taxid: 50557) restricted NCBI nr database.

2.2. Peptide structural prediction

A well-established workflow was used to predict the mature structures of *L. hesperus* peptides (e.g. Christie, 2008a; 2008b; 2015; Christie and Chi, 2015; Christie et al., 2017). In brief, all hits returned by a given BLAST search were translated using the Translate tool of ExPASy (<http://web.expasy.org/translate/>) and assessed for completeness. Proteins listed as "full-length" exhibit a start methionine and are flanked on their carboxyl (C)-terminus by a stop codon. Start codons were determined either as the first in-frame codon downstream of a stop site or in relation to homologous sequences. Proteins described here as "partial" lacked either a start methionine, referred to as C-terminal partial proteins, or a stop codon, referred to as amino (N)-terminal partial proteins. Next, each deduced full-length or N-terminal partial precursor protein was assessed for the presence of a signal peptide using the online program SignalP 3.0 (<http://www.cbs.dtu.dk/services/SignalP/>; Bendtsen et al., 2004). Prohormone cleavage sites were identified based on the information presented in Veenstra (2000) and/or by homology to known arthropod pre-/prohormone processing schemes. When present, the sulfation state of tyrosine residues was predicted by homology to known peptide isoforms or by using the online program Sulfinator (<http://www.expasy.org/tools/sulfinator/>; Monigatti et al., 2002). Disulfide bonding between cysteine residues was predicted by homology to known peptide isoforms or by using the online program DiANNA (<http://clavius.bc.edu/~clotelab/DIANNA/>; Ferrè and Clote, 2005). Other post-translational modifications (i.e., cyclization of N-terminal glutamine/glutamic acid residues and C-terminal amidation at glycine residues) were predicted by homology to known arthropod peptides. Peptide alignments were initially done using the online program MAFFT version 7 (<http://mafft.cbrc.jp/alignment/software/>; Katoh and Standley, 2013) and then further refined for publication using MUSCLE (Edgar, 2004) implemented in either Geneious R10 or Geneious Prime 2020.1.2 (Biomatters Ltd., Auckland, New Zealand). To determine amino acid conservation between selected precursor proteins or peptides, the sequences in question were aligned using MUSCLE, and amino acid identity/similarity subsequently determined based on the alignment output.

2.3. Phylogenetic analyses

The phylogenetic relationships of selected *L. hesperus* peptides, eclosion hormone (EH) and prothoracicotrophic hormone (PTTH), were inferred from multiple sequence alignments constructed in MUSCLE with evolutionary analyses based on the Neighbor-Joining method (Saitou and Nei, 1987) conducted in MEGA X (Kumar et al., 2018; Stecher et al., 2020). The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches in each tree, which were drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Poisson correction method (Zuckerkandl and Pauling, 1965) and are given in the units of the number of amino acid substitutions per site. All positions with less than 95% site coverage were eliminated, i.e., fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position (partial deletion option). The EH analysis involved 25 amino acid sequences for a total of 70 positions in the final dataset and an optimal tree with a sum of branch length = 5.2. The PTTH analysis involved 14 amino acid sequences for a total of 88 positions in the final dataset and an optimal tree with a sum of branch length = 7.6. Accession numbers for sequences used in the phylogenetic analyses are provided in Supplemental Table 2.

2.4. Cloning peptide-encoding transcripts

To increase confidence for the *in silico* peptide predictions, a subset of 31 peptide-encoding transcripts were PCR amplified, sub-cloned, and

sequenced. Oligonucleotide primers (Supplemental Table 3) designed to amplify the complete open reading frames (ORFs) were generated from nucleotide sequences of the targeted transcripts using Primer3 (Untergasser et al., 2012). Insects used for molecular analyses were acquired from an in-house (US Arid Land Agricultural Research Center, USDA Agricultural Research Services, Maricopa, AZ) colony (Brent and Hull, 2014). Total RNA was isolated in triplicate from sets of three 7-day old mixed sex adult *L. hesperus* heads using TRI Reagent Solution (Life Technologies/Ambion, Carlsbad, CA) with an RNeasy mini kit (Qiagen, Germantown, MD) and a QIAcube automated nucleic acid isolation system (Qiagen). RNA quality and quantity were determined spectrophotometrically using the Take3 module on a Synergy H4 Hybrid Multi-Mode Microplate Reader (Biotek Instruments, Winooski, VT). Total RNA (500 ng) was treated with DNase I (New England Biolabs, Ipswich, MA) to remove any residual genomic DNA. cDNAs were generated from 500-ng of DNA-free RNA with a SuperScript III First-Strand Synthesis System (Life Technologies) and custom-made random pentadecamers (Integrated DNA Technologies, San Diego, CA). Peptide precursor-encoding transcripts were amplified in a 20- μ L reaction volume using SapphireAmp Fast PCR Master Mix (Clontech Laboratories Inc., Mountain View, CA) combined with 12.5 ng cDNA and 0.2 μ M sense and antisense primers on a Biometra TRIO (Göttingen, Germany) thermocycler. PCR conditions for most of the transcripts consisted of an initial denaturation at 95 °C for 2 min followed by 40 cycles of 95 °C for 20 sec, 56 °C for 20 sec, and 72 °C for 60 sec, and concluded at 72 °C for 5 min. For the putative adipokinetic hormone/corazonin-like peptide (ACP) and periviscerokinin (PVK) transcripts, identical PCR conditions were used with the exception of annealing temperature, which was adjusted to 60 °C. PCR products were electrophoresed using a Tris-Acetate-EDTA (TAE) buffer system on 2% agarose gels stained with SYBR Safe (Life Technologies). Products were cloned into the pCR2.1TOPO TA cloning vector (Life Technologies) and sequenced using either the Arizona State University DNA Core laboratory (Tempe, AZ) or Retrogen Inc. (San Diego, CA). Consensus sequences for the 31 cloned transcripts have been deposited with NCBI under accession numbers MT210013-MT210031, MT895866-MT895875, and MW366893-MW366894. Gel images were obtained using an Azure 200 Gel Imaging Workstation (Azure Biosystems, Dublin, CA) and processed in Photoshop CS6 v13.0 (Adobe Systems Inc., San Jose, CA).

2.5. Expression profile of ALP transcripts

Total RNAs were isolated as above from two biological replicates of pooled whole eggs, nymphs (instars 1–5), and mixed sex adults (0, 7, and 20 days post-adult emergence). Day 7–9 mixed sex adults were used to provide samples from head, thorax, abdomen, midgut, hindgut, and Malpighian tubule. Additionally, the female seminal depository and ovaries and the male lateral and medial accessory glands and testes were examined. Adult maturation is typically complete by 7 days after eclosion for both females and males (Brent 2010). cDNAs were prepared from 500 ng DNase I-treated total RNAs using Superscript III reverse transcriptase as above. Each cDNA was PCR-screened for ALP transcript expression using primers (Supplemental Table 3) designed to amplify the complete ORF. A 501-bp fragment of *L. hesperus* actin (GBHO01007453.1) was likewise amplified. PCR was performed as before with cycling conditions consisting of: 95 °C for 2 min followed by 40 cycles at 95 °C for 20 s, 56 °C for 20 s, 72 °C for 60 s, and a final extension at 72 °C for 5 min. PCR products were separated on 3% agarose gels. Gel images were obtained as described above.

3. Results and discussion

3.1. Identification of putative peptide precursor-encoding transcripts

To develop a more complete *L. hesperus* peptidome, we mined an updated *L. hesperus* transcriptomic dataset (Tassone et al., 2016)

constructed from multiple tissue specific sequence reads in addition to the initial whole-body data (Hull et al., 2014). Our query set consisted of 50 known prepropeptides (Supplemental Table 1). These included all of the peptides searched for in the previous study (Christie et al., 2017) and 23 sequences corresponding to adipokinetic hormone (AKH), agatoxin-like peptides (ALPs), allatostatin CC (AST-CC), AST-CCC, arginine-vasopressin-like peptide (AVP), CCHamide 1, CCHamide 2, CNMamide, eclosion hormone (EH), ecdysis-triggering hormone (ETH), elevenin, glycoprotein hormone α 2 (GPA2), glycoprotein hormone β 4 (GPB5), natalisin, neuropeptide-like precursors 1–4 (NPLP1–4), periviscerokinin (PVK), pigment dispersing factor (PDF), proctolin, prothoracicotropic hormone (PTTH), and trissin. Because sequence similarity among prepropeptides is frequently limited to specific processed peptide regions, we included representative prepropeptides from at least two species in our search. Transcripts encoding precursors for most peptide families were identified (Table 1) with BLASTx hits supporting results of the database mining (Supplemental Table 4). Queries that failed to yield a hit were re-screened against head-specific reads, which resulted in identification of an additional nine precursors (Table 1). No transcripts were found in any of the datasets (either this study or the previous study) that encode homologs of AVP, CCHamide 2, GPA2, GPB5, inotocin, NPLP 2–4, or trissin. These absences may indicate that these genes are expressed conditionally or in such low abundance that they are poorly represented in transcriptomic datasets. However, it may also reflect the evolutionary loss observed for a number of peptides whose presence is species- or order-dependent (Hauser et al., 2010); six of the missing peptides (GPA2, GPB5, inotocin, NPLP2–4) are among those reported to be lineage-dependent. Inotocin expression is limited in heteropterans (Liutkeviciute et al., 2016), and trissin appears to be poorly conserved in hemipterans with no homologous sequence detected in *Nilaparvata lugens* (Tanaka et al., 2014), *Nezara viridula* (Lavore et al., 2018), *Rhodnius prolixus* (Ons et al., 2011), *Triatoma infestans* (Traverso et al., 2016), *Bemisia tabaci* (Li et al., 2020), or *Diaphorina citri* (Wang et al., 2018). Although identification of precursors for NPLP 2–4 has been largely limited to drosophilids, an NPLP-2 homolog is present in *Apis mellifera* (Hummon et al., 2006), and NPLP-4 is found in at least one other species of fly (Li et al., 2009). However, a recent study suggests that NPLP-2 is induced by heat stress (Rommelaere et al., 2019), and may have been missed if expressed at very low levels in *L. hesperus* sampled under normal conditions. Further, that study found NPLP-2 serves as an apolipoprotein rather than as a neuropeptide. An NPLP-3 precursor annotated in *D. citri* (Wang et al., 2018) and a number of coleopteran species (*Anoplophora glabripennis*, XP_018572912.1; *Lepitotarsa decemlineata*, XP_023023955.1; and *Onthophagus taurus*, XP_022901951) has significant sequence similarity with pupal cuticle proteins suggesting potential misannotation. For this reason, a putative *L. hesperus* NPLP-3 (GBHO01017385.1) with high homology to cuticle proteins was rejected as a *bona fide* precursor.

Our initial *L. hesperus* peptidome study yielded 119 distinct peptides corresponding to 23 peptide families (Christie et al., 2017). Our expanded peptidome is predicted to encompass 42 peptide families (Table 1) and consist of 386 peptides, of which 120 are predicted to be bioactive and 266 are precursor-related (Supplemental Fig. 1). The number of peptide families identified is comparable to the range of precursor genes annotated in other hemipteran transcriptomic/genomic datasets (Fig. 1), suggesting that the new peptidome is a more complete reflection of the *L. hesperus* peptidergic signaling system. Surprisingly, some of the variants identified via PCR-based methods (e.g. calcitonin-like diuretic hormone precursor variant B, KX584422; calcitonin-like diuretic hormone precursor variant C, KX584423; and allatotropin precursor variant B; KX584420) were not evident in the mined datasets. As with the “missing” peptide families, the absence of these variants may indicate that they are low abundance transcripts that may be conditionally, spatially, and/or temporally expressed.

Analysis of potential post-translational modifications suggests that 116 of the predicted peptides are C-terminally amidated (Supplemental

Table 1Peptide precursors identified in transcriptomic/genomic datasets for *L. hesperus* and four other hemipterans.

