Insect Biochemistry and Molecular Biology 114 (2019) 103227

Contents lists available at ScienceDirect

Insect Biochemistry and Molecular Biology

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

Evolutionary trends of neuropeptide signaling in beetles - A comparative analysis of Coleopteran transcriptomic and genomic data

Aniruddha A. Pandit[a], Shireen-Anne Davies[a], Guy Smagghe[b], Julian A.T. Dow[a,*]

[a] Institute of Molecular, Cell and Systems Biology, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow, G12 8QQ, UK
[b] Department of Plants and Crops, Faculty of Bioscience Engineering, Ghent University, Coupure Links 653, 9000, Ghent, Belgium

ARTICLE INFO

Keywords:
Coleoptera
Insect pest
Neuropeptide precursors
Mature peptides
Hormones
Bioinformatics
Phylogeny

ABSTRACT

Insects employ neuropeptides to regulate their growth & development, behaviour, metabolism and their internal milieu. At least 50 neuropeptides are known to date, with some ancestral to the insects and others more specific to particular taxa. In order to understand the evolution and essentiality of neuropeptides, we data mined publicly available high quality genomic or transcriptomic data for 31 species of the largest insect Order, the Coleoptera, chosen to represent the superfamilies' of the Adephaga and Polyphaga. The resulting neuropeptide distributions were compared against the habitats, lifestyle and other parameters. Around half of the neuropeptide families were represented across the Coleoptera, suggesting essentiality or at least continuing utility. However, the remaining families showed patterns of loss that did not correlate with any obvious life history parameter, suggesting that these neuropeptides are no longer required for the Coleopteran lifestyle. This may perhaps indicate a decreasing reliance on neuropeptide signaling in insects.

1. Introduction

The order Coleoptera from class Insecta, is by far the largest (most speciose) insect Order, and is categorized into four sub-orders spanning a long evolutionary period. Archostemata, and Myxophaga, the two smallest in terms of number of species, are the most ancient of the sub-orders appearing close to 240 million years ago (MYA) (Hunt et al., 2007). Adephaga and Polyphaga (roughly 90% of all beetle species) appeared between 225 and 200 MYA (Hunt et al., 2007). Beetles show a high diversity in terms of various parameters including age as described above; size, habitat, and even diet (Lawrence and Newton, 1982; Beutel and Leschen, 2005; Leschen et al., 2009, Mckenna et al., 2015). In terms of size, beetles could be as small as a under a couple of millimetres long, such as the Curculionoidean, *Hypothenemus hampei* to the large scarabs such as *Oryctes borbonicus* right up to the massive *Rhyzopertha dominica,* the adult of which has been known to grow above 5 cm long. While most beetles tend to be terrestrial, some members of Adephaga (either larvae or adults) are known to live in aquatic environments (Shull et al, 2001; Hunt et al., 2007). Coleopterans have diverse feeding preferences as well, ranging from the most common that is herbivorous, followed by saprophagous and xylophagous species while the ancient species preferred a predacious diet. Many species in the super families Cucujoidea, Curculionoidea and Elaeroidea have shown a polymorphic nature where the larvae and adults differ in their dietary preferences (Beutel and Leschen, 2005; Crowson, 2013; Leschen and Beutel, 2014). The diversity over the vast order of Coleopterans and their suborders, families etc. have lead to interesting evolutionary questions towards understanding the factors that may have either caused these differences or the effects of divergence over the millions of years of beetle evolution.

Neuropeptides are considered to be vital in the regulation of growth and development, reproduction, and water & ion homeostasis among other physiological processes (Alstein and Nassel, 2010; Nassel and Winther, 2010). With the advent of improved Next Generation Sequencing technologies it is now relatively commonplace to achieve the genome or transcriptome of any living organism in a short but accurate manner and more and more data is being made available globally, as and when they are obtained. Taking advantage of this DNA or mRNA data that has been made public has facilitated in the documentation of many novel genes involved in the various functionalities of the organisms. Neuropeptides and their corresponding GPCRs have now been identified in a myriad of insects including *Drosophila melanogaster* (Hewes and Taghert, 2001), *Anopheles gambiae* (Riehle et al., 2002), *Apis mellifera* (Hummon et al., 2006), the model beetle *Tribolium castaneum* (Li et al., 2008) , the rice stem borer *Chilo suppressalis* (Xu et al., 2016) and many others (Christie, 2008, 2015; Predel et al., 2010).

Abbreviations: CAPA, Capability; ACP, AKH/Corazonin related peptide

* Corresponding author.
E-mail address: Julian.Dow@glasgow.ac.uk (J.A.T. Dow).

https://doi.org/10.1016/j.ibmb.2019.103227
Received 5 June 2019; Received in revised form 30 July 2019; Accepted 21 August 2019
Available online 27 August 2019
0965-1748/

Table 1
Species under study with information on their taxonomic classification and their respective Sequence Read Archive (SRA) IDs 1 Neuropeptides already identified from the Genome of *D. ponderosae* were used 2 Neuropeptides already identified from the Genome of *T. castaneum* were used.

