1 **Prediction of a peptidome for the Western tarnished plant bug**
2 *Lygus hesperus*

3
4
5 Andrew E. Christie^{1*}, J. Joe Hull², Josh A. Richer², Scott M. Geib³, and Erica E. Tassone⁴
6

7
8 ¹Békésy Laboratory of Neurobiology, Pacific Biosciences Research Center, School of Ocean and
9 Earth Science and Technology, University of Hawaii at Manoa, 1993 East-West Road, Honolulu,
10 Hawaii, USA 96822

11
12 ²Pest Management and Biocontrol Research Unit, US Arid Land Agricultural Research Center,
13 USDA Agricultural Research Services, Maricopa, Arizona, USA 85138

14
15 ³Tropical Crop and Commodity Protection Research Unit, Daniel K. Inouye Pacific Basin
16 Agricultural Research Center, USDA Agricultural Research Services, Hilo, Hawaii, USA 96720
17

18 ⁴Plant Physiology and Genetics Research Unit, US Arid Land Agricultural Research Center,
19 USDA Agricultural Research Services, Maricopa, Arizona, USA 85138
20

21 *Correspondence to: Dr. Andrew E. Christie, Békésy Laboratory of Neurobiology, Pacific
22 Biosciences Research Center, School of Ocean and Earth Science and Technology, University of
23 Hawaii at Manoa, 1993 East-West Road, Honolulu, HI 96822 USA. Phone: 808-956-5212;
24 FAX: 808-956-6984; Email: crabman@pbrc.hawaii.edu
25

26
27 **Key words:** transcriptomics; bioinformatics; neurohormone; neuropeptide
28

Abstract

Many strategies for controlling insect pests require an understanding of their hormonal signaling agents, peptides being the largest and most diverse single class of these molecules. *Lygus hesperus* is a pest species of particular concern, as it is responsible for significant damage to a wide variety of commercially important plant crops. At present, little is known about the peptide hormones of *L. hesperus*. Here, transcriptomic data were used to predict a peptidome for *L. hesperus*. Fifty-three *L. hesperus* transcripts encoding peptide precursors were identified, with a subset amplified by PCR for sequence verification. The proteins deduced from these transcripts allowed for the prediction of a 119-sequence peptidome for *L. hesperus*. The predicted peptides include isoforms of allatostatin A, allatostatin B (AST-B), allatostatin C, allatotropin, bursicon, CCHamide, corazonin, crustacean cardioactive peptide, crustacean hyperglycemic hormone/ion transport peptide, diuretic hormone 31, GSEFLamide, insulin-like peptide, myosuppressin, neuroparsin, neuropeptide F, orcokinin, orcomyotropin, pyrokinin, short neuropeptide F, SIFamide, sulfakinin and tachykinin-related peptide. Of note were several isoforms of AST-B that possess –WX₇Wamide carboxyl-termini rather than the stereotypical –WX₆Wamide (*e.g.*, KWQDMQNPGWamide), an allatotropin ending in –SARGFamide rather than –TARGFamide (GLKNGPLNSARGFamide), a GSEFLamide ending in –GTEFLamide (TVGTEFLamide), several orcokinins with PMDEIDR– rather than NFDEIDR– amino-termini (*e.g.*, PMDEIDRAGFTHFV), and an eight rather 12 amino acid long isoform of SIFamide (PPFNGSIFamide). Collectively, the *L. hesperus* peptidome predicted here provides a resource for initiating physiological investigations of peptidergic signaling in this species, including studies directed at the biological control of this agricultural pest.

1. Introduction

Peptides are the largest and most diverse single class of molecules used for chemical communication (*e.g.*, Kastin, 2006). In insects, peptidergic signaling systems are key components in the control of essentially all aspects of physiology and behavior, and as such, are of growing interest for the development of biological control strategies for many pest species (*e.g.*, 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). Insecticidal effects have been demonstrated via the knockdown of several peptidergic systems, as well as through the application of native and/or modified peptide analogs, and targets for biological control include, among others, the peptidergic pathways involved in reproduction, development, growth and feeding (for a recent review see Audsley and Down [2015]).

Members of the genus *Lygus*, a complex of 28 morphologically similar polyphagous mirid plant bugs (Schwartz and Footitt, 1998; Wheeler, 2001; Schuh, 2013), cause significant damage via their cell-rupture (also referred to as “lacerate-and-flush”) method of feeding (Backus et al., 2005; Backus et al., 2007). In this strategy, *Lygus* species inject a mixture of proteolytic enzymes into plant vegetative and reproductive structures via specialized stylets (Shackel et al., 2005; Backus et al., 2007). Crop losses, which are mainly attributed to pre-reproductive adults and *Lygus* nymphs (Cooper and Spurgeon, 2013), arise from a combination of saliva-induced cell rupturing and plant wound responses that result in deformation of developing fruit, reduced vegetative growth, and feeding site necrosis (Strong, 1970). In the western United States, the predominant *Lygus* species is *Lygus hesperus* Knight (commonly referred to as the Western tarnished plant bug), which has an extensive host plant range (>150 species) including many economically important food, fiber, and seed crops (Scott, 1997;

Wheeler, 2001; Ritter et al., 2010; Naranjo et al., 2011). Despite its significance as an agricultural pest, essentially nothing is known about the native peptide hormones of *Lygus hesperus*.

While many methods can be used to determine arthropod peptidomes, *in silico* transcriptome/genome mining provides a rapid means for large-scale peptide discovery when appropriate molecular resources are available for a species (*e.g.*, Dircksen et al., 2011; Hauser et al., 2010; Hummon et al., 2006; Toullec et al., 2013; Ventura et al., 2014; Weaver and Audsley, 2008; Yan et al., 2012). A common strategy for this type of analysis is one based on homology to known peptide precursors, with the known protein being used to probe the target genomic or transcriptomic dataset for genes/transcripts encoding putative homologs (*e.g.* Christie, 2008a, 2008b, 2014a, 2014b, 2014c, 2014d, 2014e, 2014f, 2015a, 2015b, 2015c, 2015d, 2016a, 2016b; Christie and Chi, 2015a, 2015b, 2015c; Christie et al., 2008a, 2010a, 2011a, 2011b, 2011c; 2013, 2015; Christie and Pascual, 2016; Gard et al., 2009; Ma et al., 2009, 2010). Once identified, the gene or transcript of interest can be translated, and the deduced pre/preprohormone subjected to a simple bioinformatics workflow that utilizes freeware and homology to known arthropod precursor processing schemes to predict the structures of the mature bioactive peptides that are likely generated from it (*e.g.* Christie, 2008a, 2008b, 2014a, 2014b, 2014c, 2014d, 2014e, 2014f, 2015a, 2015b, 2015c, 2015d, 2016; Christie and Chi, 2015a, 2015b, 2015c; Christie et al., 2008a, 2010a, 2011a, 2011b, 2011c, 2013, 2015; Christie and Pascual, 2016; Gard et al., 2009; Ma et al., 2009, 2010).

The development of a transcriptome shotgun assembly (TSA) for *L. hesperus* (BioProject Nos. **PRJNA210219** and **PRJNA238835** [Hull et al., 2014]) provides a resource for peptide discovery in this species, and here these data were used for such an investigation. As the

results that follow will show, numerous peptide-encoding transcripts were identified within the *L. hesperus* TSA dataset, allowing for the prediction of over 100 distinct peptide hormones/modulators for this species. The identified peptides include many involved in the control of reproduction, development, growth and feeding in other insects, and it is likely that they play similar roles in *L. hesperus* as well. Given their putative functional roles, the identified transcripts and peptides provide a resource for initiating physiological investigations of peptidergic signaling in *L. hesperus*, including studies directed at exploring their potential use as targets for biorational-based control of this important agricultural pest.

2. Materials and methods

2.1. Database searches

Database searches were conducted on or before March 22, 2016, using methods modified from a well-vetted protocol (*e.g.* Christie, 2008a, 2008b, 2014a, 2014b, 2014c, 2014d, 2014e, 2014f, 2015a, 2015b, 2015c, 2015d, 2016a, 2016b; Christie and Chi, 2015a, 2015b, 2015c; Christie et al., 2008a, 2010a, 2011a, 2011b, 2013, 2015; Christie and Pascual, 2016; Gard et al., 2009; Ma et al., 2009, 2010). Specifically, the database of the online program tblastn (National Center for Biotechnology Information, Bethesda, MD; <http://blast.ncbi.nlm.nih.gov/Blast.cgi>) was set to “Transcriptome Shotgun Assembly (TSA)” and restricted to data from *Lygus hesperus* “(taxid:30085)”. Known insect or crustacean peptide precursors were input into tblastn as query sequences, and all hits returned by a given search were fully translated using the “Translate” tool in ExPASy (<http://web.expasy.org/translate/>) and then checked manually for homology to the target query. The complete list of peptide families searched for in this study, as well as the specific queries used, is provided in Table 1; this table also provides the BLAST-generated

maximum score and E-value for each of the transcripts identified as encoding a putative neuropeptide precursor.

2.2. Peptide prediction

The structures of mature peptides were predicted using a well-established workflow (*e.g.* Christie, 2008a, 2008b, 2014a, 2014b, 2014c, 2014d, 2014e, 2014f, 2015a, 2015b, 2015c, 2015d, 2016a, 2016b; Christie and Chi, 2015a, 2015b, 2015c; Christie et al., 2008a, 2010a, 2011a, 2011b, 2011c, 2013, 2015; Christie and Pascual, 2016; Gard et al., 2009; Ma et al., 2009, 2010). Specifically, each of the deduced precursor proteins was assessed for the presence of a signal peptide using the online program SignalP 4.1 (<http://www.cbs.dtu.dk/services/SignalP/>; Petersen et al., 2011); the D-cutoff values of SignalP 4.0 were set to “Sensitive” to better match the sensitivity of version 3.0 of this freeware program. Prohormone cleavage sites were identified based on the information presented in Veenstra (2000) and/or by homology to known arthropod pre/preprohormone processing schemes. When present, prediction of the sulfation state of tyrosine residues was conducted 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 and/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 amino (N)-terminal glutamine/glutamic acid residues and carboxyl (C)-terminal amidation at glycine residues, were predicted by homology to known arthropod peptide isoforms. Figure 1 shows two examples of mature peptide structural prediction using the workflow just described; the mature structures of all peptides predicted in this study are provided in Table 2. All protein/peptide alignments were

done using the online program MAFFT version 7 (<http://mafft.cbrc.jp/alignment/software/>; Katoh and Standley, 2013).

2.3. Peptide-encoding transcript cloning

To help validate the *in silico* peptide predictions, a subset of the putative peptide-encoding transcripts were cloned and sequenced. Oligonucleotide primers (Table 3) designed to amplify the putative open reading frames (*i.e.*, the portion of the transcript beginning with the start codon and ending with the stop codon) were designed based on the nucleotide sequences of the transcripts of interest using Primer3 (Untergasser et al., 2012). Total RNA was isolated from 7-9 day old mixed sex adult *L. hesperus* heads (obtained from an in-house-reared colony [US Arid Land Agricultural Research Center, USDA Agricultural Research Services, Maricopa, AZ]) using TRI Reagent Solution (Life Technologies/Ambion, Carlsbad, CA) according to the manufacturer's instructions. RNA quality and quantity were determined spectrophotometrically using the Take3 module on a Synergy H4 Hybrid Multi-Mode Microplate Reader (Biotek Instruments, Winooski, VT). To remove any residual genomic DNA, a 1 µg aliquot of RNA was treated with DNase I (New England Biolabs, Ipswich, MA), and then 500 ng of the DNA-free RNA was used as a template for cDNA synthesis using a SuperScript III First-Strand Synthesis System (Life Technologies) with custom-made random pentadecamers (Integrated DNA Technologies, San Diego, CA). Peptide transcripts were amplified in a 20-µL reaction volume using SapphireAmp Fast PCR Master Mix (Clontech Laboratories Inc., Mountain View, CA) with 12.5 ng cDNA and 0.2 µM sense and antisense primers on a Biometra TRIO (Göttingen, Germany) thermocycler. PCR conditions consisted of an initial denaturation at 95°C for 2 min followed by 40 cycles of 95°C for 20 sec, 53°C for 20 sec, and 72°C for 60 sec, and concluded at

72°C for 5 min. 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 at the Arizona State University DNA Core laboratory (Tempe, AZ). The resulting sequences have been deposited with NCBI GenBank under accession numbers **KX584418-KX584424** and **KX584426-KX584428**.

3. Results

3.1. In silico transcriptome mining and bioinformatics peptide prediction

The *L. hesperus* transcriptomic data publicly deposited in GenBank were searched for transcripts encoding 33 distinct peptide families (Table 1). In the interest of space, only those searches that resulted in the identification of putative precursor-encoding transcripts are described here, with the data presented in alphabetical order based on family name. All proteins listed as “full-length” exhibit a functional signal sequence (including a “start” methionine) and are flanked on their C-terminal end by a stop codon. Proteins described here as “partial” lacked a start methionine (referred to as C-terminal partial proteins), a stop codon (referred to as N-terminal partial proteins), or both of these features (referred to as internal fragment proteins).

3.1.1. Allatostatin A (AST-A)

Three transcripts were identified as encoding putative AST-A precursors (Table 1). Translation of these sequences revealed each to encode an identical 216 amino acid full-length preprohormone (Lyghe-prepro-AST-A; Fig. 2A). Fifteen peptides were predicted from Lyghe-prepro-AST-A (Table 2 and Fig. 2A), seven of which possess the C-terminal motif –

YXFGLamide (or a close approximation thereof), the hallmark of the AST-A family (Christie et al., 2010b). Of the AST-As, five are present in Lyghe-prepro-AST-A as single copies, with two copies of LYSFGLamide contained within this precursor. All eight of the linker/precursor-related peptides derived from Lyghe-prepro-AST-A possess distinct structures.

3.1.2. Allatostatin B (AST-B)

Five transcripts were identified as encoding putative AST-B precursors (Table 1). Translation of these sequences revealed each to encode an identical 357 amino acid full-length preprohormone (Lyghe-prepro-AST-B; Fig. 2B). Nineteen distinct peptides were predicted from Lyghe-prepro-AST-B (Fig. 2B and Table 2), four of which possess the C-terminal motif – WX₆Wamide, the structural hallmark of the AST-B family (Christie et al., 2010b). An additional eight peptides possess –WX₇Wamide C-termini and are considered here to be variant AST-Bs.