Peptide	<i>L. hesperus</i>			<i>N. lugens</i> ^a	<i>D. citri</i> ^a	<i>N. viridula</i> ^a	<i>A. pisum</i> ^a
	previous study ^a	current study	head specific ^b				
adipokinetic hormone (AKH)	–	–	+	+	+	+	+
AKH/corazonin-related peptide (ACP)	+	+	ns	+	+	+	–
agatoxin-like peptide (ALP)	ns	+	ns	ns	ns	ns	ns
allatostatin A (AST-A)	+	+	ns	+	+	+	+
allatostatin B/myoinhibitory peptide (AST-B/MIP)	+	+	ns	+	+	+	+
allatostatin C (AST-C)	+ ^c	–	–	+ ^c	+ ^c	–	+ ^c
allatostatin CC (AST-CC)	ns	–	+	+	+	+	+
allatostatin CCC (AST-CCC)	ns	+	ns	ns	ns	+	ns
allatotropin (AT)	+	+	ns	+	+	+	+
arginine/vasopressin-like (AVP)	ns	–	–	+	+	+	–
bursicon alpha (Burs-α)	–	+	ns	+	–	+	+
bursicon beta (Burs-β)	+	–	+	+	+	+	+
periviscerokin (PVK)/capability (CAPA)	ns	+	ns	+	+	+	+
CCHamide 1 (CCHa 1)	+	+	ns	+	+	+	+
CCHamide 2 (CCHa 2)	–	–	–	+	+	+	+
CNMamide 1 (CNMa)	ns	–	+	ns	+	+	–
corazonin (Crz)	+	+	ns	+	+	+	–
crustacean cardioactive peptide (CCAP)	+	+	ns	+	+	+	+
diuretic hormone 31 (DH31)	+	+	ns	+	+	+	+
diuretic hormone 44 (DH44)	–	+	ns	+	+	+	+
ecdysis triggering hormone (ETH)	–	–	+	+	+	+	+
eclosion hormone (EH)	–	–	++	++	++	–	++
elevenin (Ele)	ns	–	+	+	+	+	+
FMRamide (FMRFa)	–	+	ns	+	–	+	+
glycoprotein hormone A2 (GPA2)	ns	–	–	+	–	+	+
glycoprotein hormone B5 (GB5)	ns	–	–	+	+	–	+
GSEFLa/IDSLR/EFLa (GSEFLa)	+	+	ns	–	–	+	ns
inotocin (Ino)	–	–	–	ns	ns	–	ns
insulin-related peptide (IRP)	+	+	ns	++	++	+	++
ion transport peptide (ITP)	–	–	+	+	–	–	+
ion transport peptide-like (ITPL)	+	+	+	+	+	+	+
leucokinin/kinin (LK/kin)	–	+	ns	+	+	+	+
myosuppressin (MS)	+	+	ns	+	+	+	+
natalisin (NTL)	ns	+	ns	–	+	–	+
neuroparsin (NP)	+	+	ns	+	+	++	–
neuropeptide F (NPF)	+	+	ns	+	–	+	+
neuropeptide-like precursor 1 (NPLP-1)	ns	+	ns	+	+	+	+
neuropeptide-like precursor 2 (NPLP-2)	ns	–	–	–	–	–	ns
neuropeptide-like precursor 3 (NPLP-3)	ns	–	–	+	+	–	ns
neuropeptide-like precursor 4 (NPLP-4)	ns	–	–	+	–	–	ns
orcokinin (OK)	+	+	ns	+	+	+	+
pigment dispersing factor (PDF)	–	–	+	+	+	+	–
proctolin (Proc)	–	–	+	+	+	+	+
prothoracicotropic hormone (PTTH)	ns	+	ns	+	+	–	–
pyrokinin (PK)	+	+	ns	+	+	–	+
RYamide (RYa)	–	+	ns	+	+	+	+
short neuropeptide F (sNPF)	+	+	ns	+	+	+	+
SIFamide (SIFa)	+	–	+	+	+	+	+
sulfakinin (SK)	+	+	ns	+	–	+	–
tachykinin-related peptide (TRP)	+	+	ns	+	+	+	+
trissin (Tris)	ns	–	–	–	–	–	–

ns – not searched; (+) = identified; (++) = multiple genes identified; (–) = not identified.

^a Data: previous Lygus (Christie et al. 2017); *N. lugens* (Tanaka et al. 2014); *D. citri* (Wang et al. 2018); *N. viridula* (Lavore et al. 2018); *A. pisum* (Huybrechts et al. 2010).^b Data from SRA accession #s SRX1072689, SRX1155625, SRX1155629.^c Updated sequence similarities suggest initial annotations of AST-C should be AST-CCC.

Fig. 1). However, given potential differential expression of the proconvertases and/or varied utilization of peptide cleavage sites (Southey et al., 2008; Neupert et al., 2018), both the number of peptides generated and the extent of C-terminal amidation may differ from that predicted here. In addition to alternative proconvertase usage, peptide diversity can also be amplified by alternative splicing events and, less commonly, by gene duplication (Yeoh et al., 2017). Although many of the peptide family members predicted here, and previously (Christie et al., 2017), appear to be derived from single precursor genes, multiple variants are predicted for 15 families (Supplemental Fig. 1). Results from the current data mining approach coupled with our previous efforts (Christie et al., 2017) suggest alternative splicing accounts for much of the observed diversity (e.g. ALP, DH31, CCHamide, ATR, CCAP,

GSEFLamide, and orcokinin). We also found evidence suggesting alternative promoter usage may play a role in generating the CCHamide and ITP/ITP-like variants as the presumptive 5' untranslated regions for the CCHamide v3, CCHamide v4, and ITPL v2 precursors differs from that of their other variants (Fig. 2). Similar 5' untranslated region (UTR) variation, which is reported to control the spatiotemporal expression of ITPL transcripts, has been reported in *N. lugens* (Yu et al., 2016). For the predicted *L. hesperus* prepropeptides, the 5' differences introduce an N-terminal extension that blocks the respective predicted signal peptides. Sequence analyses predict (>0.99 probability) the extensions to be signal anchors. Similar signal anchors have been predicted for fruitfly, tsetse fly, and termite AST-CC peptides, which are thought to function via paracrine secretion of non-neuroendocrine cells (Veenstra, 2014).

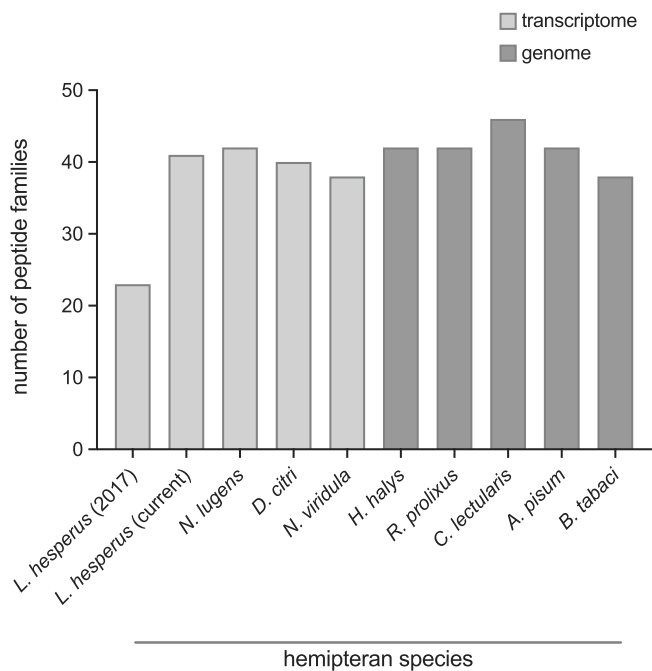


Fig. 1. Comparison of the number of peptide families predicted in various hemipteran transcriptomic or genomic datasets. Peptide family numbers are based on reported values for *L. hesperus* (Christie et al., 2017), *N. lugens* (Tanaka et al., 2014), *D. citri* (Wang et al., 2018), *N. viridula* (Lavore et al., 2018), *H. halys* (Lavore et al., 2018), *R. prolixus* (Mesquita et al., 2015), *C. lectularius* (Benoit et al., 2019), *A. pisum* (Huybrechts et al., 2010), and *B. tabaci* (Li et al., 2020).

Alternatively, the first translation initiating Met codon, which results in the signal anchor, may be bypassed for the second Met codon (i.e. the first residue in the predicted signal peptides). Regardless, the transcriptomic data support expression data for the respective precursors and the expression of both ITPL variants was confirmed by PCR cloning (accession #s MW366893-4).

Our finding that the GSEFLamide/EFLamide (also referred to as IDLSRF-like peptides) and orckinin genes appear to give rise to two distinct precursors (Supplementary Fig. 1), one containing isoforms of the peptides for which they were named, and the other having a distinct C-terminus containing completely different, structurally unrelated sets of peptides is similarly suggestive of differential expression. The presence of more than one precursor variant in these two peptide families is not without precedent (Jiang et al., 2015; Chen et al., 2015; Kotwica-Rolinska et al., 2020). The GSEFLamide encoding exon has recently been shown to be fused downstream of two exons encoding prohormone-4, with the fusion site conserved in hemipterans (Kotwica-Rolinska et al., 2020).

Precursor diversity in *L. hesperus* is also linked to gene duplication/expansion for the AKH, EH, ILP and neuroparsin families (Fig. 3 A-D). Gene duplications have been reported previously for AKH (Roller et al., 2008; Kaufmann et al., 2009; Veenstra, 2019) and EH (Veenstra, 2014). Multiple variants with varying degrees of sequence identity are predicted for *L. hesperus* ILPs and neuroparsins (Fig. 3 C-D). Typically, the number of ILP and neuroparsin paralogs vary inversely with one another, such as in the hemipteran pests *N. lugens*, *N. viridula*, *H. halys*, and *A. pisum* (Tanaka et al., 2014; Lavore et al., 2018; Huybrechts et al., 2010). This relationship between the two peptidergic systems, both of which signal through tyrosine receptor kinases, was posited as an indicator of complementary functional roles (Veenstra, 2016a) as evidenced by the synergistic effects of the two peptides on mosquito vitellogenesis (Dhara et al., 2013). In our *L. hesperus* datasets, the number of ILP and neuroparsin paralogs identified is comparable (3 – ILP; 4 –

neuroparsins), suggesting potential functional divergence of the two peptides. Alternatively, the individual paralogs may impact physiological functions via differential (spatial, temporal, and/or conditional) expression. Their role(s) in *Lygus* species, however, remain to be empirically determined.

3.2. RT-PCR amplification of select peptide precursor cDNAs

To expand our initial validation of the putative *L. hesperus* transcripts/precursor proteins, the ORFs for 31 of the predicted transcripts, including all of the precursors predicted from the head specific reads, were targeted for RT-PCR amplification. Amplimers of the expected sizes were obtained for all of the transcripts (Fig. 4). For a majority of the cloned products, nucleotide identity with the *in silico* sequences exceeded 99%. Potential variants were identified for the orckinin (B-type) product, in which all of the clones exhibited nonsynonymous nucleotide differences relative to the transcriptomic data, in which Val46 was changed to Asp and Leu62 to Gln. Similarly, nonsynonymous changes were observed in the FMRamide clones with two presumed variants identified. In variant 1, Pro58 was changed to Leu, His190 to Ser, Ala171 to Ser, and Asp196 to Asn. In variant 2, two conserved substitutions were identified; Phe193 to Tyr and Asp198 to Glu. These changes likely reflect allelic variation found in the heterogenous stock colony, and because they affect the predicted precursor related peptides, would not be expected to impact the biological activity of the encoded FMRamides. In the PCR amplified AST-CC product, a non-conserved substitution (Leu8 to Pro) was identified, but it had no effect on prediction of the signal peptide region. For the ALP variants, the longest sequence (ALP v1; MT210014), which is differentiated from the other variants by inclusion of all presumptive exons, is not represented in the expanded transcriptomic datasets (Tassone et al., 2016). A search of the earlier data (Hull et al., 2014) revealed the presence of the variant (GBRD01011935.1 and GBRD01011927.1) with no nucleotide changes between the cloned and predicted sequences. Additionally, although some of the ALP transcriptomic datasets indicated a C-terminal Glu, clones amplified using an antisense primer designed to a site in the 3' UTR failed to confirm the variable sequence, suggesting it is either an assembly artifact or that the corresponding transcript is low abundance.

3.3. Newly predicted peptide precursors

3.3.1. Peptides that function in ecdysis and development

The canonical peptidergic pathway that regulates development and ecdysis consists of five central peptide families: prothoraciotropic hormone (PTTH), ecdysis-triggering hormone (ETH), eclosion hormone (EH), crustacean cardioactive peptide (CCAP), and bursicon (de Oliveira et al., 2019). However, depending on the species, additional peptide components have been implicated including orckinins (OK) (Yamanaka et al., 2011; Wulff et al., 2017), adipokinetic hormone/corazonin-related peptide (ACP) (Hansen et al., 2010; Zandawala et al., 2015; Wahedi and Paluzzi, 2018), and corazonin (Kim et al., 2004; Žitňan et al., 2007; Hou et al., 2017). Precursors encoding each of these peptides were found either within the initial *L. hesperus* peptidome (i.e. CCAP, OK-A, bursicon β) or within the new datasets (Supplemental Fig. 1).