Species	Suborder	Super Family	Family	Subfamily	SRA ID	Tissue Sample
Gyrinus marinus	Adephaga	Gyrinoidea	Gyrinidae	Gyrininae	SRR921604	Whole Body
Carabus granulatus	Adephaga	Caraboidea	Carabidae	Carabinae	SRR596983	Head and Thorax
Amphizoa insolens	Adephaga	Dytiscoidea	Amphizoidae	–	SRR5930489	Whole Body
Rhyzopertha dominica	Polyphaga	Bostrichoidea	Bostrichidae	Dinoderinae	SRR1130516	Antennae
Anoplophora glabripennis	Polyphaga	Chrysomeloidea	Cerambycidae	Lamiinae	SRR1799851, SRR1799852	Whole Body
Callosobruchus maculatus	Polyphaga	Chrysomeloidea	Chrysomelidae	Bruchinae	SRR3113351	Whole Body
Chrysomela populi	Polyphaga	Chrysomeloidea	Chrysomelidae	Chrysomelinae	SRR1014805	Whole Body
Leptinotarsa decemlineata	Polyphaga	Chrysomeloidea	Chrysomelidae	Chrysomelinae	SRR2556962	Whole Body
Diabrotica virgifera	Polyphaga	Chrysomeloidea	Chrysomelidae	Galerucinae	SRR7818015, SRR7818014, SRR7818013	Whole Body
Cryptolaemus montrouzieri	Polyphaga	Cucujoidea	Coccinellidae	Scymninae	SRR343064	Whole Body
Propylea japonica	Polyphaga	Cucujoidea	Coccinellidae	Coccinellinae	SRR1299012	Whole Body
Harmonia axyridis	Polyphaga	Cucujoidea	Coccinellidae	Coccinellinae	SRR1283228	Whole Body
Coccinella septempunctata	Polyphaga	Cucujoidea	Coccinellidae	Coccinellinae	ERR1145724, ERR1145725, ERR1145726, ERR1145727, ERR1145728, ERR1145729	Whole Body
Adalia bipunctata	Polyphaga	Cucujoidea	Coccinellidae	Coccinellinae	ERR1145718, ERR1145719, ERR1145720, ERR1145721, ERR1145722, ERR1145723	Whole Body
Aethina tumida	Polyphaga	Cucujoidea	Nitidulidae	Nitidulinae	SRR1798556	Whole Body
Cylas brunneus	Polyphaga	Curculionoidea	Brentidae	Cyladinae	SRR3113351	Whole Body
Cylas puncticollis	Polyphaga	Curculionoidea	Brentidae	Cyladinae	SRR1611772	Whole Body
Cylas formicarius	Polyphaga	Curculionoidea	Brentidae	Cyladinae	SRR3065078	Whole Body
Hylobius abietis	Polyphaga	Curculionoidea	Curculionidae	Molytinae	SRR6765939, SRR6765940, SRR6765941	Whole Body
Dendroctonus ponderosae[1]	Polyphaga	Curculionoidea	Curculionidae	Scolytinae	–	Whole Body
Hypothenemus hampei	Polyphaga	Curculionoidea	Curculionidae	Scolytinae	SRR2163439	Whole Body
Tribolium castaneum[2]	Polyphaga	Tenebrionoidea	Tenebrionidae	Tenebrioninae	–	Whole Body
Tenebrio molitor	Polyphaga	Tenebrionoidea	Tenebrionidae	Tenebrioninae	SRR1291244	Whole Body
Agrilus planipennis	Polyphaga	Buprestoidea	Buprestidae	Agrilinae	SRR1615254, SRR16152523, SRR1615252, SRR1615251, SRR1609311, SRR1609066, SRR1604813	Whole Body
Chauliognathus flavipes	Polyphaga	Elateroidea	Cantharidae	Chauliognathinae	SRR4413771	Whole Body
Aspisoma lineatum	Polyphaga	Elateroidea	Lampyridae	Lampyrinae	SRR4407797	Whole Body
Photinus pyralis	Polyphaga	Elateroidea	Lampyridae	Lampyrinae	SRR6345452	Whole Body
Oryctes borbonicus	Polyphaga	Scarabaeoidea	Scarabaeidae	Dynastinae	SRR2970555	Whole Body
Protaetia brevitarsis	Polyphaga	Scarabaeoidea	Scarabaeidae	Cetoniinae	SRR7418790	Whole Body
Nicrophorus orbicollis	Polyphaga	Staphylinoidea	Silphidae	Nicrophorinae	SRR6286784	Whole Body
Aleochara curtula	Polyphaga	Staphylinoidea	Staphylinidae	Aleocharinae	SRR921563	Whole Body

The huge and diverse order of Coleoptera thus provides an ideal system in which to study evolutionary trends in neuropeptide adoption, or indeed tests the dogma of their essentiality. A recent preprint has performed an analysis of 17 beetle species (Veenstra, 2019); in this study, we performed comprehensive *in silico* neuropeptidomics on 31 members of the order Coleoptera which have been identified as economically significant pest species and supplemented with non-pest species for comparison and whose transcriptomic data has been made publicly available. Species were chosen to represent the major superfamilies of the key Coleoptera suborders (Table 1). From the raw sequence data available, we performed transcriptomic assemblies for each of these species, and using a Bioinformatics approach and pipelines created for the purpose, performed prediction and identification of genes of interest to create a matrix of all the neuropeptides precursors as well as mature neuropeptides found in each species. We also performed data analyses on these neuropeptidomes to identify significant ecological and evolutionary trends caused by the expression (or absence) of neuropeptides and/or hormones on the basis of their date of origin, habitat, diet and size.

2. Materials and methods

2.1. Transcriptomic and genomic read data

The aim of this study was to predict and identify neuropeptides from multiple representative species of different families and sub-families found in the Order Coleoptera and perform evolutionary trend analyses on this data. Species representing both the abundant sub-order Polyphaga as well as the more under-represented and older sub-order Adephaga were chosen. The 31 species (see Table 1), which have been studied in this project, are represented in the species tree shown in Fig. 1, which illustrates their estimated date of origin (in MYAs), along with the divergence of the various sub-orders & families. The RNA-seq datasets used in this study were obtained from the NCBI Sequence Read Archive or SRA (Leinonen et al., 2010) and were downloaded using the SRA Toolkit (https://trace.ncbi.nlm.nih.gov/Traces/sra/sra.cgi?view=software). For inclusion in the study, the tissues analysed (typically whole-body or head) were required to contain CNS (Table 1); interestingly an antennal transcriptome of one species (*Rhyzopertha dominica*) was included as the representative of the Bostrichidae, because it contained 34 neuropeptides. Preference was given to RNA-sequence raw data of adult whole tissue samples, but where other data (such as larval or other tissues) was available, it was utilized as well. The only exceptions to these were the sequence data for the well-studied insect model, the red flour beetle, *Tribolium castaneum* (Richards et al., 2008) and the mountain pine weevil *Dendroctonus ponderosae* (Keeling et al., 2013) where pre-sequenced and assembled genomic data was used for further analysis.

2.2. Transcriptome assembly

All the transcriptomic raw data that was downloaded from the SRA was taken through an assembly and analysis pipeline for neuropeptide prediction and identification. Quality assessment was done pre- and post-assembly using FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/created by Andrews, 2016). From the results of FastQC, any data that required removal of sequencing adaptors or removal of low-quality reads, duplicates, and any artefacts that may cause a reduction in the quality of the assembly was run through Cutadapt (Martin, 2011) and/or Trimmomatic (Bolger et al., 2014). This was