3.1.3. Allatostatin C (AST-C)

A single transcript was identified as encoding a putative AST-C precursor (Table 1). Translation of this sequence revealed a 119 amino acid full-length preprohormone (Lyghe-prepro-AST-C; Fig. 1A). Three distinct peptides were predicted from Lyghe-prepro-AST-C (Table 2 and Fig. 1A), including one, GQKNGRTYWRCYFNAVTCF (disulfide bridging between the two cysteine residues), that shows structural similarity to members of the AST-C family, *i.e.*, it has an unamidated C-terminus similar to the standard –PISCF motif and is predicted to possess a disulfide bridge between the cysteine residues located at positions 2 and 9 from its C-terminus (Christie et al., 2010b).

3.1.4. Allatotropin

A single transcript was identified as encoding a putative allatotropin precursor (Table 1). Translation of this sequence revealed a 98 amino acid full-length preprohormone (Lyghe-prepro-allatotropin; Fig. 2C). Three distinct peptides were predicted from Lyghe-prepro-allatotropin (Table 2 and Fig. 2C), one of which, GLKNGPLNSARGFamide, appears to be a variant allatotropin, *i.e.*, its C-terminus (–SARGFamide) is essentially identical to –TARGFamide, which is considered the structural hallmark of members of the allatotropin family (Christie et al., 2010b).

3.1.5. Bursicon β

A single transcript was identified as encoding a bursicon β precursor (Table 1). Translation of this sequence revealed a 114 amino acid C-terminal partial protein (Lyghe-pre-bursicon β ; Fig. 2D). Based on homology to known peptide sequences, this protein likely represents the C-terminus of an isoform of bursicon β (Table 2); a BLAST search of the non-redundant arthropod proteins curated in GenBank identified partner of bursicon (a synonym of bursicon β) isoform X1 from the bed bug *Cimex lectularius* (**Accession No. XP_014256948**; unpublished, i5k Genome Sequencing Initiative) as the most similar protein to the *L. hesperus* partial peptide.

3.1.6. CCHamide

Two transcripts were identified as encoding putative CCHamide precursors (Table 1). Translation of one sequence revealed a 129 amino acid full-length precursor (Lyghe-prepro-CCHamide Ia; Fig. 2E1), while translation of the other revealed a 103 amino acid full-length

protein (Lyghe-prepro-CCHamide Ia; Fig. 2E2). With the exception of a 26 amino acid insertion/deletion, Lyghe-prepro-CCHamide Ia and Ib are identical in sequence. Five distinct proteins were predicted from Lyghe-prepro-CCHamide Ia (Table 2 and Fig. 2E1), including GCSAFGHSCFGGHamide (disulfide bridging between the two cysteine residues), which possess the structural motif XCXXW/Y/FGXXCXGXHamide (disulfide bridging between the position 2 and 9 cysteine residues), which is the hallmark of members of the CCHamide family (Hansen et al., 2011). Four peptides were predicted from Lyghe-prepro-CCHamide Ib (Table 2 and Fig. 2E2), three of which (all linker/precursor-related sequences) are shared with Lyghe-prepro-CCHamide Ia. The CCHamide putatively liberated from Lyghe-prepro-CCHamide Ib is identical to that derived from Lyghe-prepro-CCHamide Ia, with the exception that it possesses an additional glycine residue at its N-terminus (Table 2).

3.1.7. *Corazonin*

A single transcript was identified as encoding a putative corazonin precursor (Table 1). Translation of this sequence revealed a 96 amino acid full-length prohormone (Lyghe-prepro-corazonin; Fig. 2F). Two peptides were predicted from Lyghe-prepro-corazonin (Table 2 and Fig. 2F), one of which, pQTFQYSRGWTNamide, is a known and broadly conserved arthropod isoform of corazonin (Christie et al., 2010b).

3.1.8. *Crustacean cardioactive peptide (CCAP)*

Three transcripts were identified as encoding putative CCAP precursors (Table 1). Translation of these transcripts revealed each to encode a full-length protein, with one prohormone being 107 amino acids long (Lyghe-prepro-CCAP Ia; Fig. 2G1), another 155

amino acids (Lyghe-prepro-CCAP Ib; Fig. 2G2), and the final protein 130 amino acids in length (Lyghe-prepro-CCAP Ic; Fig. 2G3). The three precursors are identical in sequence over much of their lengths, differing only in their C-termini. Four peptides were predicted from Lyghe-prepro-CCAP Ia and Ib, with five predicted from Lyghe-prepro-CCAP Ic (Table 2 and Figs. 2G1-2G3). All three preprohormones share one copy of PFCNAFTGCamide (a disulfide bridge between the two cysteine residues), a peptide that possesses a structure identical to authentic CCAP (Christie et al., 2010b), as well as the linker/precursor-related peptides DDVILQ and YYYYPGPGEVAEPLEPKM. The remaining peptide or peptides predicted from each protein is distinct in structure.

3.1.9. Crustacean hyperglycemic hormone(CHH)/ion transport peptide (ITP)

Fourteen transcripts were identified as encoding putative CHH/ITP precursors (Table 1). Translation of nine of these sequences revealed each to encode an identical 117 amino acid full-length precursor. This protein is most similar to ion transport peptide-like precursor from the brown marmorated stink bug *Halyomorpha halys* (**Accession No. XP_014274474**; unpublished, i5k Genome Sequencing Initiative), and thus has been named here Lyghe-prepro-ITP Ia (Fig. 2H1). Two peptides were predicted from this preprohormone (Table 2 and Fig. 2H1), one of which shows sequence similarity to known isoforms of CHH/ITP, *e.g.*, that derived from *H. halys* protein. Analysis of the putative CHH derived from Lyghe-prepro-ITP Ia suggests disulfide bonds are present between the first and fifth, second and fourth, and third and sixth cysteines, a pattern of bridging considered stereotypic of members of the CHH family (Böcking et al., 2002). Translation of the remaining five transcripts revealed a set of 158 amino acid full-length proteins that are identical in sequence to one another (Lyghe-prepro-ITP Ib; Fig. 2H2) and

differ from Lyghe-prepro-ITP Ia by a 41 amino acid N-terminal extension. This N-terminal extension generates a situation in which SignalP is unable to identify a signal sequence within the protein. At present it is unclear if these additional amino acids are the result of an artifact in the process of assembling the transcripts encoding the protein, or whether the extension has true biological significance, *e.g.*, acting as some sort of regulatory element. Regardless, the peptides predicted from Lyghe-prepro-ITP Ib are identical in structure to those derived from Lyghe-prepro-ITP Ia.

3.1.10. Diuretic hormone 31 (DH31)

Two transcripts were identified encoding DH31 precursors (Table 1). Translation of these sequences revealed each to encode an identical 111 amino acid full-length preprohormone (Lyghe-prepro-DH31; Fig. 2I). Three distinct peptides were predicted from Lyghe-prepro-DH31 (Table 2 and Fig. 2I), one of which, GLDLGLSRGFSGSQAAKHLMGLAAANYAGGPamide, possesses the structural hallmarks of a DH31 family member, *i.e.*, an overall length of 31 amino acids, 13 highly conserved residues (XXDXGLXRGXSGXXXAKXXXXXXXXANXXXGPamide), and an amidated C-terminus (Christie et al., 2010b).

3.1.11. GSEFLamide

A single transcript was identified as encoding a putative GSEFLamide precursor (Table 1). Translation of this transcript revealed a 145 amino acid full-length preprohormone (Lyghe-prepro-GSEFLamide; Fig. 2J). Eight peptides were predicted from Lyghe-prepro-GSEFLamide (Table 2 and Fig. 2J), two of which possess the structure TVGTEFLamide, which appears to be a variant GSEFLamide family member. The six linker/precursor-related peptide all possess

distinct structures. Multiple disulfide bonds were predicted by DiANNA for the linker/precursor-related peptide NSEACHPYEPFKCPGDGACISIQYLCDGAPDCLDGYDEDSRLCTAAVYFD, *i.e.*, bonding between the first and sixth, second and fourth, and third and fifth cysteines.

3.1.12. *Insulin-like peptide (ILP)*

Two transcripts were identified as encoding putative ILP precursors (Table 1). Translation of these sequences revealed them to encode identical 150 amino acid full-length preprohormones (Lyghe-prepro-ILP; Fig. 2K). Five distinct peptides were predicted from Lyghe-prepro-ILP (Table 2 and Fig. 2K), including one with sequence similarity to known ILP A-chain peptides and another with sequence similarity to known ILP B-chain peptides. Homology to other arthropod ILPs (*e.g.*, Dirksen et al., 2011) suggests disulfide bridging to be present within the ILP A-chain (the first and third cysteines bridged in GVVYDECCKKTCSINELASYCGVN; Table 2), and between it and the predicted B-chain (the first cysteine in SHQRYCGRDINEVLKLICHGVYNSMFTN bridged to the second cysteine in the A-chain and the second cysteine in this peptide bridged to the fourth cysteine in the A-chain [interpeptide bridged cysteines are not underlined in Table 2]).

3.1.13. *Myosuppressin*

A single transcript was identified as encoding a putative myosuppressin precursor (Table 1). Translation of this sequence revealed a 93 amino acid full-length preprohormone (Lyghe-prepro-myosuppressin; Fig. 2L). Two distinct peptides were predicted from Lyghe-prepro-myosuppressin (Table 2 and Fig. 2L), one of which pQDVDHVFLRFamide, possesses the

structural motif –HVFLRFamide, the hallmark of members of the myosuppressin family (Christie et al., 2010b).

3.1.14. Neuroparsin

Three transcripts were identified as encoding putative neuroparsin precursors (Table 1). Translation of these sequences revealed each to encode a full-length protein, with one prehormone being 107 amino acids long (Lyghe-pre-neuroparsin I; Fig. 2M1), another 111 amino acids (Lyghe-pre-neuroparsin II; Fig. 2M2), and the third, 131 amino acids in length (Lyghe-pre-neuroparsin III; Fig. 2M3). A single peptide was predicted from each prehormone, each possessing a distinct sequence (Table 2 and Fig. 2M). BLAST searches of the non-redundant arthropod proteins in GenBank identified a neuroparsin-A-like precursor from *H. halys* (Accession No. **XP_014291032**; unpublished, i5k Genome Sequencing Initiative) as the most similar protein to Lyghe-pre-neuroparsin I, a *C. lectularius* neuroparsin-A-like precursor (Accession No. **XP_014256240**; unpublished, i5k Genome Sequencing Initiative) as the most similar protein to Lyghe-pre-neuroparsin II, and a neuroparsin-A precursor from the termite *Zootermopsis nevadensis* (Accession No. **KDR17790**; Terrapon et al., 2014) as the most similar protein to Lyghe-pre-neuroparsin III. Analysis of the putative neuroparsin predicted from each of the three prehormones by DiANNA suggests disulfide bridging between the first and eighth, second and ninth, third and fourteenth, fourth and thirteenth, fifth and tenth, sixth and seventh, and eleventh and twelfth cysteines in the isoform derived from Lyghe-pre-neuroparsin I, between the first and fourth, second and eleventh, third and twelfth, fifth and eighth, sixth and tenth, and seventh and ninth cysteines in the isoform derived from Lyghe-pre-neuroparsin II, and between

the first and fourth, second and seventh, third and ninth, fifth and twelfth, sixth and eighth, and tenth and eleventh cysteines in the isoform derived from Lyghe-pre-neuroparsin III.

3.1.15. Neuropeptide F

A single transcript was identified as encoding a putative NPF precursor (Table 1). Translation of this sequence revealed a 102 amino acid full-length prehormone (Lyghe-prepro-NPF; Fig. 2N). Three peptides were predicted from Lyghe-prepro-NPF (Table 2 and Fig. 2N), one of which, PKTFDSPDDLRTYLNQLGQYYAVAGRPRFamide, possesses the structural hallmarks of a member of the NPF family, *i.e.*, an overall length of approximately 30 amino acids, the C-terminal motif –GRPRFamide, and tyrosine residues at positions 10 and 17 from the C-terminus (Christie et 2010b).

3.1.16. Orcokinin

Two transcripts were identified as encoding putative orcokinin precursors (Table 1). Translation of these two sequences revealed each to encode an identical 169 amino acid full-length prehormone (Lyghe-prepro-orcokinin; Fig. 2O). Nine distinct peptides were predicted from Lyghe-prepro-orcokinin (Table 2 and Fig. 2O), two of which possess the N-terminal motif NFDEIDR–, the hallmark of the orcokinin family (Christie et al., 2010b). Three other peptides derived from Lyghe-prepro-orcokinin, PMDEIDRAGFTRFQ, PMDEIDRAGFTHFV and PMDEIDRAGFDHFV, appear to be variant orcokinins, possessing PMDEIDR– rather than NFDEIDR– N-termini. The peptide FDSLNGGTFGAD, which was also predicted from Lyghe-prepro-orcokinin, shows sequence similarity to known orcomyotropins (Christie et al., 2010b), a

peptide group also produced from orcokinin precursors, and is considered here to be a member of this peptide family.

3.1.17. Pyrokinin

A single transcript was identified as encoding a putative pyrokinin precursor (Table 1). Translation of this sequence revealed a 131 amino acid full-length preprohormone (Lyghe-prepro-pyrokinin; Fig. 2P). Six distinct peptides were predicted from Lyghe-prepro-pyrokinin (Table 2 and Fig. 2P), two of which, NSASFQPRSamide and DEESQFTETSRPPFAPRLamide, possess the C-terminal motif –FXPRXamide, the hallmark of the pheromone biosynthesis activating neuropeptide/diapause hormone/pyrokinin superfamily (Rafaeli, 2009).

3.1.18. Short neuropeptide F (sNPF)

A single transcript was identified as encoding a putative sNPF precursor (Table 1). Translation of this sequence revealed a 91 amino acid full-length preprohormone (Lyghe-prepro-sNPF; Fig. 2Q). Four distinct peptides were predicted from Lyghe-prepro-sNPF (Table 2 and Fig. 2Q), one of which, DNRNPQFRLRFamide, possesses the structural hallmarks of a member of the sNPF family, *i.e.*, the C-terminal motif –RXRFamide and an overall length of 10 or so amino acids (Christie et al., 2010b).

3.1.19. SIFamide

Four transcripts were identified as encoding putative SIFamide precursors (Table 1). Translation of these sequences revealed each to encode an identical 72 amino acid full-length

precursor (Lyghe-prepro-SIFamide; Fig. 2R). Two distinct peptides, and a single tyrosine residue (not shown in Table 2), were predicted from Lyghe-prepro-SIFamide (Table 2 and Fig. 2R). One of the predicted peptides, PPFNGSIFamide, possesses the C-terminal motif –SIFamide, the defining feature of members of the SIFamide family (Christie et al., 2010b).