The *L. hesperus* PTTH homolog has characteristics typical of other insect PTTH (Fig. 5A), but like pea aphid PTTH (Barberà and Martínez-Torres, 2017), is missing one of the seven conserved Cys (i.e. Cys4) thought to be involved in intra- and intermolecular disulphide bonds. Despite this deviation, the peptide sequence clustered with other PTTH annotated peptides that have the full complement of Cys residues (Fig. 5B). While absent from a number of hemipteran datasets (e.g. *N. viridula*, *H. halys*, and *R. prolixus*), in the hemipterans in which it has been identified (*N. lugens*, BAO00973; *D. citri*, AWT50619.1; and *C. lectularius*, XP_014239433.1) the PTTH transcript encodes a peptide containing seven Cys. Given the potentially significant conformational

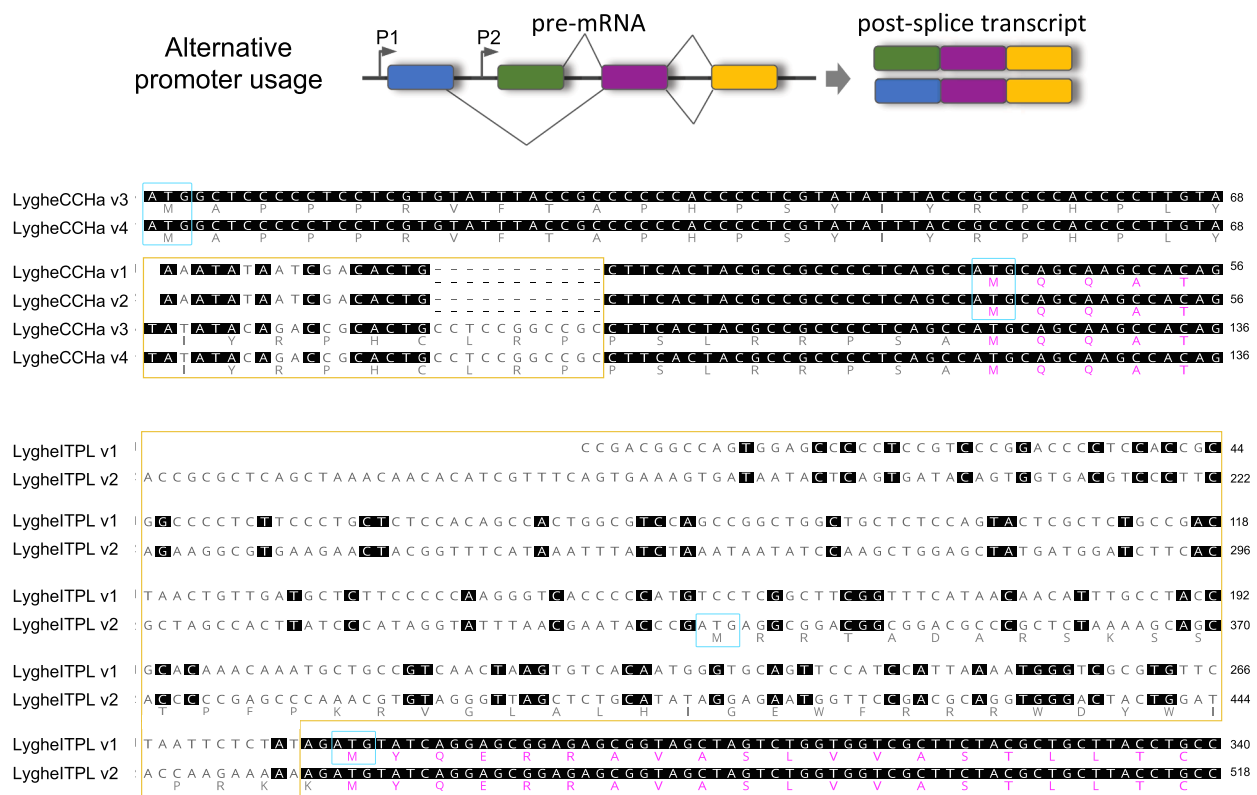


Fig. 2. Proposed alternative splicing mechanism for the *L. hesperus* CCHamide and ITP variants. Schematic of alternative promoter usage in which promoters (P1 and P2) upstream of two different exons yield mature RNA transcripts with differing 5' ends. Coding sequences for LygheCCHamide variants are shown in the upper panel. The LygheCCHamide v3 and v4 sequences correspond to nucleotides 1–136 and LygheCCHamide v1 and v2 corresponding to nucleotides –40 through 16. Not shown is the 78-nt deletion that differentiates LygheCCHamide v2 and v4. For the two LygheITP variants, the coding sequences are shown in the lower panel. The LygheITP v1 sequence corresponds to nucleotides –270 to 61, whereas LygheITP v2 correspond to nucleotide –186 to 184. The two variants are 100% identical across the remaining 293-nt coding sequence. Predicted signal anchors and signal peptides are shown in grey and pink, respectively. Regions of nucleotide variation are indicated by yellow boxes. Blue boxes indicate predicted start codons.

effects the missing Cys may have on peptide functionality, we have termed the *L. hesperus* peptide as PTH-like.

The LygheETH precursor is likewise atypical. For most species, the ETH precursor transcript encodes a >100 amino acid prepropeptide that is cleaved to yield 2–3 dodecapeptides characterized by an FFL/IKxxKxVPRL/I-amide C-terminal motif (Roller et al., 2010). In contrast, the LygheETH precursor is 96 amino acids and consists of one canonical sequence and one ETH-like sequence that lacks the C-terminal aliphatic residue and accompanying amidation (Fig. 5C). The presence of transcripts encoding similar atypical ETH precursors in the sister species *Apolygus lucorum* (KAF6207408.1) and *Lygus lineolaris* (GCXM01016614.1) suggests that the LygheETH precursor is a true transcript rather than an assembly artifact.

Two EH paralogs are present in many insect genomes, including several hemipteran species such as *Plautia stali*, *Acyrtosiphon pisum*, *N. lugens*, *D. citri*, and *B. tabaci* (Huybrechts et al., 2010; Tanaka et al., 2014; Wang et al., 2018; Li et al., 2020). Likewise, two EH precursor encoding transcripts were identified in the *L. hesperus* head dataset (Supplemental Fig. 1). The transcripts share 51% nt identity, while the encoded peptide precursors have 34% sequence identity. Both EH1 and EH2 are predicted to contain signal peptides and the characteristic six Cys motif (Fig. 5D). Phylogenetic analyses (Fig. 5E) support annotation of the two *L. hesperus* peptides, with LygheEH2 likely derived from the more ancient gene (Veenstra, 2014). Although no functional roles have been attributed to the second EH, differential expression of the two genes has been reported in another hemipteran (Wang et al., 2018).

Bursicon is a heterodimeric Cys knot peptide composed of two subunits – bursicon α and bursicon β . A bursicon β subunit (KX584428) with

the typical eleven conserved Cys residues (Supplemental Fig. 1) was identified in the initial peptidome and subsequently PCR amplified (Christie et al., 2017). A transcript encoding a full-length bursicon α precursor with the conserved Cys motif was identified in the head specific dataset (Supplemental Fig. 1). A C-terminally truncated variant with an incomplete Cys motif was also identified in the expanded transcriptomic data (Tassone et al., 2016); however, that sequence was embedded in the 3'UTR of a transcript annotated as phosphatidylinositol 4,5-bisphosphate 5-phosphatase A (JAQ15578.1). The chimeric nature of the transcript coupled with the absence of the truncated variant in *A. lucorum* and *L. lineolaris* genomic/transcriptomic datasets suggests it is likely an assembly artifact.

OK peptides have recently been linked to ecdysis in *R. prolixus* (Wulff et al., 2017; 2018). An OK type A precursor (Supplemental Fig. 1) was identified previously (KX584426; Christie et al., 2017). In the expanded dataset, we identified transcripts encoding three additional OK-like precursors; however, none of the transcripts contained complete ORFs. Two are N-terminal fragments, whereas the third is a C-terminal fragment (Supplemental Fig. 1). The first 90 nt of the two N-terminal fragments are identical to those in the LygheOK-A precursor; however, the sequence then diverges from the A-type variant with the two N-terminal fragments differentiated from each other by a 60 nt insertion. Given the similarities with the *R. prolixus* OK-B and OK-C precursors, which are differentiated by the inclusion of a small exon in OK-C (Wulff et al., 2017), we have termed the two *L. hesperus* N-terminal fragments LygheOK-B (the shorter fragment) and LygheOK-C (the fragment with the 60 nt insertion). The C-terminal fragment, which shares 40% amino acid identity with LygheOK-B and LygheOK-C, has been designated

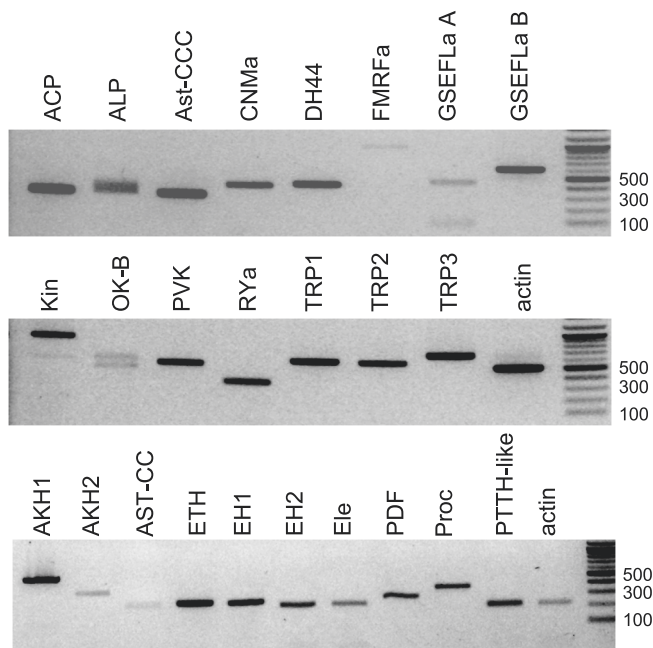


Fig. 4. RT-PCR based amplification of select peptide precursor-encoding transcripts from mixed sex *L. hesperus* heads. All amplimers were subcloned and sequence verified. A 500-bp fragment of *L. hesperus* actin was amplified as a positive control. Shown are the negative images of representative 2% agarose gels stained with SYBR safe. The bottom gel image consists of peptide precursor transcripts identified in the head specific dataset. Abbreviations are based on Yeoh et al. (2017).

(Fig. 5F) with a probable pyroglutamate residue occurring immediately after the signal peptide, a decapeptide (ACP) or undecapeptide (corazonin) with a C-terminal amidation signal, and a C-terminal precursor peptide with two conserved Cys (Supplemental Fig. 1)

3.3.2. Peptides that impact insect gut physiology

The interplay of numerous peptides (e.g. proctolin, kinins, sulfakinin, FMRFa-like peptides, diuretic hormones, periviscerokinins, RYamides, and ITP/ITPL) affect diverse aspects of insect gut physiology including myotropic activity, satiety, and osmotic balance (Audsley and Weaver, 2009; El Asrar et al., 2020). Transcripts encoding precursors for many of these peptides, including a number with atypical nonconsensus sequences, were identified in the new datasets.

Proctolin is a potent myotropic peptide that regulates gut motility (El Asrar et al., 2020). The *L. hesperus* transcript encodes a typical proctolin precursor, with the consensus pentapeptide sequence (RYLPT) bounded by a signal peptide and a dibasic cleavage site as well as a C-terminal precursor peptide (Supplemental Fig. 1).

Kinins (also referred to as leucokinin or myokinin) are pleiotropic peptides typically characterized by a C-terminal pentapeptide Fx₁x₂WG-amide (x₁ corresponds to Ser, Phe, His, Asn, or Tyr and x₂ is Ser, Pro, or Ala) that impact multiple aspects of gut physiology (Sangha et al., 2020; El Asrar et al., 2020). The Lyghekinin precursor encodes nine kinin or kinin-like peptides (Supplemental Fig. 1), which is less than that reported for *B. tabaci* (Li et al., 2020) and *R. prolixus* (Bhatt et al., 2014), but more than for *A. pisum* (Huybrechts et al., 2010) or *N. lugens* (Tanaka et al., 2014). Eight of the Lyghekinins have a typical C-terminal motif (7x FNSWG-amide; 1x FYAWG-amide) and one unique peptide, FQSWG-amide (Fig. 6A), which is also encoded by the *A. lucorum* kinin precursor (KAF6199581.1). This specific motif appears to be poorly distributed throughout the Insecta as the only other FQSWG-amide peptides listed in the DInER insect neuropeptide database were limited to a proturan species (Yeoh et al., 2017). An atypical FSxWA-amide kinin previously suggested to be distributed throughout the Heteroptera (Lavore et al.,

2018), is not present in the *L. hesperus* precursor.

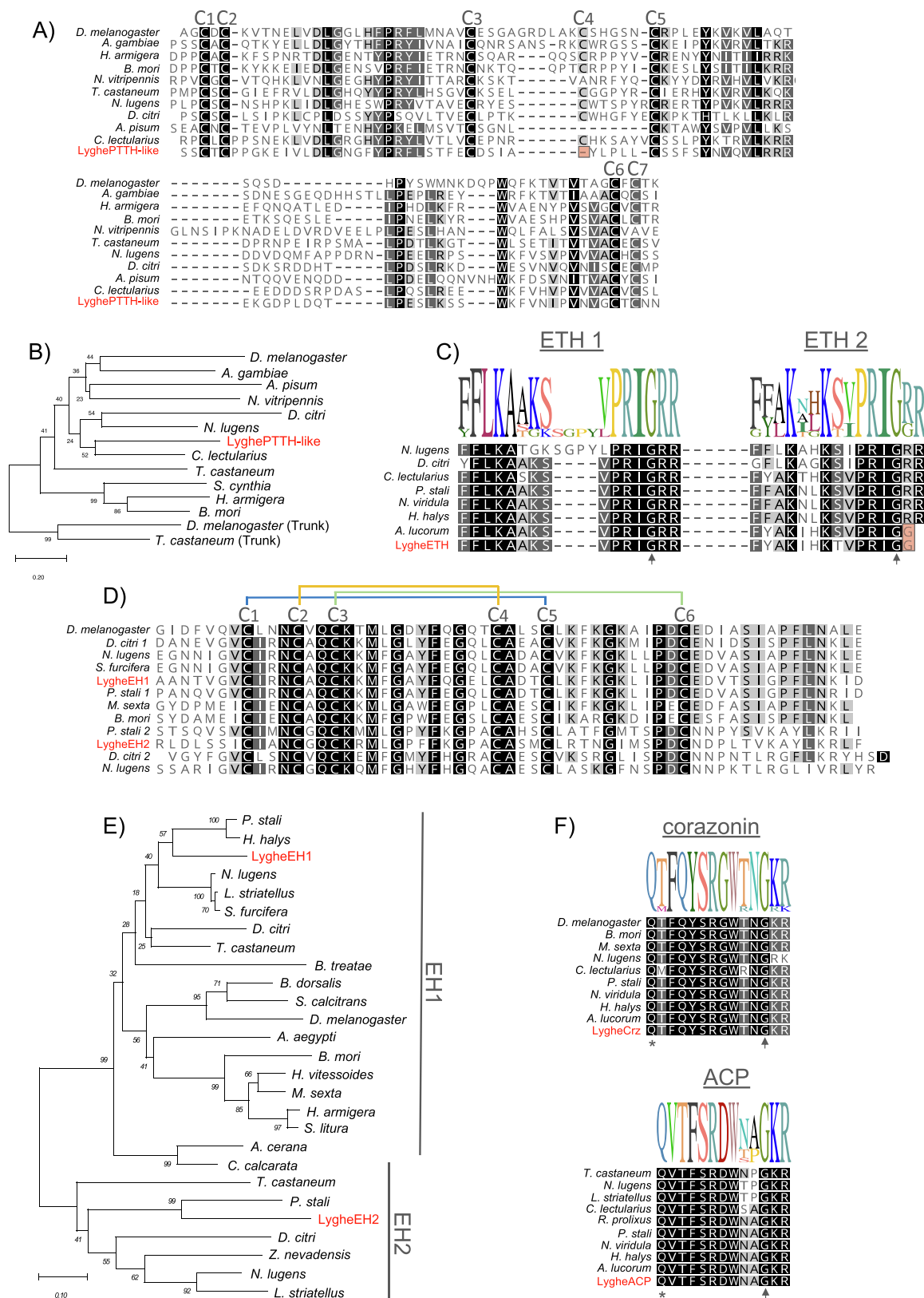
RYamides, a group of recently identified peptides typically characterized by a C-terminal FFxxxRY-amide, have been postulated to play a role in fluid balance in *Drosophila melanogaster* (Collin et al., 2011). While most RYamide precursors yield two RYamides, the LygheRYamide transcript encodes only a single FYMGSRY-amide peptide (Supplemental Fig. 1). The *L. hesperus* precursor also has an unusual non-amidated peptide, SSIFWTGSRYN, that differs from the reported consensus sequence at position 2 with a Ser rather than Gly (Fig. 6B). This difference is found in other heteropterans and the degree of phylogenetic conservation has been postulated as an indicator that the non-amidated peptide serves a physiological role (Lavore et al., 2018).