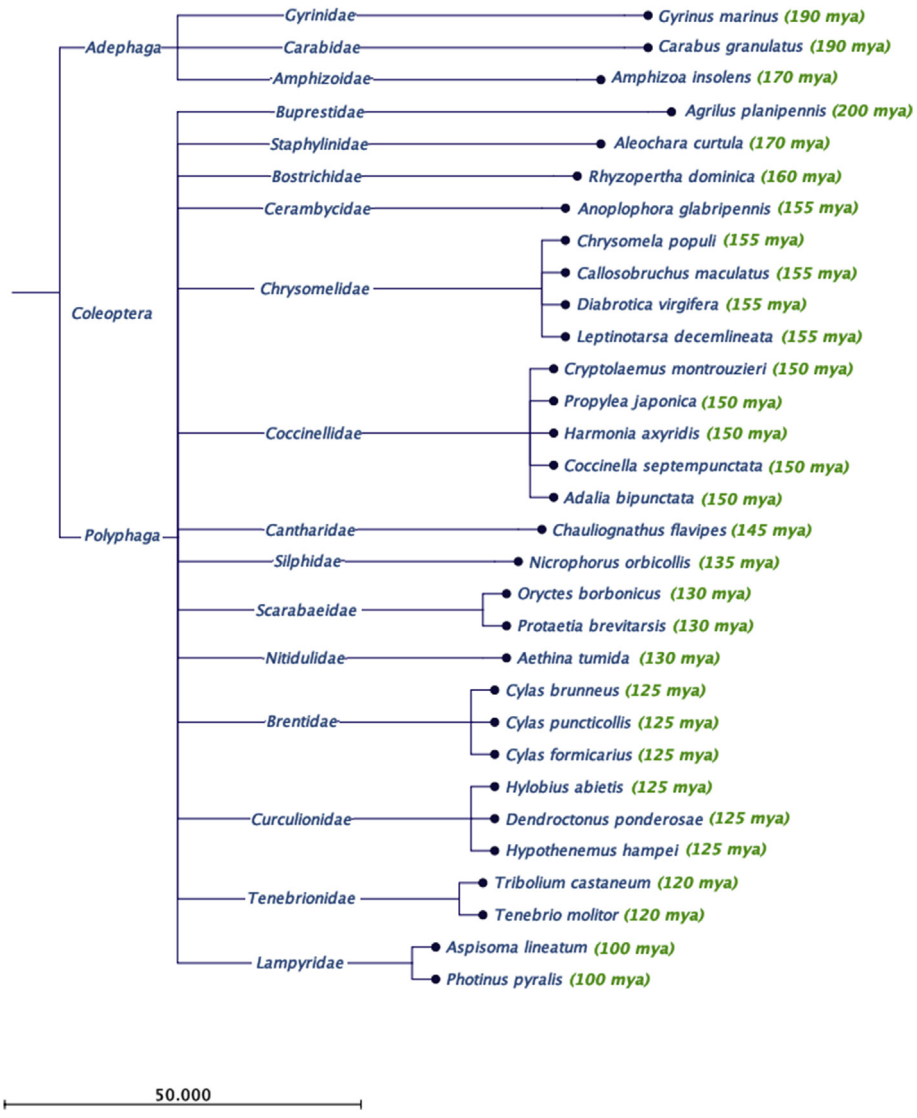


Fig. 1. A species tree illustrating all the species under study and their taxonomical nomenclature including their Order (Coleoptera), Suborders (Adephaga and Polyphaga) and their respective, underlying Families. The date of origin in the unit Million Years Ago (MYA) is denoted next to each species scientific names.

followed by digital normalization of the raw reads to increase the efficiency of the assembly by down-sampling the number of reads used in the assembly (Brown et al., 2012). This was done using the script for normalization included in the Trinity software suite (Grabherr et al, 2011) with the default settings including max coverage of 30. Following this, *de novo* transcriptome assembly was performed using the popular and robust software, Trinity version 2.2.0, that utilizes de Bruijn graphs for assembly of short reads using raw data from either paired-end sequencing or single strand sequencing. The assembly was run with default settings (Kmer 25) and adjusted for either paired end data or single strand data as per the protocol created by Haas et al. (2013). Prior to further analysis, BUSCO v3 (Benchmarking Universal Single-Copy Orthologs) was used to assess the completeness of the assembled transcriptome (Simao et al., 2015) using the insect set obtained from the OrthoDB v9 (Zdobnov et al., 2016).

2.3. Verification of data for use

Checks were carried out on suitability of the sequence data to make sure they were good enough for analyses and further neuropeptide prediction and identification. These checks were done at various stages including at the FastQC stage to check for quality of the sequencing, post assembly statistical checks to check for quality of the assembly and also using BUSCO to assess the completeness of the transcriptome. Any species whose data did not meet a “pass” criterion of quality from FastQC quality scores, and a high level of transcriptome completeness from the BUSCO analysis were discarded and other species were mined to complete all the possible family and sub-family options.

2.4. Neuropeptide precursor and mature peptide prediction

Neuropeptide precursor and mature peptide prediction and identification was done using the transcript/genomic sequences obtained from the assembly for each of the insect species following a prediction pipeline defined in Pandit et al. (2018). The transcript sequences were initially compared to precursor sequences found in the model Coleopteran *Triboleum castaneum* (Li et al., 2008), *D. ponderosae* (Keeling et al., 2013) and *Hylobius abietis* (Pandit et al., 2018) using BLASTx and tBLASTn (Altschul et al., 1990). The entire set of mature peptide sequence data for all insects available from DINeR (a Databaset for Insect Neuropeptide Research) [http://www.neurostresspep.eu/diner/infosearch] (Yeoh et al., 2017) were also downloaded. The transcript/genomic sequences from all the species under study were compared against the downloaded DINeR data using both BLASTx and tBLASTn to

Species	Complete BUSCOs (C)	Fragmented BUSCOs (F)	Missing BUSCOs (M)
Agrilus planipennis	98.8	0.5	0.7
Cylas brunneus	98.4	0.5	1.1
Aethina tumida	98.3	1.0	0.7
Photinus pyralis	97.9	1.0	1.1
Tenebrio molitor	97.8	1.1	1.1
Cylas puncticollis	97.7	0.5	1.8
Callosobruchus maculatus	97.6	0.8	1.6
Oryctes borbonicus	97.5	1.0	1.5
Protaetia brevitarsis	97.3	0.8	1.8
Anoplophora glabripennis	97.2	1.1	1.7
Nicrophorus orbicollis	96.9	1.0	2.1
Leptinotarsa decemlineata	96.9	1.3	1.8
Adalia bipunctata	96.9	1.3	1.8
Hylobius abietis	96.7	2.2	1.1
Coccinella septempunctata	96.4	1.3	2.3
Chrysomela populi	96.1	2.5	1.4
Harmonia axyridis	94.7	3.5	1.8
Hypothenemus hampei	93.3	3.9	2.8
Aleochara curtula	92.6	4.0	3.4
Diabrotica virgifera	91.9	3.9	4.3
Aspisoma lineatum	91.8	4.3	3.9
Chauliognathus flavipes	90.8	6.2	3.0
Amphizoa insolens	90.8	5.2	4.0
Cylas formicarius	89.7	8.1	2.2
Gyrinus marinus	82.1	9.0	8.8
Rhyzopertha dominica	80.8	15.5	3.7
Carabus granulatus	74.1	12.6	13.3
Cryptolaemus montrouzieri	69.4	21.2	9.4
Propylea japonica	58.4	31.7	9.8