3.1.20. *Sulfakinin*

A single transcript was identified as encoding a putative sulfakinin precursor (Table 1). Translation of this sequence revealed a 93 amino acid full-length preprohormone (Lyghe-prepro-sulfakinin; Fig. 1B). Four distinct peptides were predicted from Lyghe-prepro-sulfakinin (Table 2 and Fig. 1B), two of which, pQFNDYGHMRamide and SGGDEKFDDYGYMRamide, possess the C-terminal motif –Y_(SO₃H)GHM/LRamide (or a close approximation thereof), the defining feature of members of the sulfakinin family (Christie et al., 2010b).

3.1.21. *Tachykinin-related peptide (TRP)*

Two transcripts were identified as encoding putative TRP precursors (Table 1). Translation of one of these sequences revealed a 192 amino acid full-length preprohormone (Lyghe-prepro-TRP Ia; Fig. 2S1), while translation of the other revealed a 217 amino acid full-length protein (Lyghe-prepro-TRP Ib; Fig. 2S2); Lyghe-prepro-TRP Ia and Ib are identical in amino acid sequence over much of their lengths, differing only at their C-termini. Eleven distinct peptides were predicted from Lyghe-prepro-TRP Ia (Table 2 and Fig. 2S1), six of which possess the C-terminal motif –FXGXRamide, the hallmark of members of the TRP family (Christie et al., 2010b). The peptide pEDYGFWAGEEDHPGFYDPRamide may also be a member of the TRP family, though its C-terminus differs slightly from the standard –

FXGXRamide (Table 2). Twelve peptides were predicted from Lyghe-prepro-TRP Ib (Table 2 and Fig. 2S2), ten of which, including all TRP isoforms, are shared with Lyghe-prepro-TRP Ia.

3.2. Sequence validation of predicted peptide-encoding transcripts.

To help confirm the validity of the deduced preprohormone sequences, and hence predicted peptide isoforms, a subset of putative peptide-encoding transcripts was amplified from mixed sex adult *L. hesperus* head cDNAs, and the respective nucleotide sequences compared with the predicted assemblies. Amplimers of the expected sizes were obtained for all seven transcripts (Fig. 3A), with the nucleotide sequences of most of the amplified products identical to their respective transcriptomic sequences. Unexpectedly, multiple variants not identified in the transcriptomic data were also amplified for the allatotropin and DH31 transcripts. Two allatotropin transcripts (designated allatotropin A and allatotropin B), were amplified from the adult head cDNAs (Fig. 3B). While multiple allatotropin and allatotropin-like variants have been reported previously in a number of Lepidoptera (Horodyski et al., 2001; Abdel-Latief et al., 2003; Nagata et al., 2012), the 12 nucleotide insertion between nucleotides 92-103 that characterizes allatotropin A has no effect on the putative allatotropin peptide sequence. The 108 and 120 nucleotide insertions upstream of the dibasic cleavage site in the *L. hesperus* DH31 A and B variants (Fig. 3C) are similar to those reported for DH31 variants in *Rhodnius prolixus* (Zandawala et al., 2011). However, as with the allatotropin deletion, the insertions in both species, which likely arose from either horizontal gene transfer or transposase activity into the shorter ancestral variant (Zandawala et al., 2011), have no effect on the putative DH31 peptide.

4. Discussion

4.1. Prediction of peptidomes using transcriptome shotgun assembly datasets

Arthropods have long played significant roles in advancing our understanding of peptidergic signaling. For example, arthropod peptides were among the first members of this signaling molecule class to be isolated and fully characterized (*e.g.*, Fernlund and Josefsson, 1968, 1972), and members of this phylum were the subjects of the first formal demonstrations of neurosecretion, including release of peptide hormones (Bliss, 1951; Passano, 1951). Over the past several decades, the identification of arthropod peptidomes has moved from a strategy of one-by-one peptide identifications, typically through biochemical isolation and sequencing and/or molecular cloning, to the rapid, and essentially simultaneous deduction of large sets of peptides using mass spectrometry, most commonly via high-resolution mass profiling and/or tandem mass spectrometric sequencing (*e.g.*, Christie et al., 2010b). While a powerful tool, mass spectrometry often requires a relatively large amount of starting material to identify and fully characterize large sets of peptide structures, and thus is not particularly practical for peptidome discovery in animals that are small, rare and/or difficult to collect. For these species, *in silico* genome/transcriptome mining provides an alternative means of peptidome prediction, as much more limited starting material, often just that from a single individual, is required to generate the sequences needed to produce a genomic/transcriptomic dataset. Once in hand, simple BLAST searches can be used to mine the genome/transcriptome for peptide-encoding transcripts, and using the proteins deduced from the identified sequences, mature peptide structures can be rapidly predicted through the use of well-vetted bioinformatics workflows. Over the last decade, large peptidomes have been characterized from a wide variety of arthropods, including members of all of the extant subphyla that form this taxon (*e.g.*, Christie, 2008a, 2008b, 2014a, 2014b, 2014c, 2014d, 2014e, 2014f, 2015a, 2015b, 2015c, 2015d, 2016a, 2016b; Christie and Chi,

2015a, 2015b, 2015c; Christie et al., 2008a, 2010a, 2011a, 2011c; 2013, 2015; Christie and Pascual, 2016; Dircksen et al., 2011; Gard et al., 2009; Hauser et al., 2010; Hummon et al., 2006; Ma et al., 2009, 2010; Toullec et al., 2013; Ventura et al., 2014; Weaver and Audsley, 2008; Yan et al., 2012). Given the success of *in silico* transcriptome mining for peptide discovery in other arthropods, and the existence of a large transcriptome for *L. hesperus*, we had confidence that this strategy could be used for peptidome prediction in this species, and here such a study was conducted.

4.2. Prediction of a peptidome for *Lygus hesperus*

Using known arthropod precursor proteins as the query sequences, transcriptomic data from *L. hesperus* (**BioProject Nos. PRJNA210219 and PRJNA238835** [Hull et al., 2014]) was searched for sequences encoding putative homologs. Via these searches, 53 transcripts encoding putative peptide pre/preprohormones were identified from the approximately 64,000 transcripts extant for this species. The proteins deduced from these sequences, all but one full-length, allowed for the prediction of 119 distinct peptides, which included members of the AST-A (six isoforms), AST-B (12 isoforms), AST-C (one peptide), allatotropin (one peptide), bursicon β (one peptide, which is a partial sequence), CCHamide (two isoforms), corazonin (one peptide), CCAP (one peptide), CHH/ITP (one peptide), DH31 (one peptide), GSEFLamide (one peptide), ILP (one A- and one B-chain peptide), myosuppressin (one peptide), neuroparsin (three isoforms), NPF (one peptide), orcokinin (five isoforms), orcomyotropin (one peptide), pyrokinin (two isoforms), sNPF (one peptide), SIFamide (one peptide), sulfakinin (two isoforms) and TRP (seven isoforms) families. In addition, a large number of peptides whose structures do not place them into any generally recognized family were also predicted from the deduced

pre/preprohormones (termed here linker/precursor-related peptides). To the best of our knowledge, these are the only putative mature peptides thus far described for *L. hesperus*.

While extensive, the *L. hesperus* peptidome presented here is undoubtedly incomplete, as no transcripts encoding members of several well-known and highly conserved insect peptide families were found within the extant TSA dataset. Specifically, no transcripts encoding adipokinetic hormone, bursicon α , diuretic hormone 44, ecdysis-triggering hormone, eclosion hormone, FMRFamide-like peptide, intocin, leucokinin, pigment dispersing factor, proctolin, or RYamide were identified within the *L. hesperus* transcriptome sequences. Since members of each of these peptide groups have been identified from a variety of other insects (Ons et al., 2010; Xu et al., 2016; Tanaka et al., 2014; Predel et al., 2012; Caers et al., 2015; Audsley et al., 2015), it seems highly likely that they exist in *L. hesperus* as well. Given the size of the *L. hesperus* dataset searched here (~64,000 transcripts), the absence of transcripts encoding some peptide groups is not unprecedented as previous transcriptome mining studies suggest that peptide-encoding transcripts are relatively rare, with one typically found for every 5,000 or so sequences present in a whole organism transcriptome assembly (A.E. Christie, unpublished), e.g., those used here for prediction of the *L. hesperus* peptide structures. In fact, our discovery of 53 peptide-encoding transcripts within the extant *L. hesperus* collection is quite surprising, as given the current metric, only 13 peptide-encoding sequences would be expected to be present in this dataset. As additional transcriptomic and/or genomic data are generated for *L. hesperus*, it will be interesting to see how complete a peptidome can be deduced for this species.

4.3. The peptidome of Lygus hesperus includes a number of atypical peptide isoforms

509 As can be gleaned from the data presented here, a large number of peptide families exist
 510 in arthropods. For some of these groups, *e.g.*, the CCAP and proctolin families, essentially all
 511 species possess a single, highly conserved isoform, *i.e.*, PFCNAFTGCamide (a disulfide bridge
 512 between the two cysteine residues) for CCAP and RYLPT for proctolin (Lee et al., 2011;
 513 Orchard et al., 2011). For other families, multiple isoforms are present, but essentially all
 514 possess stereotypical structural hallmarks. For example, members of the NPF family typically
 515 exhibit an overall length of approximately 30 amino acids, possess the C-terminal motif –
 516 GRPRFamide, and have tyrosine residues at positions 10 and 17 from the C-terminus (Christie et
 517 2010b). As would be expected, most of the isoforms of the peptide families identified here
 518 appear to follow the standard “rules” for the arthropods, *e.g.*, the NPF isoform predicted for *L.*
 519 *hesperus*, PKTFDSPDDLRTYLNQLGQYYAVAGRPRFamide, is 29 amino acids long,
 520 possesses a –GRPRFamide C-terminus, and has tyrosine residues at positions 10 and 17 from its
 521 C-terminus. Interestingly, a number of *L. hesperus* peptides were predicted that are atypical for
 522 their respective families. Specifically, eight of the twelve *L. hesperus* AST-Bs possess the C-
 523 terminal motif –WX₇Wamide rather than the stereotypical –WX₆Wamide (where the Xs represent
 524 variable residues), while three of the five orcokinin predicted for this species possess
 525 PMDEIDR– rather than NFDEIDR– N-termini. Similarly, the *L. hesperus* allatotropin exhibits a
 526 –SARGFamide rather than –TARGFamide C-terminus, its GSEFLamide ends in –GTEFLamide,
 527 and the predicted SIFamide is eight rather than 12 amino acids long. In addition, one isoform of
 528 AST-A, AEGKHWNFGVamide, possesses a variant on the standard –YXFGLamide consensus
 529 motif, while two TRPs, APQMGFSAMRamide and pEDYGFWAGEEDHPGFYDPRamide,
 530 possess variants on the stereotypical –FXGXRamide C-terminus that defines members of this
 531 family. While atypical peptides similar to some of those seen in *L. hesperus* are present in at

least some other species, *e.g.*, the aphids *Aphis gossypii* and *Myzus persicae* are both predicted to possess –WX₇Wamide AST-Bs (Christie, 2008b) and a SIFamide identical that predicted here in *L. hesperus* also appears to be present in the brown planthopper *Nilaparvata lugens* (Tanaka et al., 2014), others are described here for the first time, *e.g.*, variant allatotropin, GSEFLamide and the orcokinin possessing PMDEIDR– N-termini. What, if any, significance these variant sequences may have on function remains to be determined.

4.4. The peptidome of *Lygus hesperus* as resource for biological control

While *Lygus* control strategies have largely relied on broad-spectrum insecticides, reports of resistance to traditional insecticide chemistries among *Lygus* field populations (Snodgrass et al., 2000; Snodgrass et al., 2009; Zhu and Luttrell 2012) underscore the need for new control options. Given their role in regulating critical physiological processes, targeted disruption of insect peptidergic systems offers clear promise as a pest management strategy (Scherkenbeck and Zdobinsky 2009; Altstein and Nässel 2010; Hoffmann et al., 2014; Audsley and Down 2015). Poor cuticular penetrance and rapid hemolymph-based proteolytic degradation, however, precludes the use of neuropeptides themselves. Rather, the use of peptidomimetics that incorporate chemical modifications to increase permeability and stability offer a potentially intriguing means of disrupting vital peptidergic systems *in vivo*. Peptidomimetics of different classes of insect kinins have been shown to stimulate insect diuresis (Nachman et al., 2011), disrupt ecdysis (Lange et al., 2016), impair sugar taste neurons (Kwon et al., 2016), and exhibit aphicidal activities (Smagghe et al., 2010; Nachman et al., 2011). Similarly, peptide-based agonists/antagonists of the pyrokinin neuropeptides promote irregular pupariation (Nachman et al., 2008), inhibit sex pheromone production (Nachman et al., 2009; Hariton et al., 2010), and

disrupt pupal diapause (Zhang et al., 2011; Zhang et al., 2015). Despite current delivery challenges, the exquisite specificity inherent to RNA interference (RNAi)-mediated gene knockdown offers an alternative mechanism for targeted disruption. Indeed, knockdown of peptides similar to those identified in this *L. hesperus* peptidome (e.g. allatostatins, tachykinin-related peptides, insulin-like peptides, pyrokinins, corazonin, etc) have been reported to negatively impact pupal/larval development (Abdel-latif and Hoffmann, 2014; Meng et al., 2015), mating behavior and fecundity (Jiang et al., 2013), pupation and carbohydrate metabolism (Fu et al., 2016; Defferrari et al., 2015), pheromone production (Lee et al., 2015), and pigmentation (Sugahara et al., 2015).

4.5. Summary

In silico transcriptome methods have proven to be invaluable tools for mass identification and characterization of neuropeptides from non-model organisms. The peptides predicted here for the agricultural pest species *L. hesperus* provide foundational information that can guide and inform future integrative approaches with the potential to enhance our understanding of neuropeptide structure-activity relationships, facilitate identification of cognate receptors and their mechanisms of physiological activation and action, and may ultimately uncover unique targets that can be exploited for pest control.

Acknowledgements

The authors thank Brooks Silversmith (Arid Land Agricultural Research Center [ALARC]) and Anna Cervantes (ALARC) for maintaining the *L. hesperus* colony and Lynn Forlow Jech (ALARC) for tissue dissections. Financial support for this study was provided by the National

Science Foundation (IOS-1353023 to A.E.C), the Cades Foundation of Honolulu, Hawaii (to A.E.C.) and Cotton Inc. (12-373 to J.J.H). The transcript shotgun assemblies were performed on computing resources at the USDA-ARS Daniel K. Inouye Pacific Basin Agricultural Research Center (Moana cluster; Hilo, HI) and the Extreme Science and Engineering Discovery Environment (XSEDE), which are supported by National Science Foundation (OCI-1053575XSEDE under allocation TG-MCB140032 to S.M.G). 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 U. S. Department of Agriculture. USDA is an equal opportunity provider and employer.