Sulfakinins (SK) function as satiety signals that inhibit insect feeding (Audsley and Weaver, 2009; Nagata, 2016). The SK precursor typically encodes two peptides with a conserved heptapeptide core (i.e. D/YEGHMRF-amide) that contains a sulfated Tyr. The *L. hesperus* precursor is predicted to yield two SKs (Supplemental Fig. 1). LygheSK1 has a pyroglutamate N terminus and the typical DYGHMRF-amide C terminus. The presence of an N-terminal pyroglutamate in SK1 has been previously confirmed by mass spectrometry for *R. prolixus* (Ons et al., 2011) and *Cimex lecturicus* (Predel et al., 2017). In contrast, LygheSK2 is characterized by a unique DYGYMRF-amide motif that appears to be typical of heteropterans (Fig. 6C). Although both Tyr sites in SK2 can potentially be sulfated, only the N-terminal Tyr has been confirmed to be sulfated in *C. lecturicus* (Predel et al., 2017).

The role of the diuretic hormone 44 (DH44) family of peptides, also referred to as corticotropin releasing-factor (CRF)-related diuretic hormones, in regulating osmotic balance through fluid excretion/diuresis is well established (Schooley et al., 2012). Typical of DH44 precursors, the LygheDH44 sequence has a C-terminal amidation signal and is embedded between precursor related peptides downstream of a predicted signal peptide (Supplemental Fig. 1). Comparison of the *L. hesperus* peptide with other DH44s (Fig. 6D) reveals sequence divergence at Ser8 (typically an aliphatic residue), Ser16 (typically a charged residue), and Asn29 (typically a charged residue). The role these residues may have in peptide biofunction remain to be determined.

The FMRFa-like peptides (FaLPs), which predominantly affect visceral muscle contractility (Audsley and Weaver, 2009), are characterized by a C-terminal FxRf-amide (x = Met, Ile, or Leu) sequence with multiple peptides encoded in a single precursor transcript. The presence in the *L. hesperus* precursor (Supplemental Fig. 1) of only four FaLP, all with unique C-terminal ends, appears to be uncommon among hemipterans, which typically contain multiple copies of a FaLP motif (Fig. 6E). The four *L. hesperus* FaLPs include two uncommon variations of the FaLP C terminus - FMYamide and IMRFamide. Although two sequence variants (MT210021 and MT2100212) for the LygheFaLP precursor were cloned, the significant nucleotide sequence identity (98.5%) between the two suggests the differences are the result of allelic variation rather than gene duplication.

Periviscerokinins (PVKs), initially defined as cardioacceleratory factors (e.g. cardioacceleratory peptide 2b, CAP2b) based on their bioactivity in Lepidoptera (Tublitz and Truman, 1985a; 1985b), were subsequently shown to also affect Malpighian tubule ion transport (Kean et al., 2002). PVKs are encoded by the *capability* (*capa*) gene, which typically generates a precursor transcript that contains a pair of 9–10 amino acid PVK peptides characterized by a PRV-amide C terminus and a single tryptopyrokinin peptide with a C-terminal WFGPRL-amide. Proteolytic processing of the LyghePVK precursor is predicted to yield one classic PVK, a typical tryptopyrokinin peptide, and a unique 11 amino acid PVK-like peptide with a PRS-amide C-terminus (Supplemental Fig. 1). Although also present in *A. lucorum* (KAF6216557.1), this latter peptide appears to be unique among hemipteran PVKs with the C-terminal Ser position most frequently occupied by an aliphatic amino acid (Fig. 6F). Similar variation in peptide length has been observed in number of hemipteran PVK peptides (Ahn and Choi, 2018), and the encoding gene has undergone duplication in some species such



(caption on next page)

Fig. 5. *L. hesperus* peptides associated with ecdysis and development. A) Alignment of PTTH peptide sequences. Alignment was performed with predicted signal peptides excluded. The seven conserved Cys residues potentially involved in intra- and intermolecular disulphide bonds are indicated. The Lgyus PTTH-like peptide lacks Cys4 (red box). B) Neighbor joining tree of PTTH phylogenetic relationships. The optimal tree with the sum of branch length = 7.6 is shown. Bootstrap values (1000 replicates) are shown next to the branches. The tree was rooted using the PTTH-related Trunk sequences from *D. melanogaster* and *T. castaneum*. C) Alignment of mature ETH peptide sequences. The *L. hesperus* and *A. lucorum* ETH2 sequences lack typical C-terminal Gly amidation signals (red box). D) Alignment of EH sequences. Alignment was performed with predicted signal peptides excluded. The six conserved Cys residues are indicated with the Cys-Cys pairing shown according to the biochemically determined pattern for *M. sexta*. E) Neighbor joining tree of EH phylogenetic relationships. The optimal tree with the sum of branch length = 5.2 is shown. Bootstrap values (1000 replicates) are shown next to the branches. The LygheEH2 sequence sorted to the more ancestral EH clade. F) Alignment of mature corazonin and ACP peptides. In all alignments, potential amidation sites are indicated by arrowheads and potential pyroglutamate residues are indicated by asterisks. Accession numbers of proteins used in the phylogenetic analyses are listed in Supplementary Table 3 and those used in the alignments are listed in Supplementary Table 5.

as *R. prolixus* (Paluzzi et al., 2008; Paluzzi and Orchard, 2010). A PVK-like transcript (GBRD01017626.1) was also identified in the datasets mined for the previous *L. hesperus* peptidome; however, that precursor lacks the predicted signal peptide suggesting potential transcript misassembly.

Although the antidiuretic activity of ITP on locust hindgut has been well-documented (Coast et al., 2002), the role of ITPL in fluid reabsorption is less clear (Wang et al., 2000). ITP/ITPLs are characterized by six conserved Cys and are 72 (ITP) to 79 amino acids (ITPL) in length. The two peptides are generated via alternative splicing of the ITP gene with an internal exon that encodes a stop codon retained in ITPL, but which is spliced out in ITP (Dirksen, 2009). This splicing profile is conserved in the *L. hesperus* transcripts as the exon encoding the ITP C-terminus is present in the ITPL 3'UTR (Fig S2). As indicated earlier, alternative splicing of the LygheITP/ITPL transcripts also impacts the 5' UTR and potentially results in the addition of a signal anchor upstream of the typical signal peptide in ITPL v1 (Fig. 2). Unlike many insect ITPs, the putative LygheITP lacks a typical C-terminal amidation signal (Fig. 6G). In contrast, the two LygheITPL variants have the characteristic -IKQ/HLHGAE/Dxxx C-terminus (Yu et al., 2016).

3.3.3. Other peptides with known function

Adipokinetic hormones (AKH) are 8–10 amino acid peptides that regulate energy metabolism (Nässel and Zandawala, 2019). Peptides in the AKH family are typically characterized by an N-terminal pyroglutamate, C-terminal amidation, and an aromatic acid at positions four and eight with the bioactive peptide immediately downstream of the signal peptide and a potential intermolecular disulfide bond in the C-terminal precursor related peptide. Two LygheAKH precursor transcripts were identified in our datasets (Supplementary Fig. 1). The two transcripts have 94% nucleotide identity (Fig. 3A) and 79% amino acid identity, suggesting gene duplication rather than allelic variation. Although the presence of multiple AKH encoding transcripts is not without precedence (Roller et al., 2008; Kaufmann et al., 2009; Veenstra, 2019), it seems to be limited among hemipterans with no evidence for duplication in transcriptomic data for *N. lugens* (Tanaka et al., 2014), *B. tabaci* (Li et al., 2020), *N. viridula* (Lavore et al., 2018), or *D. citri* (Wang et al., 2018). Multiple sequence alignment of LygheAKH1 and LygheAKH2 with other hemipteran AKH revealed divergence in aromatic acid usage at position four with both *L. hesperus* peptides incorporating Tyr rather than the more frequent Phe (Fig. 7A). A search of the DiNER insect neuropeptide database (Yeoh et al., 2017) underscored the uniqueness of this substitution among the 50 hemipteran AKH sequences deposited.

Pigment-dispersing factor (PDF) regulates circadian motor rhythmicity (Shafer and Yao, 2014). Among insects, PDFs are C-terminally amidated octadecapeptides characterized by a consensus sequence of NSE_{x1}INSL_{x2}LPK_{x3}x₄NDA-amide (x₁ corresponds to Leu or Ile; x₂ is Gly, Ser, or Ala; x₃ is Asn, Val, Leu, Ser, or Thr; and x₄ is Leu or Met). The LyghePDF precursor consists of a putative signal peptide, a precursor related peptide, and then the presumably bioactive peptide (Supplemental Fig. 1), which is well-conserved with other insect PDFs (Fig. 7B).

Elevenin, a recently identified peptide conserved across arthropods, has been linked with cuticular melanization (Uchiyama et al., 2018;

Wang et al., 2019). Similar to most species (Fig. 7C), the *L. hesperus* elevenin peptide is immediately downstream of the signal peptide and is characterized by two conserved Cys residues and a C-terminal RG motif that lacks amidation (Supplemental Fig. 1).

Members of the allatostatin C (AST-C) peptide family, which includes the AST-CC and AST-CCC paralogs (Veenstra, 2016b), were initially identified based on regulation of juvenile hormone biosynthesis. More recent studies, however, suggest broader physiological roles for the peptides (Urlacher et al., 2016; Bachtel et al., 2018; Villalobos-Sambucaro et al., 2016). AST-C, also referred to as PISC-AST to reflect the conserved C-terminal segment, is characterized by a pyroglutamate N terminus, the absence of C-terminal amidation, and a highly conserved CYFNPISC motif. AST-CC has the most variable sequence among the three paralogs, but can be characterized by the absence of post-translational modification at either terminus and a general C-terminal sequence of CYFNAVS/TC. In contrast, AST-CCC has a predicted C-terminal amidation site and a general sequence of KQCAFNAVSCF-amide. As noted by Lavore et al. (2018), many hemipteran peptides initially annotated as AST-C (e.g. *N. lugens*, *D. citri*, *A. pisum*, *T. infestans*) exhibit characteristics that are more consistent with AST-CCC. Similarly, the AST-C identified in our previous peptidome (i.e. GBHO01019349) aligns better with AST-CC peptides than AST-C (Fig. 7D). Here, our re-evaluation of the *L. hesperus* peptidome allowed identification of an AST-CCC precursor encoding transcript (Supplemental Fig. 1), the predicted peptide of which clearly aligns with AST-CCC annotated sequences (Fig. 7D). Similar to other hemipterans (Tanaka et al., 2014; Lavore et al., 2018; Huybrechts et al., 2010; Ons et al., 2011; Traverso et al., 2016; Predel et al., 2017; Li et al., 2020), no AST-C precursor was identified in our transcriptomic datasets. This absence may reflect loss of the AST-C gene, which has been reported for a number of insect lineages (Veenstra, 2016b).

3.3.4. Peptides with unknown function

The recently identified CNMamides (Jung et al., 2014), which have yet to be functionally characterized, have a highly conserved peptide core in hemipterans consisting of YM_{x1}LCHF_{x2}KICNM-amide (x₁ corresponds to Ser, Ala, or Thr). This core is likewise conserved in the predicted LygheCNMamide (Fig. 8A; Supplemental Fig. 1), a transcript (GBHO01023260.1) was also identified in the datasets mined for the previous *L. hesperus* peptidome.