Fig. 2. (a) BUSCO (Benchmarking Universal Single Copy Orthologs) analysis comparison of the 29 out of 31 species under study (*T. castaneum* and *D. ponderosae* were not included in this study). The table shows the percentage of complete, fragmented and missing BUSCOs when compared to the 1658 single-copy, annotated and conserved insect orthologs in the Insecta library (b) A comparison of the BUSCO scores of the species under study (denoted in Green font) with the scores of other insects whose scores have previously been calculated. The BUSCO score of each species was used in the verification of the assemblies and only those species with a satisfactory score were used in this study. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

help in the identification of putative neuropeptide precursors. The BLAST searches were done using the Galaxy Project (Afgan et al., 2016) server at the University of Glasgow (http://heighliner.cvr.gla.ac.uk/). Once a set of predicted precursor sequences for each species per neuropeptide was created, each sequence was run through the SignalP software (Peterson et al., 2011) for prediction of signal peptides. The mono and dibasic cleavage sites for the mature peptides were identified using the K and R combinations rules (Veenstra, 2000) as well as using the current version of the online software NeuroPred (Southey et al., 2006) available at http://stagbeetle.animal.uiuc.edu/cgi-bin/neuropred.py for determining putative mature peptides from the neuropeptide precursors. It should be noted that this approach identifies neuropeptide members of known families with widespread distribution across the insects; we did not attempt here to identify novel neuropeptides *ab initio*.

In addition to the above, the multiple sequence alignment tool, Clustal Omega (Sievers and Higgins, 2018) found at http://www.ebi.ac.uk/Tools/msa/clustalo/was used to generate sequence alignments and consensus generation while graphical representations of these multiple alignments in the form of sequence logos were generated using Weblogo 3.4 (Crooks et al., 2004).

2.5. Evolutionary trend analysis

The data for the various parameters studied under the evolutionary analyses including the divergence, habitat, diet, size, etc. were gathered from various websites, papers, books and databases created for this information. These include but are not limited to Encyclopedia of Life, available at http://eol.org; BugGuide, available at https://bugguide.net/; InsectaPro available at http://insecta.pro/and Wikipedia available at https://en.wikipedia.org/. Other papers and books have been listed in the References section. The heat maps were created in Microsoft Excel. Lineage data was obtained from the Taxonomy Browser available at https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi. The species tree was generated using the lineage information obtained from the NCBI Taxonomy Browser using the Common Tree generator functionality.

Multiple sequence alignments were performed using both Clustal Omega as well as the CLC Genomics Workbench 9 (https://www.qiagenbioinformatics.com/), which was also used for phylogenetic tree creation using the Neighbour Joining method with 1000 fold bootstrap resampling. Phylograms were created in the nexus file format for visualization in the CLC Genomics Workbench.

3. Results

3.1. Transcriptomics

The sequencing of each of the species was either done by Paired-end sequencing or single-read sequencing as specified in Supplementary Table 1. Fastq files containing the sequenced reads were extracted from their respective SRA files. Paired-end sequencing libraries were split into files for 'left' reads and 'right' reads which were further used for assembly. Single-read files were used as they were. Following normalization, transcriptome assembly was performed on the reads for each species separately.

The assembled reads were filtered a further two times after the assembly using scripts created in-house. In the first instance, all short contigs (< 300 bp) were removed. From the resulting data, contigs with low expression i.e. with an FPKM (Fragments Per Kilobase of transcript per Million mapped reads) value less than 1 were removed. This filtered data as well as the initial unfiltered transcript datasets were used for further identification of neuropeptide precursors and mature peptides. The summary of the assemblies is displayed in Supplementary File Table 1.

3.2. Assessment of completeness

As mentioned in the earlier section, verification of the data for further analyses was done at all the stages pre-assembly as well as post-assembly. As part of the post-assembly stage verifications, the Benchmarking Universal Single-Copy Orthologs or BUSCO scores were measured on the assembly of each species. The assessment was done by comparing each of the assemblies to the 1658 single-copy, conserved and annotated insect orthologs BUSCO library. Almost all the species had BUSCO scores above 90% (Complete and Fragmented) and a majority of them had very high scores (over 96%). The only exception to this was *Carabus granulatus* with a BUSCO score of 87% (Complete and Fragmented), which was included as there was no other species data available for that sub-order Adephaga and family Carabidae. Any species that had a very low BUSCO score was discarded and another representative of that family or sub-family was used for the analysis. Fig. 2(a) shows the BUSCO scores of the species whose transcriptome were assembled (with the exception of *T. castaneum* and *D. ponderosae*), while Fig. 2(b) displays the range of the BUSCO scores of the species we have assembled compared against the scores of other insects whose BUSCO scores have been measured and recorded in earlier studies. To a first approximation, the BUSCO scores for a particular dataset indicate