References

Abdel-latif, M., Meyering-Vos, M., Hoffmann, K.H., 2003. Molecular characterisation of cDNAs from the fall armyworm *Spodoptera frugiperda* encoding *Manduca sexta* allatotropin and allatostatin preprohormone peptides. *Insect Biochem. Mol. Biol.* 33, 467–476.

Abdel-Latif, M., Hoffmann, K.H., 2014. Functional activity of allatotropin and allatostatin in the pupal stage of a holometabolous insect, *Tribolium castaneum* (Coleoptera, Tenebrionidae). *Peptides* 53, 172–184.

Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Scherer, S.E., Li, P.W., Hoskins, R.A., Galle, R.F., George, R.A., Lewis, S.E., Richards, S., Ashburner, M., Henderson, S.N., Sutton, G.G., Wortman, J.R., Yandell, M.D., Zhang, Q., Chen, L.X., Brandon, R.C., Rogers, Y.H., Blazej, R.G., Champe, M., Pfeiffer, B.D., Wan, K.H., Doyle,

601 C., Baxter, E.G., Helt, G., Nelson, C.R., Gabor, G.L., Abril, J.F., Agbayani, A., An, H.J.,
 602 Andrews-Pfannkoch, C., Baldwin, D., Ballew, R.M., Basu, A., Baxendale, J., Bayraktaroglu, L.,
 603 Beasley, E.M., Beeson, K.Y., Benos, P.V., Berman, B.P., Bhandari, D., Bolshakov, S., Borkova,
 604 D., Botchan, M.R., Bouck, J., Brokstein, P., Brottier, P., Burtis, K.C., Busam, D.A., Butler, H.,
 605 Cadieu, E., Center, A., Chandra, I., Cherry, J.M., Cawley, S., Dahlke, C., Davenport, L.B.,
 606 Davies, P., de Pablos, B., Delcher, A., Deng, Z., Mays, A.D., Dew, I., Dietz, S.M., Dodson, K.,
 607 Doup, L.E., Downes, M., Dugan-Rocha, S., Dunkov, B.C., Dunn, P., Durbin, K.J., Evangelista,
 608 C.C., Ferraz, C., Ferriera, S., Fleischmann, W., Fosler, C., Gabrielian, A.E., Garg, N.S., Gelbart,
 609 W.M., Glasser, K., Glodek, A., Gong, F., Gorrell, J.H., Gu, Z., Guan, P., Harris, M., Harris,
 610 N.L., Harvey, D., Heiman, T.J., Hernandez, J.R., Houck, J., Hostin, D., Houston, K.A., Howland,
 611 T.J., Wei, M.H., Ibegwam, C., Jalali, M., Kalush, F., Karpen, G.H., Ke, Z., Kennison, J.A.,
 612 Ketchum, K.A., Kimmel, B.E., Kodira, C.D., Kraft, C., Kravitz, S., Kulp, D., Lai, Z., Lasko, P.,
 613 Lei, Y., Levitsky, A.A., Li, J., Li, Z., Liang, Y., Lin, X., Liu, X., Mattei, B., McIntosh, T.C.,
 614 McLeod, M.P., McPherson, D., Merkulov, G., Milshina, N.V., Mobarry, C., Morris, J., Moshrefi,
 615 A., Mount, S.M., Moy, M., Murphy, B., Murphy, L., Muzny, D.M., Nelson, D.L., Nelson, D.R.,
 616 Nelson, K.A., Nixon, K., Nusskern, D.R., Pacleb, J.M., Palazzolo, M., Pittman, G.S., Pan, S.,
 617 Pollard, J., Puri, V., Reese, M.G., Reinert, K., Remington, K., Saunders, R.D., Scheeler, F.,
 618 Shen, H., Shue, B.C., Sidén-Kiamos, I., Simpson, M., Skupski, M.P., Smith, T., Spier, E.,
 619 Spradling, A.C., Stapleton, M., Strong, R., Sun, E., Svirskas, R., Tector, C., Turner, R., Venter,
 620 E., Wang, A.H., Wang, X., Wang, Z.Y., Wassarman, D.A., Weinstock, G.M., Weissenbach, J.,
 621 Williams, S.M., Woodage, T., Worley, K.C., Wu, D., Yang, S., Yao, Q.A., Ye, J., Yeh, R.F.,
 622 Zaveri, J.S., Zhan, M., Zhang, G., Zhao, Q., Zheng, L., Zheng, X.H., Zhong, F.N., Zhong, W.,

623 Zhou, X., Zhu, S., Zhu, X., Smith, H.O., Gibbs, R.A., Myers, E.W., Rubin, G.M., Venter, J.C.,
 624 2000. The genome sequence of *Drosophila melanogaster*. *Science* 287, 2185-2195.
 625
 626 Altstein, M., 2004. Novel insect control agents based on neuropeptide antagonists: The
 627 PK/PBAN family as a case study. *J. Mol. Neurosci.* 22, 147-157.
 628
 629 Altstein, M., Nässel, D.R., 2010. Neuropeptide signaling in insects. *Adv. Exp. Med. Biol.* 692,
 630 155–165.
 631
 632 Audsley, N., Down, R.E., 2015. G protein coupled receptors as targets for next generation
 633 pesticides. *Insect Biochem. Mol. Biol.* 67, 27-37.
 634
 635 Audsley, N., Down, R.E., Isaac, R.E., 2015. Genomic and peptidomic analyses of the
 636 neuropeptides from the emerging pest, *Drosophila suzukii*. *Peptides* 68, 33–42.
 637
 638 Backus, E.A., Serrano, M.S., Ranger, C.M., 2005. Mechanisms of hopperburn: an overview of
 639 insect taxonomy, behavior and physiology. *Annu. Rev. Entomol.* 50, 125-151.
 640
 641 Backus, E.A., Cline, A.R., Ellerseick, M.R., Serrano, M.S., 2007. *Lygus hesperus* (Hemiptera:
 642 Miridae) feeding on cotton: new methods and parameters for analysis of non-sequential electrical
 643 penetration graph data. *Ann. Entomol. Soc. Am.* 100, 296–310.
 644

645 Bliss, D.E., 1951. Metabolic effects of sinus gland or eyestalk removal in the land crab,
 646 *Gecarcinus lateralis*. Anat. Rec. 111, 502-503.
 647
 648 Böcking, D., Dircksen, H., Keller, R., 2002. The crustacean neuropeptide of the CHH/MIH/GIH
 649 family: structures and biological activities. In: The Crustacean Nervous System. K Wiese (Ed.).
 650 Springer, Heidelberg. pp. 84–97.
 651
 652 Caers, J., Boonen, K., Van Den Abbeele, J., Van Rompay, L., Schoofs, L., Van Hiel, M.B., 2015.
 653 Peptidomics of neuropeptidergic tissues of the tsetse fly *Glossina morsitans morsitans*. J. Am.
 654 Soc. Mass Spectrom. 26, 2024–2038.
 655
 656 Christie, A.E., 2008a. Neuropeptide discovery in Ixodoidea: an *in silico* investigation using
 657 publicly accessible expressed sequence tags. Gen. Comp. Endocrinol. 157, 174-185.
 658
 659 Christie, A.E., 2008b. *In silico* analyses of peptide paracrine/hormones in Aphidoidea. Gen.
 660 Comp. Endocrinol. 159, 67-79.
 661
 662 Christie, A.E., 2014a. Prediction of the first neuropeptides from a member of the Remipedia
 663 (Arthropoda, Crustacea). Gen. Comp. Endocrinol. 201, 74-86.
 664
 665 Christie, A.E., 2014b. Prediction of the peptidomes of *Tigriopus californicus* and
 666 *Lepeophtheirus salmonis* (Copepoda, Crustacea). Gen. Comp. Endocrinol. 201, 87-106.
 667

668 Christie, A.E., 2014c. Peptide discovery in the ectoparasitic crustacean *Argulus siamensis*:
 669 identification of the first neuropeptides from a member of the Branchiura. Gen. Comp.
 670 Endocrinol. 204, 114-125.
 671
 672 Christie, A.E., 2014d. *In silico* characterization of the peptidome of the sea louse *Caligus*
 673 *rogercresseyi* (Crustacea, Copepoda). Gen. Comp. Endocrinol. 204, 248-260.
 674
 675 Christie, A.E., 2014e. Identification of the first neuropeptides from the Amphipoda (Arthropoda,
 676 Crustacea). Gen. Comp. Endocrinol. 206, 96-110.
 677
 678 Christie, A.E., 2014f. Expansion of the *Litopenaeus vannamei* and *Penaeus monodon*
 679 peptidomes using transcriptome shotgun assembly sequence data. Gen. Comp. Endocrinol. 206,
 680 235-254.
 681
 682 Christie, A.E., 2015a. *In silico* characterization of the neuropeptidome of the Western black
 683 widow spider *Latrodectus hesperus*. Gen. Comp. Endocrinol. 210, 63-80.
 684
 685 Christie, A.E., 2015b. Neuropeptide discovery in *Eucyclops serrulatus* (Crustacea, Copepoda): in
 686 silico prediction of the first peptidome for a member of the Cyclopoida. Gen. Comp. Endocrinol.
 687 211, 92-105.
 688
 689 Christie, A.E., 2015c. Neuropeptide discovery in *Symphylella vulgaris* (Myriapoda, Symphyla):
 690 *in silico* prediction of the first myriapod peptidome. Gen. Comp. Endocrinol. 223, 73-86.

691

692 Christie, A.E., 2015d. *In silico* prediction of a neuropeptidome for the eusocial insect

693 *Mastotermes darwiniensis*. Gen. Comp. Endocrinol. 224, 69-83.

694

695 Christie, A.E., 2016a. Prediction of *Scylla olivacea* (Crustacea; Brachyura) peptide hormones

696 using publicly accessible transcriptome shotgun assembly (TSA) sequences. Gen. Comp.

697 Endocrinol. 230-231, 1–16.

698

699 Christie, A.E., 2016b. Expansion of the neuropeptidome of the globally invasive marine crab

700 *Carcinus maenas*. Gen. Comp. Endocrinol. 235, 150-169.

701

702 Christie, A.E., Cashman, C.R., Brennan, H.R., Ma, M., Sousa, G.L., Li, L., Stemmler, E.A.,

703 Dickinson, P.S., 2008a. Identification of putative crustacean neuropeptides using *in silico*

704 analyses of publicly accessible expressed sequence tags. Gen. Comp. Endocrinol. 156, 246-264.

705

706 Christie, A.E., Chi, M., 2015a. Neuropeptide discovery in the Araneae (Arthropoda, Chelicerata,

707 Arachnida): elucidation of true spider peptidomes using that of the Western black widow as a

708 reference. Gen. Comp. Endocrinol. 213, 90-109.

709

710 Christie, A.E., Chi, M., 2015b. Identification of the first neuropeptides from the enigmatic

711 hexapod order Protura. Gen. Comp. Endocrinol. 224, 18–37.

712

713 Christie, A.E., Chi, M., 2015c. Prediction of the neuropeptidomes of members of the Astacidea
 714 (Crustacea, Decapoda) using publicly accessible transcriptome shotgun assembly (TSA)
 715 sequence data. *Gen. Comp. Endocrinol.* 224, 38–60.
 716
 717 Christie, A.E., Chi, M., Lameyer, T.J., Pascual, M.G., Shea, D.N., Stanhope, M.E., Schulz, D.J.,
 718 Dickinson, P.S., 2015. Neuropeptidergic signaling in the American lobster *Homarus*
 719 *americanus*: new insights from high-throughput nucleotide sequencing. *PLoS ONE*. 10,
 720 e0145964.
 721
 722 Christie, A.E., Cieslak, M.C., Roncalli, V., Lenz, P.H., Major, K.M., Poynton, H.C., 2016.
 723 Prediction of *Hyaella azteca* (Crustacea; Amphipoda) peptide hormones using a *de novo*
 724 transcriptome assembly. *PLoS ONE* Submitted.
 725
 726 Christie, A.E., Durkin, C.S., Hartline, N., Ohno, P., Lenz, P.H., 2010a. Bioinformatic analyses
 727 of the publicly accessible crustacean expressed sequence tags (ESTs) reveal numerous novel
 728 neuropeptide-encoding precursor proteins, including ones from members of several little studied
 729 taxa. *Gen. Comp. Endocrinol.* 167, 164-178.
 730
 731 Christie, A.E., McCoole, M.D., Harmon, S.M., Baer, K.N., Lenz, P.H., 2011c. Genomic
 732 analyses of the *Daphnia pulex* peptidome. *Gen. Comp. Endocrinol.* 171, 131-150.
 733
 734 Christie, A.E., Nolan, D.H., Garcia, Z.A., McCoole, M.D., Harmon, S.M., Congdon-Jones, B.,
 735 Ohno, P., Hartline, N., Congdon, C.B., Baer, K.N., Lenz, P.H., 2011b. Bioinformatic prediction

736 of arthropod/nematode-like peptides in non-arthropod, non-nematode members of the
 737 Ecdysozoa. Gen. Comp. Endocrinol. 170, 480-486.
 738
 739 Christie, A.E., Nolan, D.H., Ohno, P., Hartline, N., Lenz, P.H., 2011a. Identification of
 740 chelicerate neuropeptides using bioinformatics of publicly accessible expressed sequence tags.
 741 Gen. Comp. Endocrinol. 170, 144-155.
 742
 743 Christie A.E., Pascual, M.G., 2016. Peptidergic signaling in the crab *Cancer borealis*: tapping
 744 the power of transcriptomics for neuropeptidome expansion. Gen. Comp. Endocrinol. In press.
 745
 746 Christie, A.E., Roncalli, V., Wu, L.S., Ganote, C.L., Doak, T., Lenz, P.H., 2013. Peptidergic
 747 signaling in *Calanus finmarchicus* (Crustacea, Copepoda): *in silico* identification of putative
 748 peptide hormones and their receptors using a *de novo* assembled transcriptome. Gen. Comp.
 749 Endocrinol. 187, 117-135.
 750
 751 Christie, A.E., Stemmler, E.A., Dickinson, P.S., 2010b. Crustacean neuropeptides. Cell. Mol.
 752 Life Sci. 67, 4135-4169.
 753
 754 Cooper, W.R. Spurgeon, D.W., 2013. Feeding injury to cotton caused by *Lygus hesperus*
 755 (Hemiptera: Miridae) nymphs and pre-reproductive adults. Environ. Entomol. 42, 967–972.
 756