Neuropeptide-like precursor 1 (NPLP1), which was identified via mass spectrometry-based analyses of the *D. melanogaster* CNS, is predicted to encode a number of putative peptides with no conserved C-terminal motif (Baggerman et al., 2005). Three potential NPLP1-encoding transcripts (NPLP1 a-c) were identified in the *L. hesperus* datasets with two encompassing complete ORFs and the third a C-terminal fragment (Supplemental Fig. 1). The two full-length LygheNPLP1 transcripts are differentiated by a 15-nucleotide deletion in LygheNPLP1b. The shorter NPLP1c appears to be a product of alternative splicing with a different exon used for the final 98 nucleotides including the presumptive stop codon (Fig. 8B). Processing of the LygheNPLP1a and LygheNPLP1b precursors is predicted to generate twenty non-amidated peptides and four amidated peptides. Despite limited conservation of predicted peptide C-terminal ends, hemipteran

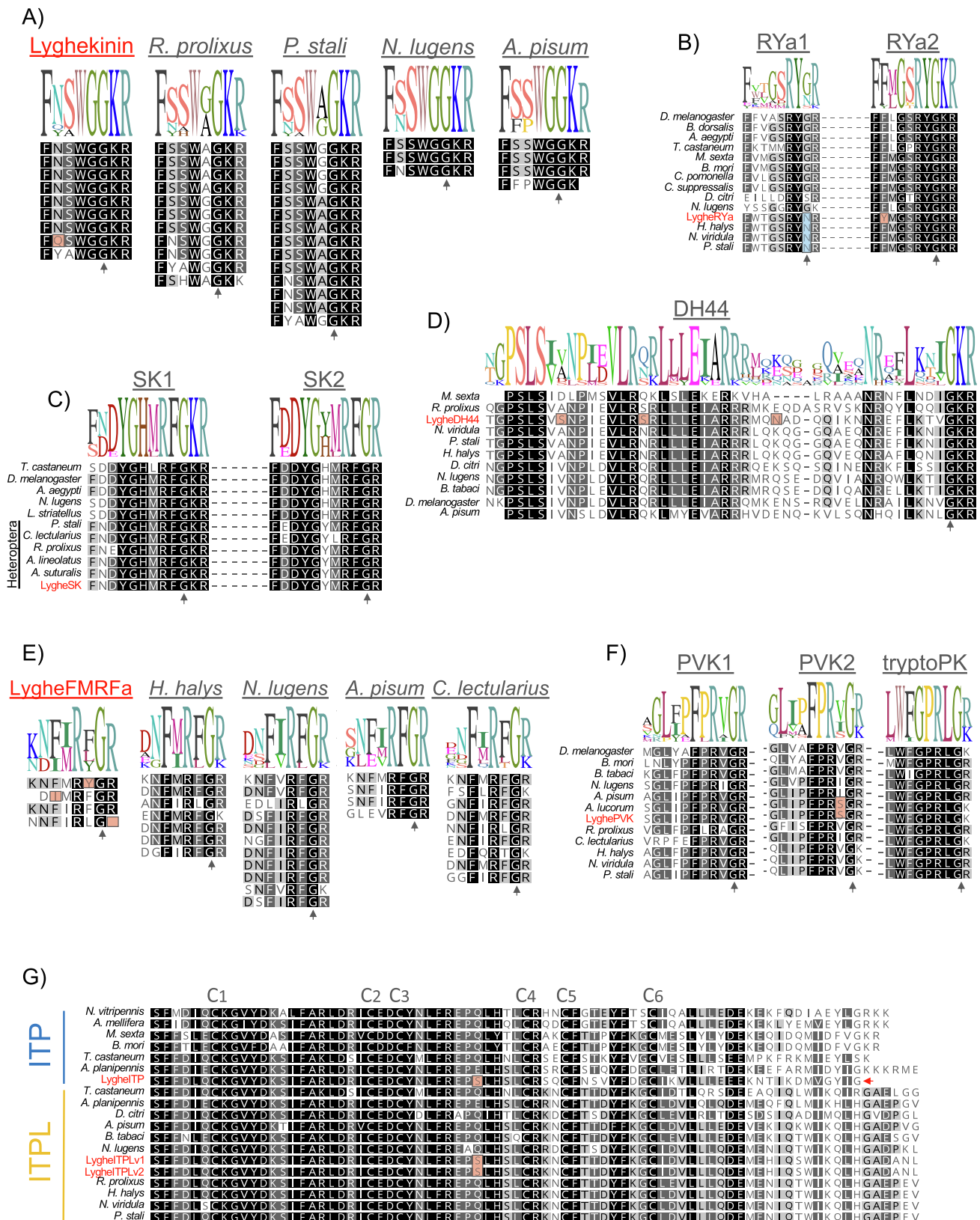


Fig. 6. *L. hesperus* peptides associated with gut physiology. A) Species specific alignment of kinins predicted for five hemipteran species. While highly represented among the other hemipterans, the FSSWA/G-amide motif is not encoded by the Lyghekinin transcript. B) Alignment of DH44 peptides. Alignment was performed with predicted signal peptides excluded. C) Alignment of mature RYamide peptides. Positions lacking a C-terminal amidation signal are indicated by a blue box. D) Alignment of FMRFa peptides predicted in the prepropeptide of five hemipteran species. E) Alignment of peptides encoded by the PVK/CAPA precursor. F) Alignment of peptides encoded by the SK precursor. G) Alignment of ITP and ITPL peptides. In all alignments, atypical amino acid usage is indicated by red boxes, potential amidation sites are indicated by arrowheads, and potential pyroglutamate residues are indicated by asterisks. The red arrow in G6 indicates absence of a C-terminal amidation signal. Accession numbers of the protein sequences used are listed in Supplementary Table 5.

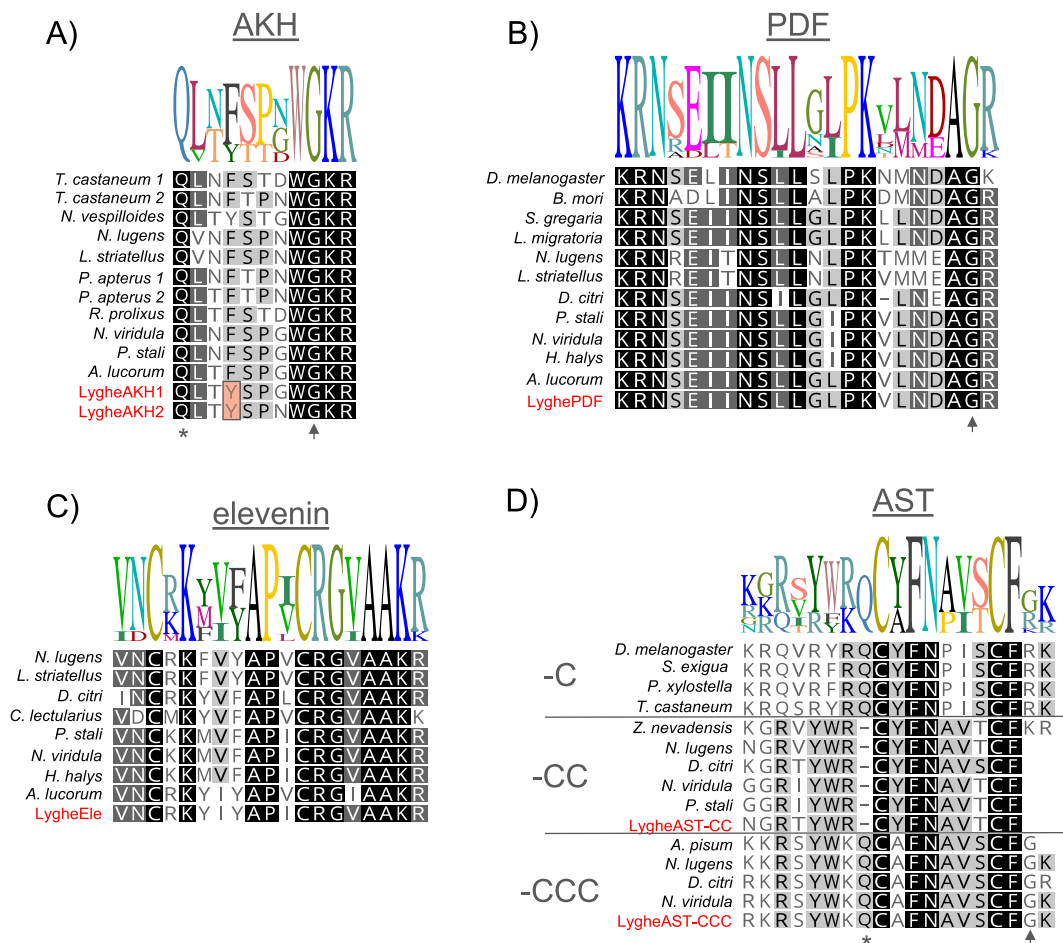


Fig. 7. *L. hesperus* peptides associated with unique biological roles. A) Alignment of AKH peptides. B) Alignment of PDF peptides. C) Alignment of elevenin-like peptides. D) Alignment of peptides from the extended AST-C peptide family. In all alignments, atypical amino acid usage is indicated by red boxes, potential amidation sites are indicated by arrowheads, and potential pyroglutamate residues are indicated by asterisks. Accession numbers of the protein sequences used are listed in Supplementary Table 5.

NPLP1 share significant sequence identity (Fig. 8C). Veenstra (2016a, b) noted the presence of an N-terminal consensus sequence, LARxGxLP (x = any amino acid), which is likewise well-represented in hemipteran NPLP1 precursors. Representation of the three other *D. melanogaster* NPLPs (i.e. NPLP2-4) in insect lineages is narrower, with an NPLP2 homolog reported in *A. mellifera* (Hummon et al., 2006) and an NPLP3 in *D. citri* (Wang et al., 2018). Consistent with this limited penetrance, no homologs were identified in any of the *L. hesperus* transcriptomic datasets.

Similar to NPLP1, agatoxin-like peptides (ALPs) were initially identified by mass spectrometry (Sturm et al., 2016), but have since been found in a number of transcriptomic and genomic datasets (Liessem et al., 2018; Badisco et al., 2011; Zhu et al., 2017; Benoit et al., 2019; Sparks et al., 2020; Veenstra, 2016a). ALPs, which share structural similarities with the agatoxin class of peptide-based spider toxins (Skinner et al., 1989; Bindokas and Adams, 1989), are characterized by a ~40 amino acid ALP region that contains a C-terminal amidation signal and eight highly conserved Cys residues. The spacing of these Cys comprise two sequence motifs, a Cx₆Cx₆CC (x = any six amino acids) principal structural motif and a Cx₆Cx₆Cx₆ extra structural motif, commonly found in spider toxins (Kozlov and Grishin, 2005). Through a combination of data mining and cDNA cloning, we identified multiple splice variants of the *L. hesperus* ALP precursor (Supplemental Fig. 1). Such variability is consistent with previously reported alternative splicing of the ALP gene (Sturm et al., 2016; Veenstra, 2016a). The splicing occurs upstream of the ALP encoding portion of the precursors,

resulting in transcripts differentiated by 66-nt, 39-nt, or 27-nt. The splicing has no effect on the peptides themselves with both Cys structural motifs maintained in all variants (Fig. 8D).

3.4. *LygheALP* transcript expression profiling

To date, no biological function has been attributed to insect ALPs. Thus, to provide potential insights into function, we used RT-PCR to examine transcript expression across development and in various adult body segments and tissues. *LygheALP* expression was, with the exception of the embryonic stage, constitutively expressed throughout development (Fig. 9A) and in all adult segments and tissues examined except ovaries (Fig. 9B). Although tissue-specific splicing is suggested, more quantitative methods such as real time PCR and/or targeted RNA-sequencing are needed to more fully explore this possibility.