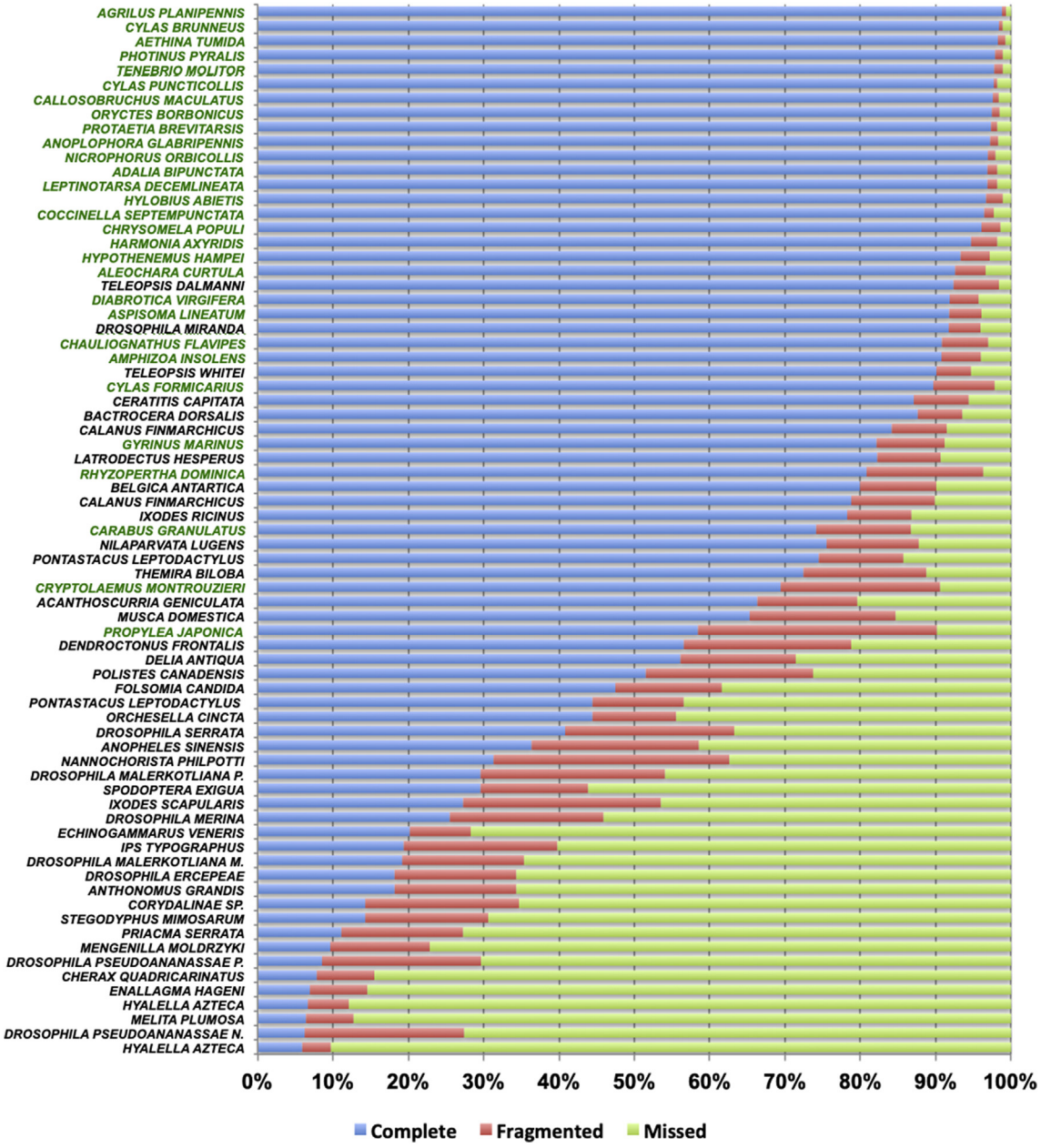


Fig. 2. (*continued*)

the percentage of single copy genes that would be present in that dataset, and thus the percentage of neuropeptides that we could expect to find in that dataset.

3.3. Neuropeptide profiles

Following the neuropeptide prediction pipelines discussed in the above section, the neuropeptide precursors that were identified in the 31 species have been shown in Fig. 3. Allatostatin B, Insulin-Like Peptides, Neuroparsins, CCHamides, Diuretic Hormone, ITG, Myosuppressins and Neuropeptide F are the most prevalent over all the species while Allatotropin, CCAP, Trissin, Natalisin and ACP are the least prevalent. Neuropeptides such as Corazonin, Allatostatin A, Allatostatin CCC and Sex Peptides were not identified, and none of these have been identified in any beetle species so far. The profiles of each neuropeptide have been discussed in further detail below.

3.4. Myoregulation

Members of the allatostatins family including Allatostatin B also known as the Myoinhibitory peptide (Schoofs et al., 1991; Blackburn et al., 1995), Allatostatin C, a cyclic neuropeptide characterized by a two cysteine residue disulphide bridge (Bendena and Tobe, 2012) and its paralog Allatostatin CC (Veenstra, 2009) have all been identified in more than 50% of the insects under study. AstB was found in all the species under study with each of the multiple mature peptides copies in the precursors easily distinguished by their C-terminal amidated end W (X_6)Wamide (see Fig. 4(a)). The mature peptide of AstC is identified by its C-terminal penta-peptide –P(VI)SCF and was identified in 27 species while the mature peptide for AstCC's C-terminal is –NAVTCF or a slight variation therein and was found in 18 species. Allatotropins, which have also been known to play a role in myostimulation during feeding (Duve et al., 1999; Masood and Orchard, 2014; Rudwall et al., 2000), have however been identified in very few of the studied species (n = 15).

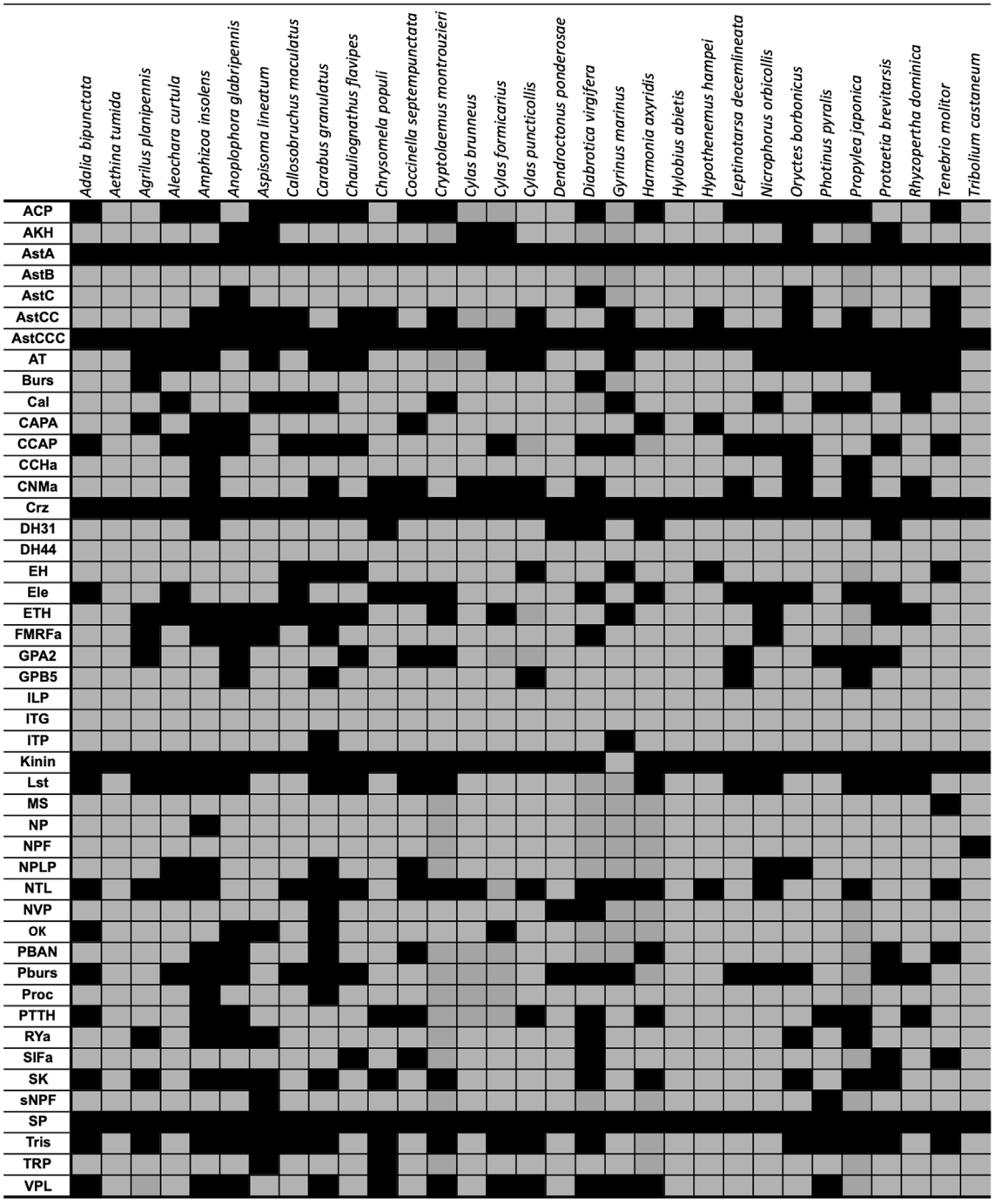


Fig. 3. A summary of the neuropeptide precursors identified in the 31 species under study. Boxes in gray are ones where the precursors were identified, while boxes in black are where no precursors were found.