757 Defferrari, M.S., Orchard, I., Lange, A.B., 2016. Identification of the first insulin-like peptide in
 758 the disease vector *Rhodnius prolixus*: Involvement in metabolic homeostasis of lipids and
 759 carbohydrates. *Insect Biochem. Mol. Biol.*, 70, 148–159.
 760
 761 Dircksen, H., Neupert, S., Predel, R., Verleyen, P., Huybrechts, J., Strauss, J., Hauser, F.,
 762 Stafflinger, E., Schneider, M., Pauwels, K., Schoofs, L., Grimmlikhuijzen, C.J., 2011.
 763 Genomics, transcriptomics, and peptidomics of *Daphnia pulex* neuropeptides and protein
 764 hormones. *J. Proteome Res.* 10, 4478-4504.
 765
 766 Fernlund, P., Josefsson, L., 1968. Chromactivating hormones of *Pandalus borealis*. Isolation and
 767 purification of the 'red-pigment-concentrating hormone'. *Biochim. Biophys. Acta* 158, 262-273.
 768
 769 Fernlund, P., Josefsson, L., 1972. Crustacean color-change hormone: amino acid sequence and
 770 chemical synthesis. *Science* 177, 173-175.
 771
 772 Ferrè, F., Clote, P., 2005. DiANNA: a web server for disulfide connectivity prediction. *Nucleic*
 773 *Acids Res.* 33,W230-W232.
 774
 775 Fu, K-Y., Zhu, T-T., Guo, W-C., Ahmat, T. and Li, G-Q. 2016. Knockdown of a putative
 776 insulin-like peptide gene LdILP2 in *Leptinotarsa decemlineata* by RNA interference impairs
 777 pupation and adult emergence. *Gene* 581, 170–177.
 778

779 Gard, A.L., Lenz, P.H., Shaw, J.R., Christie, A.E., 2009. Identification of putative peptide
 780 paracrine/hormones in the water flea *Daphnia pulex* (Crustacea; Branchiopoda; Cladocera)
 781 using transcriptomics and immunohistochemistry. *Gen. Comp. Endocrinol.* 160, 271-287.
 782
 783 Hansen, K.K., Hauser, F., Williamson, M., Weber, S.B., Grimmelikhuijzen, C.J., 2011. The
 784 *Drosophila* genes CG14593 and CG30106 code for G-protein-coupled receptors specifically
 785 activated by the neuropeptides CCHamide-1 and CCHamide-2. *Biochem. Biophys. Res.*
 786 *Commun.* 404, 184-189.
 787
 788 Hansen, K.K., Stafflinger, E., Schneider, M., Hauser, F., Cazzamali, G., Williamson, M.,
 789 Kollmann, M., Schachtner, J., Grimmelikhuijzen, C.J., 2010. Discovery of a novel insect
 790 neuropeptide signaling system closely related to the insect adipokinetic hormone and corazonin
 791 hormonal systems. *J. Biol. Chem.* 285, 10736-10747.
 792
 793 Hariton, A., Ben-Aziz, O., Davidovitch, M., Altstein, M., 2010. Bioavailability of backbone
 794 cyclic PK/PBAN neuropeptide antagonists--inhibition of sex pheromone biosynthesis elicited by
 795 the natural mechanism in *Heliothis peltigera* females. *FEBS J.* 277, 1035-1044.
 796
 797 Hauser, F., Neupert, S., Williamson, M., Predel, R., Tanaka, Y., Grimmelikhuijzen, C.J., 2010.
 798 Genomics and peptidomics of neuropeptides and protein hormones present in the parasitic wasp
 799 *Nasonia vitripennis*. *J. Proteome Res.* 9, 5296-5310.
 800

801 Hoffmann, K.H., Weidlichb, S., Wendec, F., 2014. Neuropeptide signaling and RNA
 802 interference, in: Hoffmann, K.H. (Ed.), *Insect Molecular Biology and Ecology*. CRC Press, pp.
 803 188-216.
 804
 805 Horodyski, F.M., Bhatt, S.R., Lee, K.Y., 2001. Alternative splicing of transcripts expressed by
 806 the *Manduca sexta* allatotropin (Mas-AT) gene is regulated in a tissue-specific manner. *Peptides*
 807 22, 263–269.
 808
 809 Hull, J.J., Chaney, K., Geib, S.M., Fabrick, J.A., Brent, C.S., Walsh, D., Lavine, L.C., 2014.
 810 Transcriptome-based identification of ABC transporters in the western tarnished plant bug *Lygus*
 811 *hesperus*. *PLoS One* 9, e113046.
 812
 813 Hummon, A.B., Richmond, T.A., Verleyen, P., Baggerman, G., Huybrechts, J., Ewing, M.A.,
 814 Vierstraete, E., Rodriguez-Zas, S.L., Schoofs, L., Robinson, G.E., Sweedler, J.V., 2006. From
 815 the genome to the proteome: uncovering peptides in the *Apis* brain. *Science* 314, 647-649.
 816
 817 Janssen, T., Claeys, I., Simonet, G., De Loof, A., Girardie, J., Vanden Broeck, J., 2001. cDNA
 818 cloning and transcript distribution of two different neuroparsin precursors in the desert locust,
 819 *Schistocerca gregaria*. *Insect Mol. Biol.* 10, 183-189.
 820
 821 Jiang, H., Lkhagva, A., Daubnerová, I., Chae, H-S., Šimo, L., Jung, S-H., Yoon, Y-K., Lee, N-
 822 R., Seong, J.Y., Žitňan, D., Park, Y., Kim, Y-J., 2013. Natalisin, a tachykinin-like signaling
 823 system, regulates sexual activity and fecundity in insects. *Proc. Natl. Acad. Sci.* 110, E3526–34.

824

825 Kastin, A.J., 2006. Handbook of Biologically Active Peptides, First Edition. Academic Press.

826

827 Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7:

828 improvements in performance and usability. Mol. Biol. Evol. 30, 772-780.

829

830 Kwon, H., Ali Agha, M., Smith, R.C., Nachman, R.J., Marion-Poll, F., Pietrantonio, P.V., 2016.

831 Leucokinin mimetic elicits aversive behavior in mosquito *Aedes aegypti* (L.) and inhibits the

832 sugar taste neuron. Proc. Natl. Acad. Sci. 113, 6880–6885.

833

834 Lange, A.B., Nachman, R.J., Kaczmarek, K. Zabrocki, J., 2016. Biostable insect kinin analogs

835 reduce blood meal and disrupt ecdysis in the blood-gorging Chagas' disease vector, *Rhodnius*

836 *prolixus*. Peptides 80, 108–113.

837

838 Lee, D.H., Paluzzi, J.P., Orchard, I., Lange, A.B., 2011. Isolation, cloning and expression of the

839 Crustacean Cardioactive Peptide gene in the Chagas' disease vector, *Rhodnius prolixus*. Peptides,

840 32, 475–482.

841

842 Lee, D.-W., Shrestha, S., Kim, A.Y., Park, S.J., Yang, C.Y., Kim, Y., Koh, Y.H., 2011. RNA

843 interference of pheromone biosynthesis-activating neuropeptide receptor suppresses mating

844 behavior by inhibiting sex pheromone production in *Plutella xylostella* (L.). Insect Biochem.

845 Mol. Biol. 41, 236–243.

846

847 Ma, M., Bors, E.K., Dickinson, E.S., Kwiatkowski, M.A., Sousa, G.L., Henry, R.P., Smith,
848 C.M., Towle, D.W., Christie, A.E., Li, L., 2009. Characterization of the *Carcinus maenas*
849 neuropeptidome by mass spectrometry and functional genomics. *Gen. Comp. Endocrinol.* 161,
850 320-334.

851

852 Ma, M., Gard, A.L., Xiang, F., Wang, J., Davoodian, N., Lenz, P.H., Malecha, S.R., Christie,
853 A.E., Li, L., 2010. Combining *in silico* transcriptome mining and biological mass spectrometry
854 for neuropeptide discovery in the Pacific white shrimp *Litopenaeus vannamei*. *Peptides* 31, 27-
855 43.

856

857 Meng, Q.-W., Liu, X.-P., Lü, F.-G., Fu, K.-Y., Guo, W.-C., Li, G-Q., 2015. Involvement of a
858 putative allatostatin in regulation of juvenile hormone titer and the larval development in
859 *Leptinotarsa decemlineata* (Say). *Gene* 554, 105–113.

860

861 Monigatti, F., Gasteiger, E., Bairoch, A., Jung, E., 2002. The Sulfinator: predicting tyrosine
862 sulfation sites in protein sequences. *Bioinformatics* 18, 769-770.

863

864 Nachman, R.J., Pietrantonio, P.V., Coast, G.M., 2009. Toward the development of novel pest
865 management agents based upon insect kinin neuropeptide analogues. *Ann. NY Acad. Sci.* 1163,
866 251-261.

867

868 Nachman, R.J., Zubrzak, P., Williams, H., Strey, A., Zdarek, J., 2008. A beta-amino acid
 869 pyrokinin analog induces irregular pupariation behavior in larvae of the flesh fly *Sarcophaga*
 870 *bullata*. *Pestycydy* 1-2, 95–100.
 871
 872 Nachman, R.J., Teal, P.E.A., Aziz, O.B., Davidovitch, M., Zubrzak, P., Altstein, M., 2009.
 873 An amphiphilic PK/PBAN analog is a selective pheromonotropic antagonist that penetrates the
 874 cuticle of a heliothine insect. *Peptides* 30, 616–621.
 875
 876 Nachman, R.J., Mahdian, K., Nassel, D.R., Isaac, R.E., Pryor, N., Smagghe, G., 2011.
 877 Biostable multi-Aib analogs of tachykinin-related peptides demonstrate potent oral aphicidal
 878 activity in the pea aphid *Acyrtosiphon pisum* (Hemiptera: Aphidae). *Peptides* 32, 587–594
 879
 880 Nagata, S., Matsumoto, S., Mizoguchi, A., Nagasawa, H., 2012. Identification of cDNAs
 881 encoding allatotropin and allatotropin-like peptides from the silkworm, *Bombyx mori*. *Peptides*
 882 34, 98–105.
 883
 884 Naranjo, S.E., Ellsworth, P.C., Dierig, D., 2011. Impact of *Lygus* spp. (Hemiptera:Miridae) on
 885 damage, yield, and quality of lesquerella (*Physaria fendleri*), a potential new oil-seed crop. *J.*
 886 *Econ. Entomol.* 104,1575–1583
 887
 888 Ons, S., Sterkel, M., Diambra, L., Urlaub, H., Rivera-Pomar, R., 2010. Neuropeptide precursor
 889 gene discovery in the Chagas disease vector *Rhodnius prolixus*. *Insect Mol. Biol.* 20, 29–44.
 890

891 Orchard, I., Lee, D.H., da Silva, R., Lange, A.B., 2011. The proctolin gene and biological effects
892 of proctolin in the blood-feeding bug, *Rhodnius prolixus*. *Front. Endocrinol.* 2, 59.
893

894 Passano, L.M., 1951. The X-organ-sinus gland system neurosecretory system of crabs. *Anat.*
895 *Rec.* 111, 502.
896

897 Petersen, T.N., Brunak, S., von Heijne, G., Nielsen, H., 2011. SignalP 4.0: discriminating signal
898 peptides from transmembrane regions. *Nat. Methods* 8, 785-786.
899

900 Predel, R., Russell, W.K., Russell, D.H., Suh, C.P.-C., Nachman, R.J., 2012. Neuropeptides of
901 the cotton fleahopper, *Pseudatomoscelis seriatus* (Reuter). *Peptides* 34, 39–43.
902

903 Rafaeli, A., 2009. Pheromone biosynthesis activating neuropeptide (PBAN): regulatory role and
904 mode of action. *Gen. Comp. Endocrinol.* 162, 69-78.
905

906 Ritter, R.A., Lenssen, A., Blodgett, S., Taper, M.A., 2010. Regional assemblages of *Lygus*
907 (Heteroptera: Miridae) in Montana canola fields. *J. Kansas Entomol. Soc.* 83, 297–305.
908

908 Scherkenbeck, J., Zdobinsky, T., 2009. Insect neuropeptides: structures, chemical modifications
909 and potential for insect control. *Bioorg. Med. Chem.* 17, 4071-4084.
910

911 Schuh, R.T., 2013. On-line systematic catalog of plant bugs (Insecta: Heteroptera: Miridae).
912 <http://research.amnh.org/pbi/catalog/>.
913

914 Schwartz, M.D., Foottit, R.G., 1998. Revision of the nearctic species of the genus *Lygus* Hahn
 915 with a review of the palearctic species (Heteroptera: Miridae). Associated Publishers.
 916
 917 Scott, D.R., 1977. An annotated listing of host plants of *Lygus hesperus* Knight. Bull. Entomol.
 918 Soc. Am. 23, 19-22.
 919
 920 Shackel, K.A., Celorio-Mancera, M.P., Ahmadi, H., Greve, L.C., Teuber, L.R., Backus, E.A.,
 921 Labavitch, J. M., 2005. Micro-injection of *Lygus* salivary gland proteins to simulate feeding
 922 damage in alfalfa and cotton flowers. Arch. Insect Physiol. Biochem. 58, 69-83.
 923
 924 Smagghe, G., Mahdian, K., Zubrzak, P., Nachman, R.J., 2010. Antifeedant activity and high
 925 mortality in the pea aphid *Acyrtosiphon pisum* (Hemiptera: Aphidae) induced by biostable
 926 insect kinin analogs. Peptides. 31, 498–505
 927
 928 Snodgrass, G.L., Scott, W.P., 2000. Seasonal changes in pyrethroid resistance in tarnished plant
 929 bug (Heteroptera: Miridae) populations during a three-year period in the Delta area of Arkansas,
 930 Louisiana, and Mississippi. J. Econ. Entomol. 93, 441–446.
 931
 932 Snodgrass, G. L., Gore, J., Abel, C.A., Jackson, R. 2009. Acephate resistance in populations of
 933 the tarnished plant bug (Heteroptera: Miridae) from the Mississippi River Delta. J. Econ.
 934 Entomol. 102, 699-707.
 935