Structural features shared with spider toxins and expression in hymenopteran venom glands (Bouazid et al., 2014; Torres et al., 2014; Liu et al., 2015) are consistent with a toxin-like function for insect ALPs. However, direct detection of ALPs in the neuroendocrine systems of insects from diverse orders (Sturm et al., 2016) and the presence of ALP transcripts in non-venomous species (Badisco et al., 2011; Liessem et al., 2018) suggests the physiological role of ALPs extends beyond toxins. This alternative use also seems to be the case for *L. hesperus*, in which the *LygheALP* transcript is expressed during nymph and adult development as well as throughout most of the adult body (Fig. 9A,B). The toxin-associated activities of ALPs, which affect neurotransmitter release via

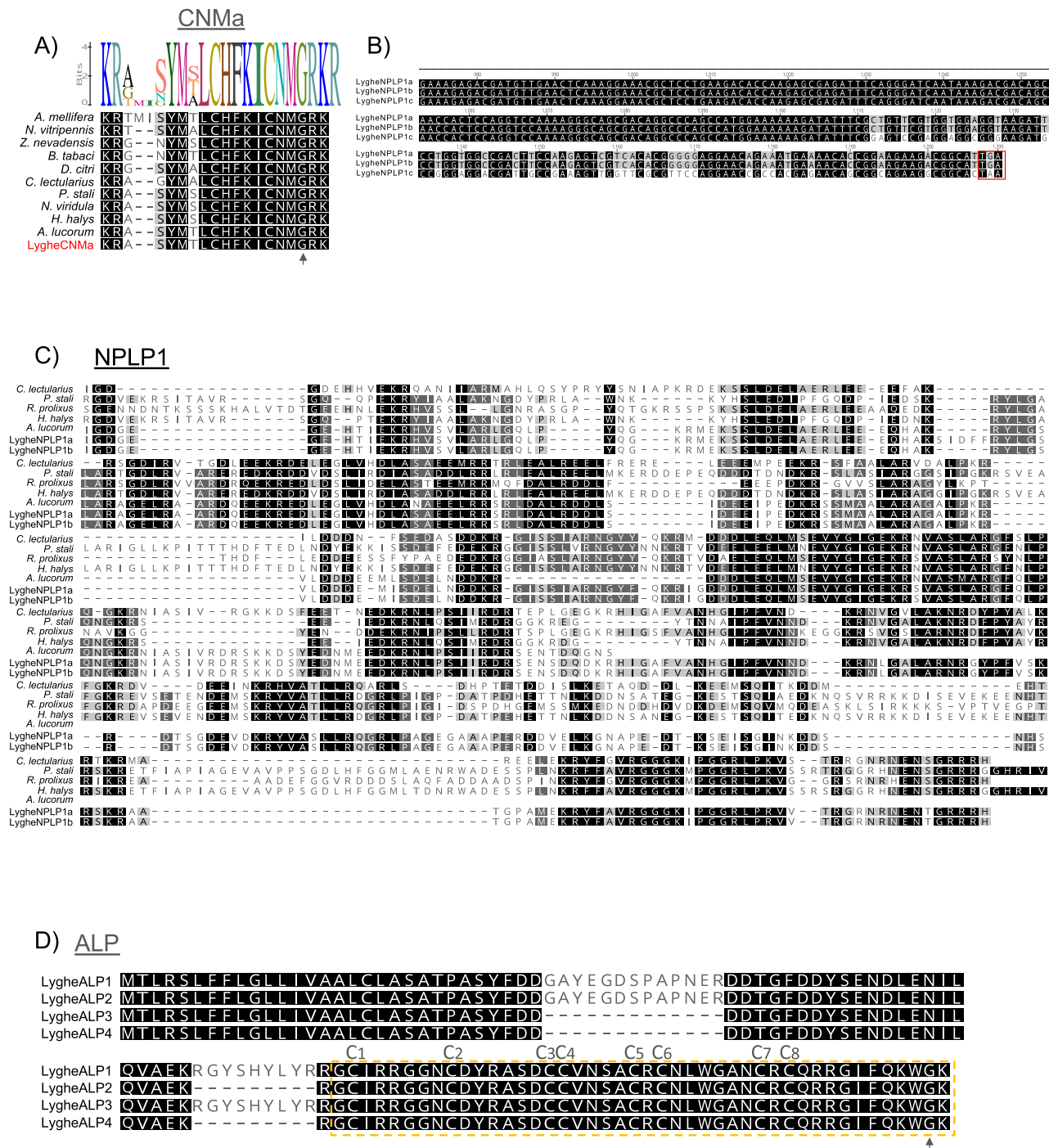


Fig. 8. *L. hesperus* peptides with unknown function. A) Alignment of CNMa peptides. B) Nucleotide sequence alignment of LygheNPLP1 transcript variants. Because sequence divergence in the variants is restricted to the 3' end, only nucleotides 973–1209 (i.e. stop codon) are shown. The positions of the predicted stop codons are indicated by a red box. C) Alignment of NPLP1 peptides. Alignment was performed with predicted signal peptides excluded. D) Alignment of the four LygheALP precursor peptides. Location of the eight Cys residues highly conserved in ALP across species are indicated. The position of the ALP peptide in the prepropeptide is demarcated by a yellow dashed box. In all alignments, potential amidation sites are indicated by arrowheads. Accession numbers of the protein sequences used are listed in Supplementary Table 5.

modification of various receptor-activated and/or voltage-gated channels (Adams, 2004), could be derived from an ancestral ALP functionality associated with ion channel modulation (Mobli et al., 2017). Ubiquitously expressed cysteine-rich peptides have been reported to function in pathogen-defense (Undheim et al., 2016; Daly and Wilson, 2018) and regulation of the melanization process (Tsukamoto et al., 1992; Shi et al., 2006). The absence of ALP transcripts in *L. hesperus* ovaries and embryonic tissue could indicate a potential transportation

repressor role for the peptides with potential to impede oogenesis and developmental processes.

4. Conclusions

In silico-based methods of mining transcriptomic datasets have proven to be invaluable for mass identification and characterization of neuropeptides from non-model organisms, such as *L. hesperus*. The

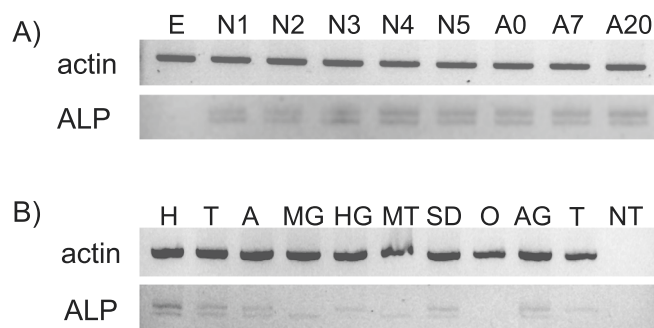


Fig. 9. LygheALP transcript expression profile. (A) Endpoint RT-PCR amplification of LygheALP transcripts across *L. hesperus* development. Abbreviations: egg, E; 1st-5th instar nymphs are N1-N5, and adults aged 0, 7, and 20 days post-eclosion are A0, A7, and A20. (B) Endpoint RT-PCR amplification of LygheALP transcripts in various adult body segments and tissues. Abbreviations - H, head; T, thorax; A, abdomen; MG, midgut; HG, hindgut; MT, Malpighian tubule; SD, seminal depository; O, ovary; AG, accessory gland; T, testes; NT, no template. Shown are the negative images of representative 3% agarose gels stained with SYBR safe. The presence of multiple ALP amplicons relative to that seen in Fig. 4 reflects the resolving power of the higher percentage agarose gel for differentiating the ALP splice variants.

peptides predicted here significantly expand current resources and further foundational information that can guide and inform future integrative approaches that have the potential to enhance our understanding of how neuropeptides impact critical physiological processes. This knowledge can then be applied to uncovering unique targets for pest control. In addition, the confirmation of expression and transcriptional profiling of ALPs in a non-venomous pest further implicate these poorly characterized, cysteine-rich peptides in non-toxin associated activities and open new avenues of research to characterize their *in vivo* biological function.

Funding

This work was supported by funds from the National Science Foundation (IOS-1856307; AEC), the Cades Foundation (AEC), and base CRIS funding from the US Department of Agriculture (2020-22620-022-00D; JJH and CSB).

CRediT authorship contribution statement

J. Joe Hull: Conceptualization, Investigation, Writing - original draft, Writing - review & editing. **Roni J. Gross:** Investigation. **Colin S. Brent:** Conceptualization, Resources, Writing - review & editing. **Andrew E. Christie:** Conceptualization, Writing - original draft.

Acknowledgements

Lisa Baldwin is thanked for reading and editing an earlier version of this article. Mention of trade names or commercial products in this article is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the US Department of Agriculture; the USDA is an equal opportunity provider and employer.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ygcn.2020.113708>.

References

Adams, M.E., 2004. Agatoxins: ion channel specific toxins from the American funnel web spider, *Agelenopsis aperta*. *Toxicon* 43, 509–525.

- Ahn, S.-J., Choi, M.-Y., 2018. Identification and characterization of capa and pyrokinin genes in the brown marmorated stink bug, *Halyomorpha halys* (Hemiptera): Gene structure, immunocytochemistry, and differential expression. *Arch. Insect Biochem. Physiol.* 99, e21500.
- Altstein, M., 2004. Novel insect control agents based on neuropeptide antagonists: The PK/PBAN family as a case study. *J. Mol. Neurosci.* 22, 147–157.
- Altstein, M., Nüssel, D.R., 2010. Neuropeptide signaling in insects. *Adv. Exp. Med. Biol.* 692, 155–165.
- Audsley, N., Down, R.E., 2015. G protein coupled receptors as targets for next generation pesticides. *Insect Biochem. Mol. Biol.* 67, 27–37.
- Audsley, N., Weaver, R.J., 2009. Neuropeptides associated with the regulation of feeding in insects. *Gen. Comp. Endocrinol.* 162, 93–104.
- Bachtel, N.D., Hovsepian, G.A., Nixon, D.F., Eleftherianos, I., 2018. Allatostatin C modulates nociception and immunity in *Drosophila*. *Sci. Rep.* 8, 7501–7511.
- Badisco, L., Huybrechts, J., Simonet, G., Verlinden, H., Marchal, E., Huybrechts, R., Schoofs, L., De Loof, A., Vanden Broeck, J., 2011. Transcriptome analysis of the desert locust central nervous system: production and annotation of a *Schistocerca gregaria* EST database. *PLoS ONE* 6, e17274.
- Baggerman, G., Boonen, K., Verleyen, P., De Loof, A., Schoofs, L., 2005. Peptidomic analysis of the larval *Drosophila melanogaster* central nervous system by two-dimensional capillary liquid chromatography quadrupole time-of-flight mass spectrometry. *J. Mass Spectrom.* 40, 250–260.
- Bao, C., Yang, Y., Huang, H., Ye, H., 2015. Neuropeptides in the cerebral ganglia of the mud crab, *Scylla paramamosain*: transcriptomic analysis and expression profiles during vitellogenesis. *Sci. Rep.* 5, 17055.
- Barberá, M., Martínez-Torres, D., 2017. Identification of the *prothoracicotrophic hormone* (*Pth*) coding gene and localization of its site of expression in the pea aphid *Acyrtosiphon pisum*. *Insect Mol. Biol.* 26, 654–664.
- Bendtsen, J.D., Nielsen, H., von Heijne, G., Brunak, S., 2004. Improved prediction of signal peptides: SignalP 3.0. *J. Mol. Biol.* 340, 783–795.
- Benoit, J.B., Adelman, Z.N., Reinhardt, K., Dolan, A., Poelchau, M., Jennings, E.C., Szuter, E.M., Hagan, R.W., Gujjar, H., Shukla, J.N., Zhu, F., Mohan, M., Nelson, D.R., Rosendale, A.J., Derst, C., et al., 2019. Unique features of a global human ectoparasite identified through sequencing of the bed bug genome. *Nat. Commun.* 7, 10165.
- Bindokas, V.P., Adams, M.E., 1989. Omega-Aga-I: a presynaptic calcium channel antagonist from venom of the funnel web spider, *Agelenopsis aperta*. *J. Neurobiol.* 20, 171–188.
- Bhatt, G., da Silva, R., Nachman, R.J., Orchard, I., 2014. The molecular characterization of the kinin transcript and the physiological effects of kinins in the blood-gorging insect, *Rhodnius prolixus*. *Peptides* 53, 148–158.
- Brent, C.S., 2010. Reproduction of the western tarnished plant bug, *Lygus hesperus*, in relation to age, gonadal activity and mating status. *J. Insect Physiol.* 56, 28–34.
- Brent, C.S., Hull, J.J., 2014. Characterization of male-derived factors inhibiting female sexual receptivity in *Lygus hesperus*. *J. Insect Physiol.* 60, 104–110.
- Bouzid, W., Verdenaud, M., Klopp, C., Ducancel, F., Noirot, C., Vétillard, A., 2014. De novo sequencing and transcriptome analysis for *Tetramorium bicarinatum*: a comprehensive venom gland transcriptome analysis from an ant species. *BMC Genomics* 15, 987.
- Chen, J., Choi, M.S., Mizoguchi, A., Veenstra, J.A., Kang, K., Kim, Y.J., Kwon, J.Y., 2015. Isoform-specific expression of the neuropeptide orckinin in *Drosophila melanogaster*. *Peptides* 68, 50–57.
- Christie, A.E., 2008a. Neuropeptide discovery in Ixodoidea: an *in silico* investigation using publicly accessible expressed sequence tags. *Gen. Comp. Endocrinol.* 157, 174–185.
- Christie, A.E., 2008b. *In silico* analyses of peptide paracrine/hormones in Aphidoidea. *Gen. Comp. Endocrinol.* 159, 67–79.
- Christie, A.E., 2011. Crustacean neuroendocrine systems and their signaling agents. *Cell Tissue Res.* 345, 41–67.
- Christie, A.E., 2015. *In silico* prediction of a neuropeptidome for the eusocial insect *Mastotermes darwiniensis*. *Gen. Comp. Endocrinol.* 224, 69–83.
- Christie, A.E., 2016. Prediction of *Scylla olivacea* (Crustacea; Brachyura) peptide hormones using publicly accessible transcriptome shotgun assembly (TSA) sequences. *Gen. Comp. Endocrinol.* 230–231, 1–16.
- Christie, A.E., Chi, M., 2015. Neuropeptide discovery in the Araneae (Arthropoda, Chelicerata, Arachnida): elucidation of true spider peptidomes using that of the Western black widow as a reference. *Gen. Comp. Endocrinol.* 213, 90–109.
- Christie, A.E., Hull, J.J., Richer, J.A., Geib, S.M., Tassone, E.E., 2017. Prediction of a peptidome for the western tarnished plant bug *Lygus hesperus*. *Gen. Comp. Endocrinol.* 243, 22–38.
- Christie, A.E., Stemmler, E.A., Dickinson, P.S., 2010. Crustacean neuropeptides. *Cell. Mol. Life Sci.* 67, 4135–4169.
- Coast, G.M., Orchard, I., Phillips, J.E., Schooley, D.A., 2002. Insect diuretic and antidiuretic hormones. *Adv. Insect Physiol.* 29, 279–409.
- Collin, C., Hauser, F., Krogh-Meyer, P., Hansen, K.K., de Valdivia, E.G., Williamson, M., Grimmelikhuijzen, C.J.P., 2011. Identification of the *Drosophila* and *Tribolium* receptors for the recently discovered insect RYamide neuropeptides. *Biochem. Biophys. Res. Commun.* 412, 578–583.
- Daly, N.L., Wilson, D., 2018. Structural diversity of arthropod venom toxins. *Toxicon* 152, 46–56.
- de Oliveira, A.L., Calcino, A., Wanninger, A., 2019. Ancient origins of arthropod moulting pathway components. *eLife* 8, 489.
- Dhara, A., Eum, J.-H., Robertson, A., Gulia-Nuss, M., Vogel, K.J., Clark, K.D., Graf, R., Brown, M.R., Strand, M.R., 2013. Ovary ecdysteroidogenic hormone functions independently of the insulin receptor in the yellow fever mosquito, *Aedes aegypti*. *Insect Biochem. Mol. Biol.* 43, 1100–1108.