Primarily the functionality of FMRFamide related peptides (FaRP) family members include myostimulation but given the multi-functionality of neuropeptides, they are also involved in ecdysis related behaviour, fluid secretion, even myoinhibition (Duve et al., 1992; Orchard et al., 2001; Kim et al., 2006; Nässel and Winther, 2010; Sedra and Lange, 2014; Suggs et al., 2016). Mature peptides sequences of FMRFamide are found as multiple copies in the precursor and are marked by their C-terminal FMRFa motifs. Myosuppressins, known to have inhibitory effects on the heart and visceral muscles in locusts and fruit flies (Orchard et al., 2001; Dickerson et al., 2012) have been identified in all except for *T. molitor* where it has not been identified. The mature peptide (QDVDHVFLRF–*amide*) shows a high level of conservation throughout all of the species in which it was identified (see Fig. 4(b)). Proctolin, which holds the distinction of being the first neuropeptide to be identified in insects (Starratt and Brown, 1975), also shows a highly conserved motif –RYLPT (Fig. 6 (a)) over all the species in this study where it was identified (n = 28).

Neuropeptide F (NPF) and short Neuropeptide F (sNPF) belong to the FaRP family as well but have been found to have multifunctional roles in insects including feeding and reproduction. Most of the NPF precursors identified in this study show two splice variants (one long and one short) as seen in Supplementary Table 3, and these variants have been observed in other insects (Xu et al., 2016; Veenstra, 2014; Derst et al., 2016). sNPFs on the other hand are smaller sequences with

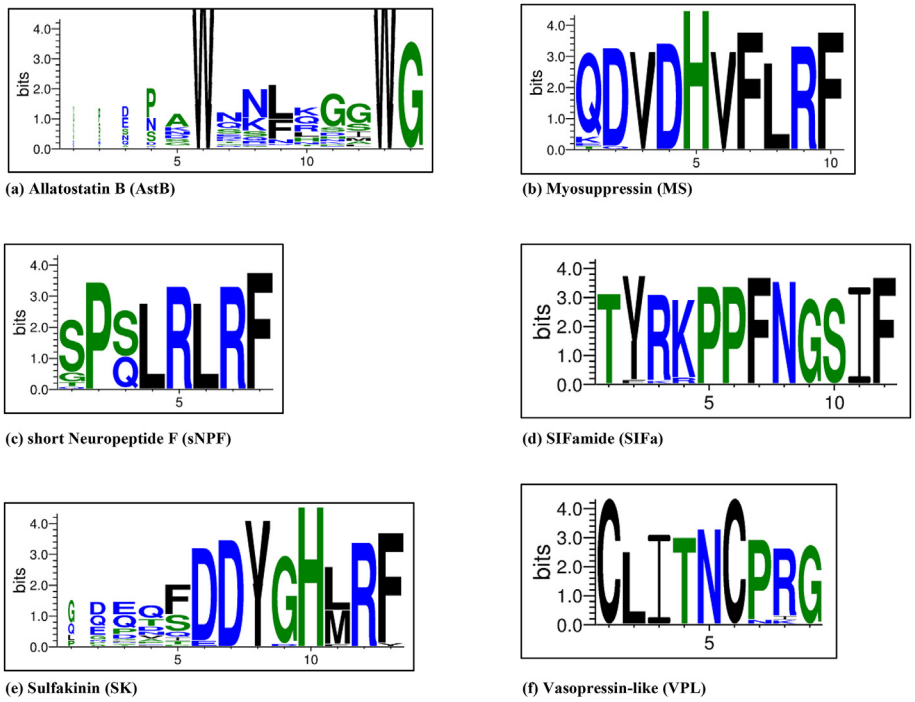


Fig. 4. Weblogos created from the alignments of the mature peptides for (a) Allatostatin B, (b) Myosuppressin (c) short Neuropeptide F (d) SIFamide (e) Sulfakinin and (f) Vasopressin-like from the various insect species under study. All sequences are presented in the N-terminus to C-terminus direction. In AstB, the characteristic W(X6)Wamide C-terminal is well conserved throughout all the sequences. For the Sulfakinin weblogo created from the alignment the highly conserved amidated motif HLRF on the C-terminal is seen, while for MS, sNPF, SIFa and VPL, all the mature peptides that were identified in the species under study were well-conserved as can be seen from the above Weblogos.

a distinctive and highly conserved RLRFamide C-terminal (Broeck, 2001; Mertens et al., 2002) as observed in Fig. 4(c).

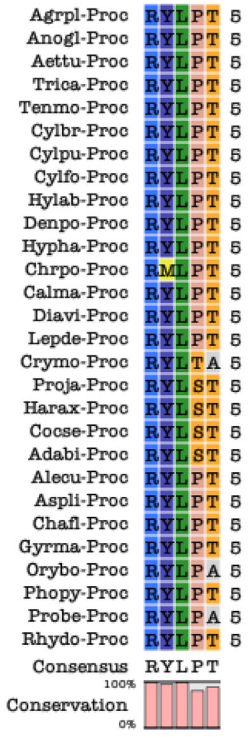
Kinin, first isolated from *L. maderae* (Holman et al 1986, 1987) is known to play an important role in myostimulation in visceral, reproductive muscles as well as in the secretory functionality of renal tubules (Coast et al., 1990; Schoofs et al., 1993; Dow, 2009). Although basal to the insects, Kinin has not been found in any beetle; but Veenstra (2019) identified both kinin and its receptor in the Adephagan *Pogonus chalceus*. Here, we have also identified a putative Kinin precursor in a second Adephaga member *Gyrinus marinus* (Fig. 5) by comparing against the sequence identified in *P. chalceus*. The presence of this putative Kinin precursor would need to be verified experimentally using other methods to confirm its presence to show evolutionary significance given that no other Adephaga members or any of the Polyphaga members in this study have shown the presence of Kinin.