936 Stafflinger, E., Hansen, K.K., Hauser, F., Schneider, M., Cazzamali, G., Williamson, M.,
 937 Grimmelikhuijzen, C.J., 2008. Cloning and identification of an oxytocin/vasopressin-like
 938 receptor and its ligand from insects. *Proc. Natl. Acad. Sci. USA.* 105, 3262-3267.
 939
 940 Strong, F.E., 1970. Physiology of injury caused by *Lygus hesperus*. *J. Econ. Entomol.* 63, 808-
 941 814.
 942
 943 Sugahara, R., Saeki, S., Jouraku, A., Shiotsuki, T., Tanaka, S., 2015. Knockdown of the
 944 corazonin gene reveals its critical role in the control of gregarious characteristics in the desert
 945 locust. *J. Insect Physiol.* 79, 80–87.
 946
 947 Tanaka, Y., Suetsugu, Y., Yamamoto, K., Noda, H., Shinoda, T., 2014. Transcriptome analysis
 948 of neuropeptides and G-protein coupled receptors (GPCRs) for neuropeptides in the brown
 949 planthopper *Nilaparvata lugens*. *Peptides* 53, 125-133.
 950
 951 Taylor, P.A, Bhatt, T.R., Horodyski, F.M., 1996. Molecular characterization and expression
 952 analysis of *Manduca sexta* allatotropin. *Eur. J. Biochem.* 239, 588-596.
 953
 954 Terrapon, N., Li, C., Robertson, H.M., Ji, L., Meng, X., Booth, W., Chen, Z., Childers, C.P.,
 955 Glastad, K.M., Gokhale, K., Gowin, J., Gronenberg, W., Hermansen, R.A., Hu, H., Hunt, B.G.,
 956 Huylmans, A.K., Khalil, S.M., Mitchell, R.D., Munoz-Torres, M.C., Mustard, J.A., Pan, H.,
 957 Reese, J.T., Scharf, M.E., Sun, F., Vogel, H., Xiao, J., Yang, W., Yang, Z., Yang, Z., Zhou, J.,
 958 Zhu, J., Brent, C.S., Elsik, C.G., Goodisman, M.A., Liberles, D.A., Roe, R.M., Vargo, E.L.,

959 Vilcinskas, A., Wang, J., Bornberg-Bauer, E., Korb, J., Zhang, G., Liebig, J., 2014. Molecular
 960 traces of alternative social organization in a termite genome. *Nat. Commun.* 5, 3636.
 961
 962 Tingey, W.M., Pillimer, E.A., 1977. *Lygus* bugs: crop resistance and physiological nature of
 963 feeding injury. *Bull. Entomol. Soc. Am.* 23, 277-287.
 964
 965 Toullec, J.Y., Corre, E., Bernay, B., Thorne, M.A., Cascella, K., Ollivaux, C., Henry, J., Clark,
 966 M.S., 2013. Transcriptome and peptidome characterisation of the main neuropeptides and
 967 peptidic hormones of a euphausiid: the Ice Krill, *Euphausia crystallorophias*. *PLoS One* 8,
 968 e71609.
 969
 970 Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B.C., Remm, M., Rozen, S.G.,
 971 2012. Primer3 - new capabilities and interfaces. *Nucleic Acids Res.* 40, e115.
 972
 973 Van Hiel, M.B., Van Loy, T., Poels, J., Vandersmissen, H.P., Verlinden, H., Badisco, L., Vanden
 974 Broeck, J., 2010. Neuropeptide receptors as possible targets for development of insect pest
 975 control agents. *Adv. Exp. Med. Biol.* 692, 211-226.
 976
 977 Veenstra, J.A., 2000. Mono- and dibasic proteolytic cleavage sites in insect neuroendocrine
 978 peptide precursors. *Arch. Insect. Biochem. Physiol.* 43, 49-63.
 979

980 Ventura, T., Cummins, S.F., Fitzgibbon, Q., Battaglione, S., Elizur, A., 2014. Analysis of the
 981 central nervous system transcriptome of the eastern rock lobster *Sagmariasus verreauxi* reveals
 982 its putative neuropeptidome. PLoS One 9, e97323.
 983

984 Weaver, R.J., Audsley, N., 2008. Neuropeptides of the beetle, *Tenebrio molitor* identified using
 985 MALDI-TOF mass spectrometry and deduced sequences from the *Tribolium castaneum* genome.
 986 Peptides 29, 168-178.
 987

988 Wheeler, A.G., 2001. Biology of the plant bugs (Hemiptera: Miridae): pests, predators,
 989 opportunists. Comstock Publishing Associates.
 990

991 Xie, Y., Zhang, L., Zhang, C., Wu, X., Deng, X., Yang, X., Tobe, S.S., 2015. Synthesis,
 992 biological activity, and conformational study of N-methylated allatostatin analogues inhibiting
 993 juvenile hormone biosynthesis. J. Agric. Food Chem. 63, 2870-2876.
 994

995 Xu, G., Gu, G.-X., Teng, Z.-W., Wu, S.-F., Huang, J., Song, Q.-S., Ye, G.-Y., Fang, Q., 2016.
 996 Identification and expression profiles of neuropeptides and their G protein-coupled receptors in
 997 the rice stem borer *Chilo suppressalis*. Sci. Rep. 1–15.
 998

999 Yan, X.C., Chen, Z.F., Sun, J., Matsumura, K., Wu, R.S., Qian, P.Y., 2012. Transcriptomic
 1000 analysis of neuropeptides and peptide hormones in the barnacle *Balanus amphitrite*: evidence of
 1001 roles in larval settlement. PLoS One 7, e46513.
 1002

1003 Zandawala, M., Paluzzi, J.-P., Orchard, I., 2011. Isolation and characterization of the cDNA
 1004 encoding DH31 in the kissing bug, *Rhodnius prolixus*. Mol. Cell. Endocrinol. 331, 79–88.
 1005
 1006 Zhang, Q., Nachman, R.J., Kaczmarek, K., Zabrocki, J., Denlinger, D.L., 2011.
 1007 Disruption of insect diapause using agonists and an antagonist of diapause hormone
 1008 Proc. Natl. Acad. Sci. USA 108, 16922–16926.
 1009
 1010 Zhang, C., Qu, Y., Wu, X., Song, D., Ling, Y., Yang, X., 2015a. Design, synthesis and aphicidal
 1011 activity of N-terminal modified insect kinin analogs. Peptides 68, 233-238.
 1012
 1013 Zhang, C., Qu, Y., Wu, X., Song, D., Ling, Y., Yang, X., 2015b. Eco-friendly insecticide
 1014 discovery via peptidomimetics: design, synthesis, and aphicidal activity of novel insect kinin
 1015 analogues. J. Agric. Food Chem. 63, 4527-4532.
 1016
 1017 Zhang, Q., Nachman, R.J., Kaczmarek, K., Kierus, K., Zabrocki, J., Denlinger, D.L., 2015.
 1018 Development of neuropeptide analogs capable of traversing the integument: A case study using
 1019 diapause hormone analogs in *Helicoverpa zea*. Insect Biochem. Mol. Biol. 67, 87–93.
 1020
 1021 Zhu, Y. C., Luttrell, R., 2012. Variation of acephate susceptibility and correlation with esterase
 1022 and glutathione S-transferase activities in field populations of the tarnished plant bug, *Lygus*
 1023 *lineolaris*. Pestic. Biochem. Physiol. 103, 202-209.
 1024
 1025

Figure legends

Figure 1. Two examples of the *in silico* workflow used for the prediction of putative mature *Lygus hesperus* peptide structures. **(A)** Predicted processing scheme for prepro-allatostatin C (AST-C). The structure of the putative mature AST-C isoform is shown in red, with those of two putative mature linker/precursor-related peptides shown in blue. In this panel, the inverted red bracket indicates a predicted disulfide bond between the two cysteine residues in the putative mature AST-C, while “Y_(SO₃H)” indicates the presence of a sulfated tyrosine residue in one of the linker/precursor-related peptides. **(B)** Predicted processing scheme for prepro-sulfakinin. The structures of two putative mature sulfakinin isoforms are shown in red, with those of two putative mature linker/precursor-related peptides shown in blue. In this panel, “Y_(SO₃H)” indicates the presence of sulfated tyrosine residues in the sulfakinin isoforms, while “pQ” indicates the presence of pyroglutamic acid in one of these two peptides.

Figure 2. Putative *Lygus hesperus* pre/preprohormones deduced from transcriptome shotgun assembly sequence data. **(A)** Prepro-allatostatin A. **(B)** Prepro-allatostatin B. **(C)** Prepro-allatotropin. **(D)** The carboxyl-terminal portion of pre-bursicon β. **(E1)** Prepro-CCHamide Ia. **(E2)** Prepro-CCHamide Ib. **(F)** Prepro-corazonin. **(G1)** Prepro-crustacean cardioactive peptide Ia. **(G2)** Prepro-crustacean cardioactive peptide Ib. **(G3)** Prepro-crustacean cardioactive peptide Ic. **(H1)** Prepro-crustacean hyperglycemic hormone Ia. **(H2)** Prepro-crustacean hyperglycemic hormone Ib. **(I)** Prepro-diuretic hormone 31. **(J)** Prepro-GSEFLamide. **(K)** Prepro-insulin-like peptide. **(L)** Prepro-myosuppressin. **(M1)** Pre-neuroparsin I. **(M2)** Pre-neuroparsin II. **(M3)** Pre-neuroparsin III. **(N)** Prepro-neuropeptide F. **(O)** Prepro-orcokinin. **(P)** Prepro-pyrokinnin. **(Q)** Prepro-short neuropeptide F. **(R)** Prepro-SIFamide. **(S1)** Prepro-tachykinin-related peptide

Ia. (**S2**) Prepro-tachykinin-related peptide Ib. In this figure, signal peptides are shown in gray, while all mono/dibasic cleavage loci are shown in black. For each sequence, the isoform(s) of the peptide for which the precursor is named is/are shown in red, with all linker/precursor related peptides shown in blue. The “+” symbol in **D** indicates the presence of additional, unknown, amino acid residues at the amino (N) terminus of pre-bursicon β . It should be noted that the prepro-ITP shown in **H2** is an N-terminally extended variant of the shorter precursor shown in **H1**. This N-terminal extension (highlighted in yellow) generates a situation in which the signal peptide is no longer predicted for the longer protein using SignalP, while one is predicted by this program for the shorter preprohormone. The extension may be the result of an artifact in the process of assembling the transcript encoding the protein in question, or may have true biological significance, *e.g.*, it may function as a potential regulatory element. An isoform of orcomyotropin has been colored green in **O**.

Figure 3. RT-PCR based amplification of select peptide precursor-encoding transcripts from mixed sex *L. hesperus* heads. (**A**) Image of a 2% agarose gel with amplimers of the expected sizes for each of six precursor-encoding transcripts: DH31, 324 nucleotides; sNPF, 276 nucleotides; pyrokinin/PBAN, 396, nucleotides; allatotropin, 366 nucleotides; allatostatin A, 651 nucleotides; orcokinin, 510 nucleotides; bursicon β , 405 nucleotides. All amplimers were subcloned and sequence verified. Splice variants were identified for DH31 and allatotropin. Asterisks indicate non-specific amplification products. A 500-bp fragment of *L. hesperus* actin was amplified as a positive control for cDNA template quality. (**B**) Sequence alignment of the two *L. hesperus* allatotropin preprohormone variants deduced from the PCR amplimers. The predicted signal peptide and cleavage sites are indicated by the green and orange boxes,

1072 respectively. The predicted allatotropin sequence is shown in the pink box. (C) Sequence
1073 alignment of the three *L. hesperus* DH31 preprohormone variants deduced from the PCR
1074 amplimers. The predicted signal peptide and cleavage sites are indicated by the green and orange
1075 boxes, respectively. The predicted DH31 sequence is shown in the pink box.

Table 1

Table 1. Putative <i>Lygus hesperus</i> peptide-encoding transcripts identified via <i>in silico</i> transcriptome shotgun assembly (TSA) mining						
Peptide target family	TSA transcript/protein identifications				tblastn search statistics	
	Transcript accession number	Transcript length*	Deduced protein length†	Deduced protein type‡	BLAST score	E-value
Adipokinetic hormone						
Adipokinetic hormone-corazonin-like peptide						
Allatostatin A	GBRD01001611	791	216	F	57.4	7e-08
	GBHO01034447	1059	216	F	57.4	1e-08
	GBHO01034445	1117	216	F	57.4	7e-08
Allatostatin B	GBRD01006049	1271	357	F	43.9	0.012
	GBRD01006050	1303	357	F	43.9	0.012
	GBHO01039135	1703	357	F	43.9	0.013
	GBHO01039136	1710	357	F	43.9	0.013
	GBHO01039134	1742	357	F	43.9	0.013
Allatostatin C	GBHO01019349	902	119	F	62.4	1e-09
Allatotropin	GBRD01018034	709	98	F	191	7e-60
Bursicon α						
Bursicon β	GBHO01000699	375	114	C	171	3e-51
CCHamide	GBHO01017144	1300	129	F	40.4	0.049
	GBHO01017145	1222	103	F	37.4	0.57
Corazonin	GBHO01027723	690	96	F	42.7	0.008
Crustacean cardioactive peptide	GBRD01016318	468	107	F	68.9	2e-12
	GBHO01014887	606	155	F	68.9	3e-12
	GBHO01014886	3929	130	F	69.7	2e-11
Crustacean hyperglycemic hormone/ion transport peptide	GBHO01006806	1535	117	F	99.8	2e-22
	GBHO01023508	1689	117	F	99.8	2e-22
	GBHO01023514	1624	117	F	99.8	2e-22
	GBHO01023513	1763	117	F	99.8	2e-22
	GBHO01006805	1698	117	F	99.8	2e-22
	GBRD01015340	2739	117	F	99.8	4e-22
	GBRD01015337	2805	117	F	99.8	4e-22
	GBRD01015333	2567	117	F	99.8	4e-22
	GBRD01015338	2633	117	F	99.8	4e-22
	GBHO01040138	1839	158	F	100	2e-22
	GBRD01015334	2586	158	F	100	2e-22
	GBRD01015339	2652	158	F	100	2e-22
	GBRD01015336	2815	158	F	100	3e-22
	GBRD01015335	2881	158	F	100	3e-22
Diuretic hormone 31	GBHO01007397	1544	111	F	78.2	5e-15

	GBHO01007401	2564	111	F	78.2	9e-15
Diuretic hormone 44						
Ecdysis-triggering hormone						
Eclosion hormone						
FMRFamide-like peptide						
GSEFLamide	GBHO01039466	544	145	F	45.1	0.002
Insulin-like peptide	GBHO01008344	606	150	F	46.6	2e-04
	GBRD01004401	617	150	F	46.6	2e-04
Intocin						
Leucokinin						
Myosuppressin	GBHO01020034	693	93	F	48.5	2e-05
Neuroparsin	GBHO01012600	1051	107	F	60.8	3e-09
	GBHO01026222	548	111	F	54.3	1e-07
	GBHO01027674	529	131	F	41.2	0.007
Neuropeptide F	GBHO01037799	1173	102	F	57.8	4e-08
Orcokinin	GBRD01007133	1033	169	F	89.0	2e-18
	GBHO01017219	1106	169	F	89.0	2e-18
Pigment dispersing hormone						
Proctolin						
Pyrokinin	GBHO01031681	640	131	F	41.6	0.022
RYamide						
Short neuropeptide F	GBHO01034190	840	91	F	62.8	3e-10
SIFamide	GBHO01035740	4096	72	F	80.9	3e-16
	GBHO01009864	4159	72	F	80.9	3e-16
	GBHO01011321	4750	72	F	80.9	4e-16
	GBHO01008283	4813	72	F	80.9	4e-16
Sulfakinin	GBHO01022480	607	93	F	59.3	5e-09
Tachykinin-related peptide	GBHO01042633	1358	192	F	71.2	1e-11
	GBHO01042634	1329	217	F	70.9	1e-11

*Length in nucleotides

†Length in amino acids

‡Deduced protein type: F, full-length; C, carboxyl-terminal partial.