- Dirksen, H., 2009. Insect ion transport peptides are derived from alternatively spliced genes and differentially expressed in the central and peripheral nervous system. *J. Exp. Biol.* 212, 401–412.
- Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucl. Acids Res.* 32, 1792–1797.
- El Asrar, R.A., Cools, D., Vanden Broeck, J., 2020. Role of peptide hormones in insect gut physiology. *Curr. Opin. Insect Sci.* 41, 71–78.
- Ferré, F., Clote, P., 2005. DiANNA: a web server for disulfide connectivity prediction. *Nucl. Acids Res.* 33, W230–W232.
- Hansen, K.K., Stafflinger, E., Schneider, M., Hauser, F., Cazzamali, G., Williamson, M., Kollmann, M., Schachtner, J., Grimmelikhuijzen, C.J.P., 2010. Discovery of a novel insect neuropeptide signaling system closely related to the insect adipokinetic hormone and corazonin hormonal systems. *J. Biol. Chem.* 285, 10736–10747.
- Hauser, F., Neupert, S., Williamson, M., Predel, R., Tanaka, Y., Grimmelikhuijzen, C.J., 2010. Genomics and peptidomics of neuropeptides and protein hormones present in the parasitic wasp *Nasonia vitripennis*. *J. Proteome Res.* 9, 5296–5310.
- Hou, Q.-L., Jiang, H.-B., Gui, S.-H., Chen, E.-H., Wei, D.-D., Li, H.-M., Wang, J.-J., Smaghe, G., 2017. A role of corazonin receptor in larval-pupal transition and pupariation in the Oriental fruit fly *Bactrocera dorsalis* (Hendel) (Diptera: Tephritidae). *Front. Physiol.* 8, 460.
- Hull, J.J., Chaney, K., Geib, S.M., Fabrick, J.A., Brent, C.S., Walsh, D., Lavine, L.C., 2014. Transcriptome-based identification of ABC transporters in the western tarnished plant bug *Lygus hesperus*. *PLoS One* 9, e113046.
- Hummon, A.B., Richmond, T.A., Verleyen, P., Baggerman, G., Huybrechts, J., Ewing, M. A., Vierstraete, E., Rodriguez-Zas, S.L., Schoofs, L., Robinson, G.E., Sweedler, J.V., 2006. From the genome to the proteome: uncovering peptides in the Apis brain. *Science* 314, 647–649.
- Huybrechts, J., Bonhomme, J., Minoli, S., Prunier-Leterme, N., Dombrovsky, A., Abdelatif, M., Robichon, A., Veenstra, J.A., Tagu, D., 2010. Neuropeptide and neurohormone precursors in the pea aphid, *Acyrtosiphon pisum*. *Insect Mol. Biol.* 19, 87–95.
- Isaac, R.E., Taylor, C.A., Hamasaka, Y., Nässel, D.R., Shirras, A.D., 2004. Proctolin in the post-genomic era: new insights and challenges. *Invert. Neurosci.* 5, 51–64.
- Jiang, H., Kim, H.G., Park, Y., 2015. Alternatively spliced orckinin isoforms and their functions in *Tribolium castaneum*. *Insect Biochem. Mol. Biol.* 65, 1–9.
- Jung, S.-H., Lee, J.-H., Chae, H.-S., Seong, J.Y., Park, Y., Park, Z.Y., Kim, Y.-J., 2014. Identification of a novel insect neuropeptide, CNMa and its receptor. *FEBS Lett* 588, 2037–2041.
- Kastin, A.J., 2006. Handbook of Biologically Active Peptides, first ed. Academic Press.
- Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780.
- Kaufmann, C., Merzendorfer, H., Gäde, G., 2009. The adipokinetic hormone system in Culicidae (Diptera: Culicidae): molecular identification and characterization of two adipokinetic hormone (AKH) precursors from *Aedes aegypti* and *Culex pipiens* and two putative AKH receptor variants from *A. aegypti*. *Insect Biochem. Mol. Biol.* 39, 770–781.
- Kean, L., Cazenave, W., Costes, L., Broderick, K.E., Graham, S., Pollock, V.P., Davies, S.-A., Veenstra, J.A., Dow, J.A.T., 2002. Two nitridergic peptides are encoded by the gene *capability* in *Drosophila melanogaster*. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* 282, R1297–R1307.
- Kim, Y.-J., Spalovská-Valachová, I., Cho, K.-H., Zitnanova, I., Park, Y., Adams, M.E., Zitnan, D., 2004. Corazonin receptor signaling in ecdysis initiation. *Proc. Natl. Acad. Sci.* 101, 6704–6709.
- Konopińska, D., Rosiński, G., 1999. Proctolin, an insect neuropeptide. *J. Pept. Sci.* 5, 533–546.
- Kotwica-Rolinska, J., Kristofova, L., Chvalova, D., Pauchová, L., Provazník, J., Hejnkova, M., Sehadová, H., Lichý, M., Vaněková, H., Dolezel, D., 2020. Functional analysis and localisation of a thyrotropin-releasing hormone-type neuropeptide (EFLa) in hemipteran insects. *Insect Biochem. Mol. Biol.* 122, 103376.
- Kozlov, S., Grishin, E., 2005. Classification of spider neurotoxins using structural motifs by primary structure features. Single residue distribution analysis and pattern analysis techniques. *Toxicol. Anal. Chem.* 46, 672–686.
- Kumar, S., Stecher, G., Li, M., Niyaz, C., Tamura, K., 2018. MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. *Mol. Biol. Evol.* 35, 1547–1549.
- Lavore, A., Perez-Gianmarco, L., Esponda-Behrens, N., Palacio, V., Catalano, M.I., Rivera-Pomar, R., Ons, S., 2018. *Nezara viridula* (Hemiptera: Pentatomidae) transcriptomic analysis and neuropeptidomics. *Sci. Rep.* 8, 1–15.
- Li, A., Rinehart, J.P., Denlinger, D.L., 2009. Neuropeptide-like precursor 4 is uniquely expressed during pupal diapause in the flesh fly. *Peptides* 30, 518–521.
- Li, J.J., Shi, Y., Lin, G.L., Yang, C.H., Liu, T.-X., 2020. Genome-wide identification of neuropeptides and their receptor genes in *Bemisia tabaci* and their transcript accumulation change in response to temperature stresses. *Insect Sci.* 83, 409–412.
- Liessem, S., Ragionieri, L., Neupert, S., Büschges, A., Predel, R., 2018. Transcriptomic and neuropeptidomic analysis of the stick insect, *Carausius morosus*. *J. Proteome Res.* 17, 2192–2204.
- Liu, Z., Chen, S., Zhou, Y., Xie, C., Zhu, B., Zhu, H., Liu, S., Wang, W., Chen, H., Ji, Y., 2015. Deciphering the venom transcriptome of killer-wasp *Vespa velutina*. *Sci. Rep.* 5, 9454.
- Liutkeviciute, Z., Koehbach, J., Eder, T., Gil-Mansilla, E., Gruber, C.W., 2016. Global map of oxytocin/vasopressin-like neuropeptide signalling in insects. *Sci. Rep.* 6, 39177.
- Mesquita, R.D., Vionette-Amaral, R.J., Lowenberger, C., Rivera-Pomar, R., Monteiro, F. A., Minx, P., Spieth, J., Carvalho, A.B., Panzera, F., Lawson, D., Torres, A.Q., Ribeiro, J.M.C., Sorgine, M.H.F., Waterhouse, R.M., Montague, M.J., Abad-Franch, F., Alves-Bezerra, M., Amaral, L.R., Araujo, H.M., Araujo, R.N., Aravind, L., Atella, G.C., Azambuja, P., Berni, M., Bittencourt-Cunha, P.R., Braz, G.R.C., Calderón-Fernández, G., Carareto, C.M.A., Christensen, M.B., Costa, I.R., Costa, S.G., Dansa, M., Dumas-Filho, C.R.O., De Paula, I.F., Dias, F.A., Dimopoulos, G., Emrich, S.J., Esponda-Behrens, N., Fampa, P., Fernandez-Medina, R.D., da Fonseca, R.N., Fontenelle, M., Fronick, C., Fulton, L.A., Gandara, A.C., Garcia, E.S., Genta, F.A., Giraldo-Calderón, G.I., Gomes, B., Gondim, K.C., Granzotto, A., Guarneri, A.A., Guigó, R., Harry, M., Hughes, D.S.T., Jablonka, W., Jacquin-Joly, E., Juárez, M.P., Koerich, L.B., Lange, A.B., Latorre-Estivalis, J.M., Lavore, A., Lawrence, G.G., Lazoski, C., Lazzari, C.R., Lopes, R.R., Lorenzo, M.G., Lugon, M.D., Majerowicz, D., Marcet, P.L., Mariotti, M., Masuda, H., Megy, K., Melo, A.C.A., Missirlis, F., Mota, T., Noriega, F.G., Nouzova, M., Nunes, R.D., Oliveira, R.L.L., Oliveira-Silveira, G., Ons, S., Orchard, I., Pagola, L., Paiva-Silva, G.O., Pascual, A., Pavan, M.G., Pedrini, N., Peixoto, A.A., Pereira, M.H., Pike, A., Polycarpo, C., Prodocimi, F., Ribeiro-Rodrigues, R., Robertson, H.M., Salerno, A.P., Salmon, D., Santesmasses, D., Schama, R., Seabra-Junior, E.S., Silva-Cardoso, L., Silva-Neto, M.A. C., Souza-Gomes, M., Sterkel, M., Taracena, M.L., Tojo, M., Tu, Z.J., Tubio, J.M.C., Ursic-Bedoya, R., Venancio, T.M., Walter-Nuno, A.B., Wilson, D., Warren, W.C., Wilson, R.K., Huebner, E., Dotson, E.M., Oliveira, P.L., 2015. Genome of *Rhodnius prolixus*, an insect vector of Chagas disease, reveals unique adaptations to hematophagy and parasite infection. *Proc. Natl. Acad. Sci. U.S.A.* 112, 14936–14941.
- Mobli, M., Undheim, E.A.B., Rash, L.D., 2017. Modulation of ion channels by cysteine-rich peptides: from sequence to structure. *Adv. Pharmacol.* 79, 199–223.
- Monigatti, F., Gasteiger, E., Bairoch, A., Jung, E., 2002. The Sulfinator: predicting tyrosine sulfation sites in protein sequences. *Bioinformatics* 18, 769–770.
- Nachman, R.J., Pietrantonio, P.V., Coast, G.M., 2009. Toward the development of novel pest management agents based upon insect kinin neuropeptide analogues. *Ann. NY Acad. Sci.* 1163, 251–261.
- Nagata, S., 2016. Sulfakinin. In *Handbook of Hormones*. pp. 381–382. Academic Press.
- Naranjo, S.E., Ellsworth, P.C., Dierig, D., 2011. Impact of *Lygus* spp. (Hemiptera:Miridae) on damage, yield, and quality of lespedeza (*Physaria fendleri*), a potential new oil-seed crop. *J. Econ. Entomol.* 104, 1575–1583.
- Nässel, D.R., Zandawala, M., 2019. Recent advances in neuropeptide signaling in *Drosophila*, from genes to physiology and behavior. *Prog. Neurobiol.* 179, 101607.
- Neupert, S., Marciniak, P., Köhler, R., Nachman, R.J., Suh, C.-P.-C., Predel, R., 2018. Different processing of CAPA and pyrokinin precursors in the giant mealworm beetle *Zophobas atratus* (Tenebrionidae) and the boll weevil *Anthonomus grandis grandis* (Curculionidae). *Gen. Comp. Endocrinol.* 258, 53–59.
- Ons, S., Sterkel, M., Diambra, L., Urlaub, H., Rivera-Pomar, R., 2011. Neuropeptide precursor gene discovery in the Chagas disease vector *Rhodnius prolixus*. *Insect Mol. Biol.* 20, 29–44.
- Paluzzi, J.P., Orchard, I., 2010. A second gene encodes the anti-diuretic hormone in the insect, *Rhodnius prolixus*. *Mol. Cell. Endocrinol.* 317, 53–63.
- Paluzzi, J.P., Russell, W.K., Nachman, R.J., Orchard, I., 2008. Isolation, cloning, and expression mapping of a gene encoding an antidiuretic hormone and other CAPA-related peptides in the disease vector, *Rhodnius prolixus*. *Endocrinology* 149, 4638–4646.
- Predel, R., Neupert, S., Derst, C., Reinhardt, K., Wegener, C., 2017. Neuropeptidomics of the bed bug *Cimex lectularius*. *J. Proteome Res.* 17, 440–454.
- Ritter, R.A., Lensen, A., Blodgett, S., Taper, M.A., 2010. Regional assemblages of *Lygus* (Heteroptera: Miridae) in Montana canola fields. *J. Kansas Entomol. Soc.* 83, 297–305.
- Roller, L., Yamanaka, N., Watanabe, K., Daubnerová, I., Žitňan, D., Kataoka, H., Tanaka, Y., 2008. The unique evolution of neuropeptide genes in the silkworm *Bombyx mori*. *Insect Biochem. Mol. Biol.* 38, 1147–1157.
- Roller, L., Žitňanová, I., Dai, L., Šimo, L., Park, Y., Satake, H., Tanaka, Y., Adams, M.E., Žitňan, D., 2010. Ecdysis triggering hormone signaling in arthropods. *Peptides* 31, 429–441.
- Rommelaere, S., Boquete, J.-P., Piton, J., Kondo, S., Lemaitre, B., 2019. The exchangeable apolipoprotein Nlp2 sustains lipid flow and heat acclimation in *Drosophila*. *Cell Reports* 27, 886–899.
- Saitou, N., Nei, M., 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4, 406–425.
- Sangha, V., Lange, A.B., Orchard, I., 2020. Identification and cloning of the kinin receptor in the Chagas disease vector, *Rhodnius prolixus*. *Gen. Comp. Endocrinol.* 289, 113380.
- Scherkenbeck, J., Zdobinsky, T., 2009. Insect neuropeptides: structures, chemical modifications and potential for insect control. *Bioorg. Med. Chem.* 17, 4071–4084.
- Schooley, D.A., Horodyski, F.M., Coast, G.M., 2012. Hormones controlling homeostasis in insects. In: *Insect Endocrinology*. Academic Press, pp. 366–429.
- Scott, D.R., 1977. An annotated listing of host plants of *Lygus hesperus* Knight. *Bull. Entomol. Soc. Am.* 23, 19–22.
- Shafer, O.T., Yao, Z., 2014. Pigment-dispersing factor signaling and circadian rhythms in insect locomotor activity. *Curr. Opin. Insect Sci.* 1, 73–80.
- Shi, L., Li, B., Paskewitz, S.M., 2006. Cloning and characterization of a putative inhibitor of melanization from *Anopheles gambiae*. *Insect Mol. Biol.* 15, 313–320.
- Skinner, W.S., Adams, M.E., Quistad, G.B., KataokaCesarin, B.J., Enderlin, F.E., Schooley, D.A., 1989. Purification and characterization of two classes of neurotoxins from the funnel web spider, *Agelenopsis aperta*. *J. Biol. Chem.* 264, 2150–2155.
- Sparks, M.E., Bansal, R., Benoit, J.B., Blackburn, M.B., Chao, H., Chen, M., Cheng, S., Childers, C., Dinh, H., Doddapaneni, H.V., Dugan, S., Elpidina, E.N., Farrow, D.W., Friedrich, M., Gibbs, R.A., et al., 2020. Brown marmorated stink bug, *Halymorpha halys* (Stål), genome: putative underpinnings of polyphagy, insecticide resistance potential and biology of a top worldwide pest. *BMC Genomics* 21, 227.
- Southey, B.R., Sweedler, J.V., Rodriguez-Zas, S.L., 2008. Prediction of neuropeptide cleavage sites in insects. *Bioinformatics* 24, 815–825.