3.5. Reproduction

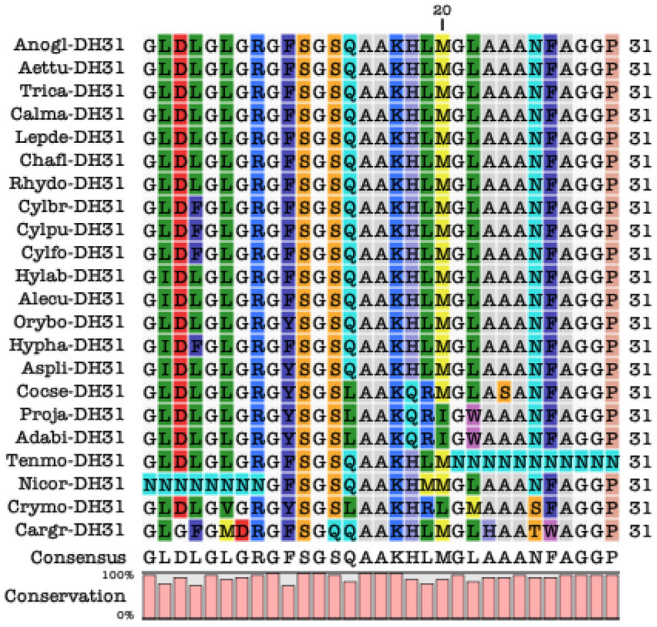
Neuropeptides that play a major role in reproduction include the ubiquitous Neuroparsin (NP) which was found in 30 of the species under study, SIFamide found in 26, Pheromone Biosynthesis Activating Neuropeptide (PBAN) found in 24 species while the least prevalent one was Natalisin or NTL (n = 13). NP was first isolated in *Locusta migratoria* and is a homo-dimer comprised of two bridged cysteine-rich peptides (Girardie et al., 1989). SIFamide has been known to cause an increase in promiscuity in male *Drosophila,* while making females more receptive to mating (Tehrzaz et al., 2007). Its function in Coleopterans has not been studied in detail yet, however, the neuropeptide has clearly been identified in most of the species under study (Fig. 4 (d)). PBAN is found on a precursor that is responsible for encoding other neuropeptides such as Pyrokinin (PK) and Diapause hormone (DH). PBAN regulates pheromone biosynthesis (Raina and Menn, 1993) and works in tandem with DH, which is responsible for inhibition of insect diapause (Zhang et al., 2011). NTLs share an ancestry with tachykinin-related peptides (Jiang et al., 2013) and although they are not abundantly identified in most of the species in this study, multiple copies are found for each of the species where the precursor for Natalisin was identified.

```
pogch_kinin     MYYKNSVIQTSKILLQIILYILIQFANTNCMTIPNDNSNMLQTKSIRNRLNLLEDDNIFG   60
gyrma_kinin     ------------------------------------------------------------   0

pogch_kinin     TETPTDTKTYNSGNDMENNNNGDIFTKETDNENIFTEDSDIDDKKLNVNKIVSLLLDPNE   120
gyrma_kinin     ------------------------------------------------------------   0

pogch_kinin     VTPKMIEKLSQFKTFYLSNSNDKTLINLCTLSRNYINDVCLLIQSSNRSNGLYNPYIYQR   180
gyrma_kinin     ----------------MSKPDD--------------------------WEIKNKIAYQM   17
                              :*: :*                              : *   **

pogch_kinin     CVKLL--------PLLDFYHCVNEKQNYDETAKENEFKRVPFHAWGGKRFRQVATEKVQS   232
gyrma_kinin     LMDASKQVDDEDAKHVKFLGWI--GKRFSHANGNNIPKRIPFHAWGGKRSNQVGTEKIQA   75
                 :.            :.*   :   :.:..:  :*  **:********* .**.***:*:

pogch_kinin     SKVPFNAWGGKRQMNYDDGDE-----------------YCKRSKSKFHSWGGKRTVRET   274
gyrma_kinin     SRVPFNAWGGKRTYNMNSDESTPDNMYEINTPSSVLPYLVYKRSRTKFHSWGGKRNY---   132
                *:**********  * :..:.                    ***::*********.

pogch_kinin     ANVL*       278
gyrma_kinin     -----       132
```

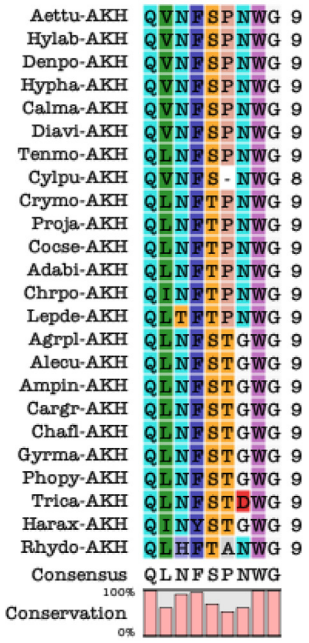
Fig. 5. Pairwise sequence alignment carried out on the putative precursor sequences Kinin (or Leucokinin as it is commonly denoted) of *Pogonus chalceus* (as identified by Veenstra, 2019) and *Gyrinus marinus*, both members of Sub order Adephaga. The putative mature peptide sequences for each are highlighted above in gray and the cleavage sites in red. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)



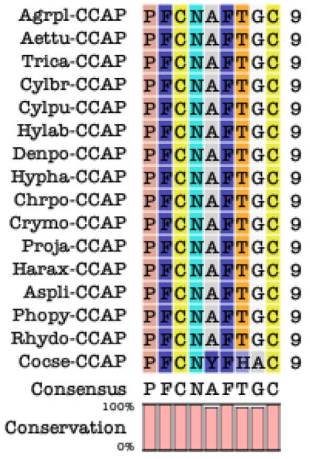
(a) Proctolin



(b) Diuretic Hormone (DH31)



(c) Adipokinetic Hormone (AKH)



(d) Crustacean Cardioactive Peptide (CCAP)

Fig. 6. Multiple sequence alignments showing alignments of the (a) Proctolin (Proc) (b) Diuretic Hormone (DH31) (c) Adipokinetic Hormone or AKH and (d) Crustacean Cardioactive Peptide or CCAP mature peptide sequences of the various species in which they were identified. The consensus sequences of each show a very high percentage of similarity especially for CCAP as well as for Proc where the peptide sequence is highly conserved.

3.6. Feeding

RYamides, characterized by their C-terminal RYamide ending, have been found to reduce an insect's motivation to feed (Ida et al., 2011a) triggered by satiety. Each precursor usually encodes two mature peptides for RYamide and this has been the case for most of the species under study where the precursors were identified (n = 24) with a few exceptions. Sulfakinins (SK) are also members of the RFamide family, but additionally characterized by a sulfated C-terminal (YGHMRFamide) [see Fig. 4(e)]. Like the RYamides, these neuropeptides are also responsible for feeding inhibition and are normally encoded as two mature peptides per precursor. Although broadly present, they were not

found in two predatory species. Limostatins (Lst) have been studied in *Drosophila* and have been known to be responsible for production of ILPs (Alfa et al., 2015), although they haven't been studied extensively in Coleopterans. The mature peptides of Trissin that have been identified in the species in this study (n = 14) show a high degree of conservation. It is structurally composed of six highly conserved cysteine residues forming three disulfide bridges (Ida et al., 2011b).