Query proteins: adipokinetic hormone (AKH), *Drosophila melanogaster* prepro-AKH (**AAF47846**; Adams et al., 2000) and *Nilaparvata lugens* prepro-AKH (**BAO00932**; Tanaka et al., 2014); adipokinetic hormone-corazonin-like peptide (ACP), *Tribolium castaneum* prepro-ACP (**NP_001159497**; Hansen et al., 2010) and *N. lugens* prepro-ACP (**BAO00933**; Tanaka et al., 2014); allatostatin A (AST-A), *D. melanogaster* prepro-AST-A (**AAF56331**; Adams et al., 2000); allatostatin B (AST-B); *D. melanogaster* prepro-AST-B (**AAF49354**; Adams et al., 2000); allatostatin C (AST-C), *Hyalomma azteca* prepro-AST-C III (Christie et al., 2016); allatotropin, *Manduca sexta* prepro-allatotropin (**AAB08758**; Taylor et al., 1996); bursicon α , *D. melanogaster* pre-bursicon α (**AAF55915**; Adams et al., 2000); bursicon β , *D. melanogaster* partner of bursicon (**AAF53396**; Adams et al., 2000); CCHamide, *D. melanogaster* prepro-CCHamide-2 (**AAF54942**; Adams et al., 2000); corazonin, *D. melanogaster* prepro-corazonin (**AAF55046**; Adams et al., 2000); crustacean cardioactive peptide (CCAP), *D. melanogaster*

prepro-CCAP, isoform A ([AAF56045](#); Adams et al., 2000); crustacean hyperglycemic hormone/ion transport peptide (ITP), *D. melanogaster* prepro-ITP, isoform C ([ABI31114](#); Adams et al., 2000); diuretic hormone 31, *D. melanogaster* prepro-DH31, isoform A ([AAF52685](#); Adams et al., 2000); diuretic hormone 44 (DH44), *D. melanogaster* prepro-DH44 ([AAF54421](#); Adams et al., 2000) and *N. lugens* prepro-CRF-related diuretic hormone 45 ([BAO00945](#); Tanaka et al., 2014); ecdysis-triggering hormone (ETH), *D. melanogaster* prepro-ETH ([AAF47275](#); Adams et al., 2000) and *N. lugens* prepro-ETH ([BAO00949](#); Tanaka et al., 2014); eclosion hormone (EH), *N. lugens* pre-EH 1 ([BAO00950](#); Tanaka et al., 2014) and *N. lugens* pre-EH 2 ([BAO00951](#); Tanaka et al., 2014); FMRFamide-like peptide (FLP); *D. melanogaster* prepro-FLP ([AAF58874](#); Adams et al., 2000) and *N. lugens* prepro-FLP ([BAO00954](#); Tanaka et al., 2014); GSEFLamide, *Tigriopus californicus* prepro-GSEFLamide Ia (Christie, 2014b); insulin-like peptide (ILP), *D. melanogaster* prepro-ILP 2 ([AAF50204](#); Adams et al., 2000); intocin, *T. castaneum* prepro- vasopressin-like peptide ([ABX52000](#); Stafflinger et al., 2008); leucokinin, *D. melanogaster* prepro-leucokinin ([AAF49731](#); Adams et al., 2000) and *N. lugens* prepro-kinin ([BAO00961](#); Tanaka et al., 2014); myosuppressin, *D. melanogaster* prepro-myosuppressin, isoform A ([AAF56283](#); Adams et al., 2000); neuroparsin, *Schistocerca gregaria* pre-neuroparsin 1 ([CAC38869](#); Janssen et al., 2001); neuropeptide F (NPF), *N. lugens* prepro-NPF ([BAO00965](#); Tanaka et al., 2014); orcokinin, *N. lugens* prepro-orcokinin ([BAO00969](#); Tanaka et al., 2014); pigment dispersing hormone, *N. lugens* prepro-pigment dispersing factor ([BAO00970](#); Tanaka et al., 2014); proctolin, *N. lugens* prepro-proctolin ([BAO00972](#); Tanaka et al., 2014); pyrokinin, *N. lugens* prepro-pyrokinin ([BAO00974](#); Tanaka et al., 2014); RYamide, *N. lugens* prepro-RYamide ([BAO00975](#); Tanaka et al., 2014); short neuropeptide F (sNPF), *N. lugens* prepro-sNPF ([BAO00976](#); Tanaka et al., 2014); SIFamide, *N. lugens* prepro-SIFamide ([BAO00977](#); Tanaka et al., 2014); sulfakinin, *N. lugens* prepro-sulfakinin ([BAO00978](#); Tanaka et al., 2014); tachykinin-related peptide (TRP), *N. lugens* prepro-TRP ([BAO00979](#); Tanaka et al., 2014).

Table 2

Table 2. Predicted peptidome of <i>Lygus hesperus</i>	
Peptide group	Predicted peptide structure
Allatostatin A	LPVYNFGLa LYSFGLa MKQYSFGLa APKQYSFGLa AEGKHWNFGVa SMMYNFGLa
Allatostatin A linker/precursor-related peptide	SADNDQPLDYKLYEGDTGLS AAYSIVSEY SPGDE SGDAEDDY _(SO3H) EDEIKDEY _(SO3H) N APYDHEPRLYEDY _(SO3H) EAPDEV AGENa DDEQRQa SDGPQPEKDSSSSSLLVN
Allatostatin B	AWNMQATWa KWQDMQNPWa GWKDMQGPSWa GWQDMRQPSWa GWKDMQTSAWa GWQDMQGQWa GWQDMQGNWa GWQDMQGLNWa WQDMPTSAWa GWKELQTTGWa AWNDIQSSGWa GWKEMPSTGWa
Allatostatin B linker/precursor-related peptide	YDNNKVRSLPD'TNEEDRDAKVV STECGNTGWQDPKSDSE TPTPPEFDKDIQASLQTSFQED DALEDQMAD GWGDMKISGYa SPQDHEDFDDLEEDWELPSKSEVDIEDLLSAAEWPLGKEEGANRND HLYEPLDDQALFGKILEGLYKASWTNRNFNPNKLAEEAGTSAQRN
Allatostatin C	GQKNGR ¹ TYWRCYFNAVTC ² F
Allatostatin C linker/precursor-related peptide	

	SPTSRTSDDQTPDLANSQISYDDY _(SO3H) PVVVP AALLLDRLMVALEKAVESEAAEKGMTNRI PQDDTWKMDLQ
Allatotropin	GLKNGPLNSARGFa
Allatotropin linker/precursor-related peptide	SAPQMLISRYYDRSRST SPHTSEETNVPIEWMADEMASNPELARLMVRQFIDVDRDGF
Bursicon β +NDEACETLPSEIHI IKEEFDELGRLSRTCNGDIAVNKCEGACTSQVQPSVITPTGFLKECYCCRESFLRERIVTLTHCYDPDGVRLEKEGVATMDIKLKEPSDCKCFKCGDYSR	
CCHamide	GCSAFGHSCFGGHa GGCSAFGHSCFGGHa
CCHamide linker/precursor-related peptide	SPYAGDRDFDVPF PL GSFQGERLLLVEPFDSSPIETASAF SDSSSLMEPQLQQAEQRLHQRLPNTDLFKQWLQAYRGSPSLLDGQ
Corazonin	pQTFQYSRGWTNa
Corazonin linker/precursor-related peptide	SGLSPKDTACELHRVKSLLLEGRPITSYFWPCNYRQFLDANGRILSPVEDLSKIPIEISDEQK
Crustacean cardioactive peptide	PFCNAFTGCa
Crustacean cardioactive peptide linker/precursor-related peptide	DDVILQ YYYPGPGEVAEPEPKM SDESMAALVDLSSEPAVEDLSRQILSEAKLWEALQEARMELINRKHQL SDESMAALVDLSSEPAVEDLSRQILSEAKLWEALQEARMELINRKHQTVPWFLDEDQPIHTWRGAPIILD CYSEPGLDSDLILNSICSNGANFTSS SDESMAALVDLSSEPAVEDLSRQILSEAKLWEALQEARMELINRKHQGD RMQQQLPLPLTGI SQYVYT
Crustacean hyperglycemic hormone/molt-inhibiting hormone	SFFDLQCKGVYDKSIFARLDRICEDCYNLFREPSLHSLCRKNCFTTDYFKGCIDVLLLQDEMEHIQSWIKQLHGADANL
Crustacean hyperglycemic hormone precursor-related peptide	GPASRIIMGHPLN
Diuretic hormone 31	GLDLGLSRGFSGSQAAKHLMGLAAANYAGGPa
Diuretic hormone 31 linker/precursor-related peptide	QESPYPRHRSYLGE LSDPD AEVMLEILAKLGQTIMRANDLEKYS LTS HV
GSEFLamide	

GSEFLamide linker/precursor-related peptide	<p>TVGTEFLa</p> <p>IDLSRLYGHVAV</p> <p>NSEACHPYEPFKCPGDGACISIQYLCDGAPDCLDGYDEDSRLCTAAVYFD</p> <p>NRWA</p> <p>CSSPGSDCDa</p> <p>LPDY_(S03R)KS</p> <p>YPPI</p>
Insulin-like peptide	<p>SHQRYCGRDINEVLKLICHGVNSMFTN</p> <p>GVYDECCKKTC_SINELASYCGVN</p>
Insulin-like peptide linker/precursor-related peptide	<p>SYPAEAFa</p> <p>NGGDQEIRDFGEYQVFPNE</p> <p>EDLTRPADMDAFVAAIQNLRYPFASKATSLTMLPRTNGIYPRE</p>
Myosuppressin	
Myosuppressin linker/precursor-related peptide	<p>pQDVDHVFLRFa</p> <p>APGPDCSATTIQDLPPRVRTMCAAFYQLSNALQTYIDEQS_{NY}QPLVREGNQIYETGV</p>
Neuroparsin	<p>SFYPCRPCVGDECDLEPEDCKYGTARDPCNRLICAAGPGERCGGRDNHIGKCGEGMNCRCGTCRG_CSTVRFLQGFIDCEWNHHMCNS</p> <p>CKNYCLDESCRIDMQDGPCLDGEVEVCGR_TVC_AKGPGESCGGIDSWRGKCGRGMHCQSGRCSGCSTTRYLN_GAKPEEACRKWDKGDMLSGR</p> <p>pQRYRSIIAPLSMTCQELCVGPGCNTEPEVEECSEWGHTKICGRRICYKGPDEVCGMPYNRYGLCAENLHCHHGGSII_GKCKGCVYVNGVSNCSHYEHSFTDPRSHHHYRY</p>
Neuropeptide F	
Neuropeptide F linker/precursor-related peptide	<p>PKTFDSPDDLRTYLNQLGQYYAVAGRPRFa</p> <p>QMPADAMARPA</p> <p>TSSRLHLPLIEGLSPRYPYNDPSELYDLLYQPLAD</p>
Orcokinin	<p>NFDEIDRAGFRSFN</p> <p>NFDEIDRAGFTSFA</p> <p>PMDEIDRAGFTRFQ</p> <p>PMDEIDRAGFTHFV</p> <p>PMDEIDRAGFDHFV</p>
Orcokinin linker/precursor-related peptide	<p>LPTPELQRNAYDDFGMANYYDRLS</p> <p>QAKARQYVRNLDPLSGSTFGTS</p> <p>EVDELPA_{TQH}</p>
Orcomyotropin	<p>FDSLNGGTFGAD</p>
Pyrokinin	<p>NSASFQPRSa</p>

Pyrokinin linker/precursor-related peptide	DEESQFTETSRSPPFAPRLa MEEGGRDM SVGNNFGGGSWSQDAHTNGNEFNGDEAIADLLKVSPWTFYANQGKLTFARES VLHYSPRF ITLN
Short neuropeptide F	DNRNPQFRLRFa
Short neuropeptide F linker/precursor-related peptide	SPTSLDY _(SO3H) DSVRDLYEGLLKASHPDWDHRLV SAPSIPEDHRMEPS MDY _(SO3H) F EK
SIFamide	PPFNGSIFa
SIFamide linker/precursor-related peptide	TGPPEYEQAGKALSTMCEIASEACSAWFPTQDNN
Sulfakinin	pQFN DY _(SO3H) GHMRFa SGGDEKFDDY _(SO3H) GYMRFa
Sulfakinin linker/precursor-related peptide	SPAPRPYP RNP HHLPRAIELNFQDDDDALE GFV
Tachykinin-related peptide	APAMGFQGM Ra APVMGFQGM Ra APQMGFSAM Ra pEDYGFWAGEEDHPGFYDPRa APSGFFGM Ra VPSAFFGM Ra GPSGFMGV Ra
Tachykinin-related peptide linker/precursor-related peptide	QERHPMDFIGSQVKMDGQPPAPDIEEF DDGLATLSADFDIEEF DGADDLDALMQIIRESAS QGGDSYHV QEMEDMLASRV FAVPTGLANPYQRPNDLELV
Abbreviations: a, amidated C-terminus; +, likely partial peptide with additional unknown amino acids; pQ, pyroglutamic acid; <u>C</u> , disulfide bonded cysteine (see text for descriptions of bonding partners in peptides with more than one bridged cysteine); Y _(SO3H) , sulfated tyrosine.	
Prediction of sulfation state of tyrosine residues and disulfide bridging between cysteines was done only for full-length peptides.	