- Stecher, G., Tamura, K., Kumar, S., 2020. Molecular Evolutionary Genetics Analysis (MEGA) for macOS. *Mol. Biol. Evol.* 37, 1237–1239.
- Sterkel, M., Oliveira, P.L., Urlaub, H., Hernández-Martínez, S., Rivera-Pomar, R., Ons, S., 2012. OKB, a novel family of brain-gut neuropeptides from insects. *Insect Biochem. Mol. Biol.* 42, 466–473.
- Sturm, S., Ramesh, D., Brockmann, A., Neupert, S., Predel, R., 2016. Agatoxin-like peptides in the neuroendocrine system of the honey bee and other insects. *J. Proteomics* 132, 77–84.
- Tanaka, Y., Suetsugu, Y., Yamamoto, K., Noda, H., Shinoda, T., 2014. Transcriptome analysis of neuropeptides and G-protein coupled receptors (GPCRs) for neuropeptides in the brown planthopper *Nilaparvata lugens*. *Peptides* 53, 125–133.
- Tassone, E.E., Geib, S.M., Hall, B., Fabrick, J.A., Brent, C.S., Hull, J.J., 2016. *De novo* construction of an expanded transcriptome assembly for the western tarnished plant bug, *Lygus hesperus*. *Gigascience* 5, 6.
- Torres, A.F.C., Huang, C., Chong, C.-M., Leung, S.W., Prieto-da-Silva, A.R.B., Havt, A., Quinet, Y.P., Martins, A.M.C., Lee, S.M.Y., Rádis-Baptista, G., 2014. Transcriptome analysis in venom gland of the predatory giant ant *Dinoponera quadricaps*: insights into the polypeptide toxin arsenal of Hymenopterans. *PLoS ONE* 9, e87556.
- Traverso, L., Sierra, I., Sterkel, M., Francini, F., Ons, S., 2016. Neuropeptidomics in *Triatoma infestans*. Comparative transcriptomic analysis among triatomines. *J. Physiol. (Paris)* 110, 83–98.
- Tsukamoto, T., Ichimaru, Y., Kanegae, N., Watanabe, K., Yamaura, I., Katsura, Y., Funatsu, M., 1992. Identification and isolation of endogenous insect phenoloxidase inhibitors. *Biochem. Biophys. Res. Commun.* 184, 86–92.
- Tublitz, N.J., Truman, J.W., 1985a. Insect cardioactive peptides. I. Distribution and molecular characteristics of two cardioacceleratory peptides in the tobacco hawkmoth *Manduca sexta*. *J. Exp. Biol.* 114, 365–379.
- Tublitz, N.J., Truman, J.W., 1985b. Insect cardioactive peptides. II. Neurohormonal control of heart activity by two cardioacceleratory peptides in the tobacco hawkmoth *Manduca sexta*. *J. Exp. Biol.* 114, 381–395.
- Uchiyama, H., Maehara, S., Ohta, H., Seki, T., Tanaka, Y., 2018. Elevenin regulates the body color through a G protein-coupled receptor NLA42 in the brown planthopper *Nilaparvata lugens*. *Gen. Comp. Endocrinol.* 258, 33–38.
- Undheim, E.A.B., Mobli, M., King, G.F., 2016. Toxin structures as evolutionary tools: Using conserved 3D folds to study the evolution of rapidly evolving peptides. *Bioessays* 38, 539–548.
- Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B.C., Remm, M., Rozen, S. G., 2012. Primer3—new capabilities and interfaces. *Nucleic Acids Res.* 40, e115–e115.
- Urlacher, E., Soustelle, L., Parmentier, M.-L., Verlinden, H., Gherardi, M.-J., Fourmy, D., Mercer, A.R., Devaud, J.-M., Massou, I., 2016. Honey bee allatostatins target galanin/somatostatin-like receptors and modulate learning: a conserved function? *PLoS ONE* 11, e0146248.
- Van Hiel, M.B., Van Loy, T., Poels, J., Vandersmissen, H.P., Verlinden, H., Badisco, L., Vanden Broeck, J., 2010. Neuropeptide receptors as possible targets for development of insect pest control agents. *Adv. Exp. Med. Biol.* 692, 211–226.
- Veenstra, J.A., 2000. Mono- and dibasic proteolytic cleavage sites in insect neuroendocrine peptide precursors. *Arch. Insect Biochem. Physiol.* 43, 49–63.
- Veenstra, J.A., 2014. The contribution of the genomes of a termite and a locust to our understanding of insect neuropeptides and neurohormones. *Front. Physiol.* 5, 454.
- Veenstra, J.A., 2016a. Similarities between decapod and insect neuropeptidomes. *PeerJ* 4, e2043–e2139.
- Veenstra, J.A., 2016b. Allatostatins C, double C and triple C, the result of a local gene triplication in an ancestral arthropod. *Gen. Comp. Endocrinol.* 230–231, 153–157.
- Veenstra, J.A., 2019. Coleoptera genome and transcriptome sequences reveal numerous differences in neuropeptide signaling between species. *PeerJ* 7, e7144–e7239.
- Villalobos-Sambucaro, M.J., Diambra, L.A., Noriega, F.G., Ronderos, J.R., 2016. Allatostatin-C antagonizes the synergistic myostimulatory effect of allatotropin and serotonin in *Rhodnius prolixus* (Stal). *Gen. Comp. Endocrinol.* 233, 1–7.
- Wahedi, A., Paluzzi, J.-P., 2018. Molecular identification, transcript expression, and functional deorphanization of the adipokinetic hormone/corazonin-related peptide receptor in the disease vector, *Aedes aegypti*. *Sci. Rep.* 8, 1–13.
- Wang, S.L., Wang, W.W., Ma, Q., Shen, Z.F., Zhang, M.-Q., Zhou, N.M., Zhang, C.-X., 2019. Elevenin signaling modulates body color through the tyrosine-mediated cuticle melanism pathway. *FASEB J.* 33, 9731–9741.
- Wang, Y.J., Zhao, Y., Meredith, J., Phillips, J.E., Theilmann, D.A., Brock, H.W., 2000. Mutational analysis of the C-terminus in ion transport peptide (ITP) expressed in *Drosophila* Kc1 cells. *Arch. Insect Biochem.* 45, 129–138.
- Wang, Z., Zhou, W., Hameed, M., Liu, J., Zeng, X., 2018. Characterization and expression profiling of neuropeptides and G-protein-coupled receptors (GPCRs) for neuropeptides in the asian citrus psyllid, *Diaphorina citri* (Hemiptera: Psyllidae). *Int. J. Mol. Sci.* 19, 3912–3919.
- Wheeler, A.G., 2001. Biology of the Plant Bugs (Hemiptera: Miridae): Pests, Predators, Opportunists. Comstock Publishing Associates.
- Wulff, J.P., Sierra, I., Sterkel, M., Holtorf, M., Van Wielendaele, P., Francini, F., Broeck, J. V., Ons, S., 2017. Orcokinin neuropeptides regulate ecdysis in the hemimetabolous insect *Rhodnius prolixus*. *Insect Biochem. Mol. Biol.* 81, 91–102.
- Wulff, J.P., Capriotti, N., Ons, S., 2018. Orcokinins regulate the expression of neuropeptide precursor genes related to ecdysis in the hemimetabolous insect *Rhodnius prolixus*. *J. Insect Physiol.* 108, 31–39.
- Xie, Y., Zhang, L., Zhang, C., Wu, X., Deng, X., Yang, X., Tobe, S.S., 2015. Synthesis, biological activity, and conformational study of N-methylated allatostatin analogues inhibiting juvenile hormone biosynthesis. *J. Agric. Food Chem.* 63, 2870–2876.
- Yamanaka, N., Roller, L., Zitnan, D., Satake, H., Mizoguchi, A., Kataoka, H., Tanaka, Y., 2011. Bombyx orckinins are brain-gut peptides involved in the neuronal regulation of ecdysteroidogenesis. *J. Comp. Neurol.* 519, 238–246.
- Yeoh, J.G.C., Pandit, A.A., Zandawala, M., Nüssel, D.R., Davies, S.-A., Dow, J.A.T., 2017. DInER: Database for Insect Neuropeptide Research. *Insect Biochem. Mol. Biol.* 86, 9–19.
- Yu, B., Li, D.-T., Wang, S.-L., Xu, H.-J., Bao, Y.-Y., Zhang, C.-X., 2016. Ion transport peptide (ITP) regulates wing expansion and cuticle melanism in the brown planthopper, *Nilaparvata lugens*. *Insect Mol. Biol.* 25, 778–787.
- Zandawala, M., Haddad, A.S., Hamoudi, Z., Orchard, I., 2015. Identification and characterization of the adipokinetic hormone/corazonin-related peptide signaling system in *Rhodnius prolixus*. *FEBS J.* 282, 3603–3617.
- Zhang, C., Qu, Y., Wu, X., Song, D., Ling, Y., Yang, X., 2015a. Design, synthesis and aphicidal activity of N-terminal modified insect kinin analogs. *Peptides* 68, 233–238.
- Zhang, C., Qu, Y., Wu, X., Song, D., Ling, Y., Yang, X., 2015b. Eco-friendly insecticide discovery via peptidomimetics: design, synthesis, and aphicidal activity of novel insect kinin analogues. *J. Agric. Food Chem.* 63, 4527–4532.
- Zhu, J., Jiang, F., Wang, X., Yang, P., Bao, Y., Zhao, W., Wang, W., Lu, H., Wang, Q., Cui, N., Li, J., Chen, X., Luo, L., Yu, J., Kang, L., Cui, F., 2017. Genome sequence of the small brown planthopper, *Laodelphax striatellus*. *GigaScience* 6, 1–12.
- Žitňan, D., Kim, Y.-J., Žitňanová, I., Roller, L., Adams, M.E., 2007. Complex steroid-peptide-receptor cascade controls insect ecdysis. *Gen. Comp. Endocrinol.* 153, 88–96.
- Zuckerkandl, E., Pauling, L., 1965. Evolutionary divergence and convergence in proteins. Edited in *Evolving Genes and Proteins* by V. Bryson and H.J. Vogel, pp. 97–166. Academic Press, New York.