Table 3

Table 3. Oligonucleotide primers used to amplify select <i>L. hesperus</i> neuropeptide precursor-encoding transcripts		
Precursor	Primers	
	Sense (5'-3')	Antisense (5'-3')
sNPF	atgaaagtggctttaaccttcg	ttattttcaaagtagtcattctcga
Orcokinin	atgttgacgctcgctattct	ctaatactgggtggcgg
Pyrokinin	atggtaacctgacagcg	ttaattcaatgttattctgcgaaacc
AST-A	atggttacttcaggaaagggc	ctaattgacgaggagcgacg
DH31	atggtcacaacgccgt	ttacacgtgccgcctcc
Bursicon β	atggaaatcgttccttagcc	tcattctagaataatctccacatttga
Allatotropin	atgcagtgggtacgaacg	ctaattgtccgtctggtcg
Actin	atgtgcgacgaagaagttg	gtcacggccagccaaatc
Abbreviations: sNPF, short neuropeptide F; AST-A, allatostatin A; DH31, diuretic hormone 31.		

Figure 1

A. Prepro-allatostatin C (from GBHO01019349)

MLESHGNLRQLLCWWLVLTALTFSLSA**CS**PTSRTSDDQTPDLANSQISYDDYPVVVPKRAALLDRLMVALEKAVESEAAEKGMTNRIPQDDTWKMDLQRRGQKNGRTYWRCYFNAVTCF

↓ Signal peptidase (cleavage locus highlighted above)

SPTSRTSDDQTPDLANSQISYDDYPVVVPK**RA**AALLDRLMVALEKAVESEAAEKGMTNRIPQDDTWKMDLQ**RG**QKNGRTYWRCYFNAVTCF

↓ Prohormone convertase (cleavage loci highlighted above)

SPTSRTSDDQTPDLANSQISYDDYPVVVP**KR** AALLDRLMVALEKAVESEAAEKGMTNRIPQDDTWKMDLQ**RR** GQKNGRTYWRCYFNAVTCF

↓ Carboxypeptidase (cleavage loci highlighted above)

SPTSRTSDDQTPDLANSQISYDD**Y**PVVVP AALLDRLMVALEKAVESEAAEKGMTNRIPQDDTWKMDLQ GQKNGRTYWRCYFNAVTCF

↓ Tyrosylprotein sulfotransferase (tyrosine target highlighted above)

SPTSRTSDDQTPDLANSQISYDD**Y**_(S03H) PVVVP GQKNGRTYWR**C**YFNAV**T**C**F**

↓ Enzymatic disulfide bond formation (bridged cysteines highlighted above)

GQKNGRTYWRCYFNAV**T**C**F**

B. Prepro-sulfakinin (from GBHO01022480)

MSRHVMTTILLTVGVYLLQLNTSRLCA**GS**PAPRPYPNPNPHLPRAIELNFQDDDDALEKRQFN DYGHMRF GKRSGGDEKFDDYGYMRFGRGFV

↓ Signal peptidase (cleavage locus highlighted above)

SPAPRPYPNPNPHLPRAIELNFQDDDDALEK**RQ**FN DYGHMRF GKR**S**GGDEKFDDYGYMRF G**R**GFV

↓ Prohormone convertase (cleavage loci highlighted above)

SPAPRPYPNPNPHLPRAIELNFQDDDDALE**KR** QFN DYGHMRF G**KR** SGGDEKFDDYGYMRF G**R** **GFV**

↓ Carboxypeptidase (cleavage loci highlighted above)

SPAPRPYPNPNPHLPRAIELNFQDDDDALE QFN DYGHMRF **G** SGGDEKFDDYGYMRF **G**

↓ Peptidylglycine-α-amidating monooxygenase (amidation loci highlighted above)

QFN**DY**GHMRFamide SGGDEKFDD**Y**GYMRFamide

↓ Tyrosylprotein sulfotransferase (tyrosine targets highlighted above)

QFN**DY**_(S03H)GHMRFamide SGGDEKFDD**Y**_(S03H)GYMRFamide

↓ Glutaminyl cyclase (cyclization locus highlighted above)

pQFND**Y**_(S03H)GHMRFamide

A. Prepro-allatostatin A (from GBRD01001611, GBHO01034447 and GBHO01034445)

MVTSGKGMILHVVALVLEIAGQVTA**SADNDQPLDYKLYEGDTGLSKRAAYSIVSEYKRLPVYNFGLGKRSPGDERRLYSFGLGKRSGDAEDDYED**
EIKDEYNKRMKQYSFGLGKRAPKQYSFGLGKRAEGKHNFGVGKRAPYDHEPRLYEDYEAPDEVKRAGENGRLYSFGLGKRDEQRQGRSMMYNFG
LGKRSDGPQPEKDSSSSSLLVN

B. Prepro-allatostatin B (from GBRD01006049, GBRD01006050, GBHO01039135, GBHO01039136 and GBHO01039134)

MLQLIITILVGLATLACTQC**YDNNKVRSLPDTNEEDRDAKVVKRSTECNGTGWQDPKSDSEKRAWNNMQATWGKRKWQDMQNPWGKRTPPTPEFDK**
DIQASLQTSFQEDKRGWKDMQGPSWGKRGWQDMRQPSWGKRGWKDMQTSAWGKRGWQDMQGWGKRGWQDMQGNWGKRGWQDMQGLNWGKRWRQDMP
TSAWGKRGWKELQTTGWGKRDALEDQMDKRGWDMKISGYGKRAWNDIQSSGWGKRGWKEMPSTGWGKRSPQDHEDFDDLEEDWELPSKSEVDIED
LLSAAEWPLGKEEGANRNRHLYEPLDDQALFGKILEGLYKASWTNRNFPNKNLAEAGTSAQRNR

C. Prepro-allatotropin (from GBRD01018034)

MQWVRTSVCLIVLIALMYETVHA**SAPQMLISRYDRSRSTRGLKNGPLNSARGFGKRSPHTSEETNVPIEWMADEMASNPELARLMVRQFIDVDRDG**
F

D. Pre-bursicon β (C-terminal partial protein from GBHO01000699)

+NDEACETLPSEIHIIKEEFDELGRLSRTCNGDIAVNKCEGACTSQVQPSVITFTGFLKECYCCRESFLRERIVTLTHCYDPDGVRLEKEGVATMDI
KLKEPSDCKCFKCGDYSR

E1. Prepro-CCHamide Ia (from GBHO01017144)

MQQATVLAALLAVVVICSVHSSPYAGDRDFDVFPF**RKKPLRRGSFQGERLLLVFPDSSPIETASAFKRGC****SAFGHSCFGGHGKRSDSSSLMEPQLQQ**
AEQRLHQRLPNTDLFKQWLQAYRGSPSLDQG

E2. Prepro-CCHamide Ib (from GBHO01017145)

MQQATVLAALLAVVVICSVHSSPYAGDRDFDVFPF**RKKPLRRGG****CSAFGHSCFGGHGKRSDSSSLMEPQLQQA****EQRLHQRLPNTDLFKQWLQAYRGSP**
SLDQG

F. Prepro-corazonin (from GBHO01027723)

MGLHLTFALVVCVVSSALG**QTFQYSRGWTNGKRSGLSPKDTACELHRVKSLEGRPITSYFWPCNYRQFLDANGRLSPVEDLSKIPIEISDEQK**

G1. Prepro-crustacean cardioactive peptide Ia (from GBRD01016318)

MQLIYSCLLPALFSVIYA**DDVILQKRYYPGPGEVAEPLPKMKKPF****CNAFTGCGKKRSDESMAALVDLSSEPAVEDLSRQILSEAKLWEALQEAR**
MELINRKHQL

G2. Prepro-crustacean cardioactive peptide Ib (from GBHO01014887)

MQLIYSCLLPALFSVIYA**DDVILQKRYYPGPGEVAEPLPKMKKPF****CNAFTGCGKKRSDESMAALVDLSSEPAVEDLSRQILSEAKLWEALQEAR**
MELINRKHQTPVWFLEDDQPIHTWRGAPIILDCYSEGLDSDLILNSICNGANFTSS

G3. Prepro-crustacean cardioactive peptide Ic (from GBHO01014886)

MQLIYSCLLPALFSVIYA**DDVILQKRYYPGPGEVAEPLPKMKKPF****CNAFTGCGKKRSDESMAALVDLSSEPAVEDLSRQILSEAKLWEALQEAR**
MELINRKHQGDRLMQQLPLPLTGIRKRSQYVYT

H1. Prepro-crustacean hyperglycemic hormone Ia (from GBHO01006806, GBHO01023508, GBHO01023514, GBHO01023513, GBHO01006805, GBRD01015340, GBRD01015337, GBRD01015333, GBRD01015338)

MYQERRAVASLVVASTLLTCLLAG**PASRIIMGHPLNKR****SFFDLQCKGVYDKSIFARLDRICEDCYNLFREPSLHSLCRKNCFTTDYFKGCIDVLLQ**
DEMEHIQSWIKQLHGADANL

H2. Prepro-crustacean hyperglycemic hormone Ib (from GBHO01040138, GBRD01015334, GBRD01015339, GBRD01015336, and GBRD01015335)

MRRTADARSKSSTPEPKRVGLALHIGEFRRRWYWIPIRKKMYQERRAVASLVVASTLLTCLLAG**PASRIIMGHPLNKR****SFFDLQCKGVYDKSIFAR**
LDRICEDCYNLFREPSLHSLCRKNCFTTDYFKGCIDVLLQDEMEHIQSWIKQLHGADANL

I. Prepro-diuretic hormone 31 (from GBHO01007397 and GBHO01007401)

MVHNNAVLTVTVSLVMGTLLLSAA**QESPYPRHRSYLGE LDSDPDAEVMLEILAKLGQTIMRANDLEKYSILTSKRGLDLGLSRGFSGSQA****AAKHLMGLAA**
ANYAGGPGRRRRHV

J. Prepro-GSEFLamide (from GBHO01039466)

MSCSHSALVLAAGVVCCLAWPHSVMA**IDLSRLYGHVAVKR****NSEACHPYEPFKCPGDGACISIQYLCDGAPDCLDGYEDSRLCTAAVYFDKRN****RWAR**
KKCSSPGSDCDGKR**TVGTEFLGKRRLPDYKSKR****TVGTEFLGKR****RYPPPI**

K. Prepro-insulin-like peptide (from GBHO01008344 and GBRD01004401)

MPSQSIYVFLSLSALMCNALSSYPAEAFGKRSHQRYCGRDINEVLKLIICHGVYNSMFTNKRNGGDQEIIRDFGEYQVFPNERREDLTRPADMDAFVAAIQNLRYPFASKATSLTMLPRTNGIYPREKRGVYDECCCKTCSINELASYCGVN

L. Prepro-myosuppressin (from GBHO01020034)

MTCAWMLACLVLGSAVSLSAPGPDCSATTIQDLPPRVRTMCAAFYQLSNALQTYIDEQSNYQPLVREGNQIYETGVKRDVDHVFLRFGRRR

M1. Pre-neuroparsin I (from GBHO01012600)

MSWVSCFLFVVMITFLGVQG SFYPCRPCVGDECDLEPEDCKYGTARDPCNRLICAAGPGERCGGRDNHIGKCGEGMNCRCGTCRGCSTVRFLQGFI DCEWNHMCNS

M2. Pre-neuroparsin II (from GBHO01026222)

MNNVGSALLLLFSLAGLARA CKNYCLDESCRIDMQDGPCLDGEVEVCGRTVCAKGPGESCGGIDSWRGKCGRMHCQSGRCSCGSTTRYLNGAKPEEACRKWDKGDMLSGR

M3. Pre-neuroparsin III (from GBHO01027674)

MRSATAVFLVAVAAIATTS CQYRSIIAPLSMTCQELCVGPCNTEPEVEECSEWGHTKICGRRICYKGPDEVCGMPYNRYGLCAENLHCHHGGS IIGKCKGCVYVNGVSNCSHYEHSFTDPRSHHHYR

N. Prepro-neuropeptide F (from GBHO01037799)

MNAWTSWLWACVVVFCLAGLA QMPADAMARPARPKTFDSPDDLRTYLNQLGQYAVAGRPRFGKR TSSRLHLPLIEGLSPRYPYNDPSELYDLLYQPLAD

O. Prepro-orcokinin (from GBRD01007133 and GBHO01017219)

MLTLAILALAAVSTSA LPTPELQRNAYDDFGMANYYDRLSRQAKARQYVRNLDPLSGSTFGTSKRFD SLNGGTFGADKRNFDEIDRAGFRSFNKKNFDEIDRAGFTSFAKRPMD EIDRAGFTRFQKRPMD EIDRAGFTHFVKRPMD EIDRAGFDHFVKREVD ELPATQH

P. Prepro-pyrokinin (from GBHO01031681)

MVNLTATLQAALFVLLLTVIHG MEEGGRDMR SVGNNGGGSWSQDAHTNGNEFNGDEAIA D L L K V S P W T F Y A N Q G K L T F A R E S R N S A S F Q P R S G R D E E S Q F T E T S R S P P F A P R L G R V L H Y S P R F R R I T L N

Q. Prepro-short neuropeptide F (from GBHO01034190)

MKVALTFVGCFLGTMIMLAPEAMS SPTS LDYDSVRDLYEGLLKASHPDWDHRLV R K D N R N P Q F R L R F G R R S A P S I P E D H R M E P S R M D Y F E K

R. Prepro-SIFamide (from GBHO01035740, GBHO01009864, GBHO01011321 and GBHO01008283)

MSRFLVASCLMVVLVLLILDVTFAYKK PPFNGSIFGKR TGPPEYEQAGKALSTMCEIASEACSAWFPTQDNN

S1. Prepro-tachykinin-related peptide Ia (from GBHO01042633)

MRMSLESSVVLACLVAVCLC QERHPMDFIGSQVKMDGQPPAPDIEEFKRAPAMGFQGMRGKKDDGLATLSADAFDIEEFKRAPVMGFQGMRGKKAPQMGFSAMRGKKEDYGFWAGEEDHPGFYDPRGKKAPSGFFGMRGKKVPSAFFGMRGKKGPSGFMGVRGKKDGADDLDALMQIIRESASRQGGDSYHV

S2. Prepro-tachykinin-related peptide Ib (from GBHO01042634)

MRMSLESSVVLACLVAVCLC QERHPMDFIGSQVKMDGQPPAPDIEEFKRAPAMGFQGMRGKKDDGLATLSADAFDIEEFKRAPVMGFQGMRGKKAPQMGFSAMRGKKEDYGFWAGEEDHPGFYDPRGKKAPSGFFGMRGKKVPSAFFGMRGKKGPSGFMGVRGKKDGADDLDALMQIIRESASRQEMEDMLASRVRRFAVPTGLANPYQRPNDLELV

Figure 3

Figure 3