3.7. Water and ion regulation

Diuretic hormones were identified in the various species in this study. They play a major role in fluid secretion as has been widely studied (Coast et al., 2001, Mirabeau and Joly, 2013; Zandawala, 2012) but are also functional in muscle contractions, reproduction and molting. The sequences of the mature peptides for DH31 (named thus due to the peptide being 31 amino acids long) in the various species where it was identified show a high conservation and this can be seen in Fig. 6(b). The corticotropin-releasing factor-like diuretic hormone (CRF-DH) peptides are however more prevalent in the species under study, with all 31 species displaying their presence. The glycoproteins GPA2 and GPB5 contain 10 cysteine residues, which are highly conserved, and have been recently identified to have a function in ion transport (Paluzzi et al., 2014). Precursors and the encoding mature peptides were identified for both of these GPA2 (n = 22) and GPB5 (n = 26) in the species under study. The Ion Transport Peptide (ITP) whose main functionality is modulation of ion transport (Audsley et al., 1992; King et al., 1999) and was one of the most prevalent neuropeptides predicted in the species under study (n = 29). Capability/CAP2B or CAPA neuropeptide precursors usually contain two CAPA peptides and a pyrokinin peptide. The CAPA peptides (CAPA-1 and CAPA-2) are involved in various functionality such as anti-diuresis (Paluzzi et al., 2008), myotropicity (Wegener et al., 2002) but an intensive study by Halberg et al., (2015) showed that CAPA peptides were involved in the stimulation of secretion in Malpighian tubules of several insects, whereas the functionality of CAPA-PK is still unknown. The highly conserved Vasopressin-like peptide (Fig. 4(f)) was seen in close to 60% of the species under study and has also been found to be involved in triggering diuresis in *T. castaneum* (Aikins et al., 2008). Kinin peptides were covered in a previous section.

3.8. Moulting and development

Bursicon (alpha and beta) have been found to induce wing expansion, tanning and plays an important role in ecdysis in various insects (Luo et al., 2005; Mendive et al., 2005; Dai et al., 2008; Peabody et al., 2008). Burs alpha was identified in 26 while Burs beta or partner of Bursicon was identified in only 16 of the 31 insects under study. One of the main peptides involved in ecdysis in insects (Ewer, 2005) and which is responsible for the release of the ecdysis triggering hormone (ETH) is the Eclosion hormone or EH. EH and ETH were both identified in the species under study however, EH was found in a considerable number more (n = 24) than ETH (n = 17). Adipokinetic hormone/corazonin related peptide or ACP was not identified in many of the species under study (n = 14), but this could be due to its high similarity with Adipokinetic hormone (AKH), which is found in 24 of the 31 species under study (Fig. 6(c)). Unlike AKH, which is known to be multifunctional with its roles in heart rate stimulation, lipid synthesis, locomotory and myotropic activity and protein inhibition (Gäde, 1997; Gäde and Auerswald, 2003; Bharucha et al., 2008; Gäde, 2009; Nässel and Winther, 2010; Hauser and Grimmelikhuijzen, 2014), the functionality of ACP has not been clearly defined although some studies by Zandawala et al. (2015) have suggested their role in development and ecdysis. Prothoracicotropic hormone (PTTH) regulates the production and release of ecdysome, which is responsible for the coordination of molting and metamorphosis (Gilbert et al., 2002). It also regulates development with the production of ecdysone (Rewitz et al., 2009).

3.9. Multifunctional neuropeptides

Certain neuropeptides are considered multifunctional either because they play multiple roles in the various regulatory processes on an insect, or they have shown different roles in different insects. Insulin-like peptides or ILP precursors contain numerous encoding genes ranging from 1 to 38 (Mizoguchi and Okamoto, 2013). In the species under study, they range from 1 to 4 in some cases. These were, like AstB, one of the most prevalent neuropeptides identified across all the studied Coloepterans (n = 31). ILPs are involved in quite a few regulatory processes in insects including but not limited to growth and development, reproduction and metabolism. This also is the case in Tachykinin-related peptides (TRP), the precursors of whom encode multiple peptides paracopies and like ILPs are part of multiple functionalities in the insects including locomotion, modulation of olfactory input, regulation of insulin producing neurons, metabolism and even aggression in male flies (Kahsai et al., 2010; Birse et al., 2011; Asahina et al., 2014; Ko et al., 2015). 29 of the Coleopterans studied in this project showed the presence of precursors and the encoded mature peptides for TRP. Orcokinins which show myostimulatory properties in crustaceans have been identified in Coloeopterans (26 out of the 31 species in this study) and other insect orders but their functionality differs in the various species. It has been found to play a role in the circadian clock in cockroaches, have neurosensory functionality in certain locusts and have also been known to play a function in gut activities (Pascual et al., 2004; Hofer et al., 2005; Hofer and Homberg, 2006). CCAP or the crustacean cardioactive peptide first discovered in *Carcinus maenas* (Stangier et al., 1987) as its name suggests is responsible for a multitude of cardio active functions in insect species. Although it was only identified in 16 of the species under study, they all show a high conservation of the peptide sequence PFCNAFTGCamide (Fig. 6(d)).

3.10. Neuropeptides with unknown function

There are two CCHamide genes each coded as CCH1 and CCH2 where the function of the latter is known to cause feeding irregularities (Ren et al., 2015; Sano et al., 2015) whereas the function of CCH1 is not yet known. The precursors for CCHamide were quite prevalent and the two distinct peptides were found in 28 of the 31 species under study, as were the precursors for ITG-like peptides (n = 31) and NVP-like peptides (n = 28) both of whose function has not been determined yet. A study by Veenstra (2014) showed that a gene that encodes calcitonin has been identified in locust and termite genomes. In the precursors of the Coleopteran species under study, interestingly in some cases we found three transcript variants that code for different mature Calcitonin peptides. CNMamides are named accordingly based on their amidated Cysteine-Aspartic Acid-Methionine C-terminal (Jung et al., 2014) and this can be observed in the 19 species in which it was identified. Similarly the two-cysteine residue peptide Elevenin identified in other insects (Veenstra, 2010) was identified in 18 of the species under study. The functions of either of these have not been identified so far.

3.11. Ecological and evolutionary trend analysis

From the results of the neuropeptide identification above, we have performed trend analysis on various parameters including phylogenetic divergence, habitat, diet and size to understand any evolutionary trends over these parameters.

The oldest beetles are known to have originated approximately 240 MYA with the two oldest known sub-orders being Archostemata and Myxophaga. Adephaga is believed to have diverged around 200 MYA while Polyphaga, that contains the largest number of beetle species, is believed to have diverged around 225 MYA. By ordering the species based on their dates of origin and arranging their neuropeptidomes from most prevalent to least prevalent, we can see that neuropeptides are more easily identified in more phylogenetically recent beetle

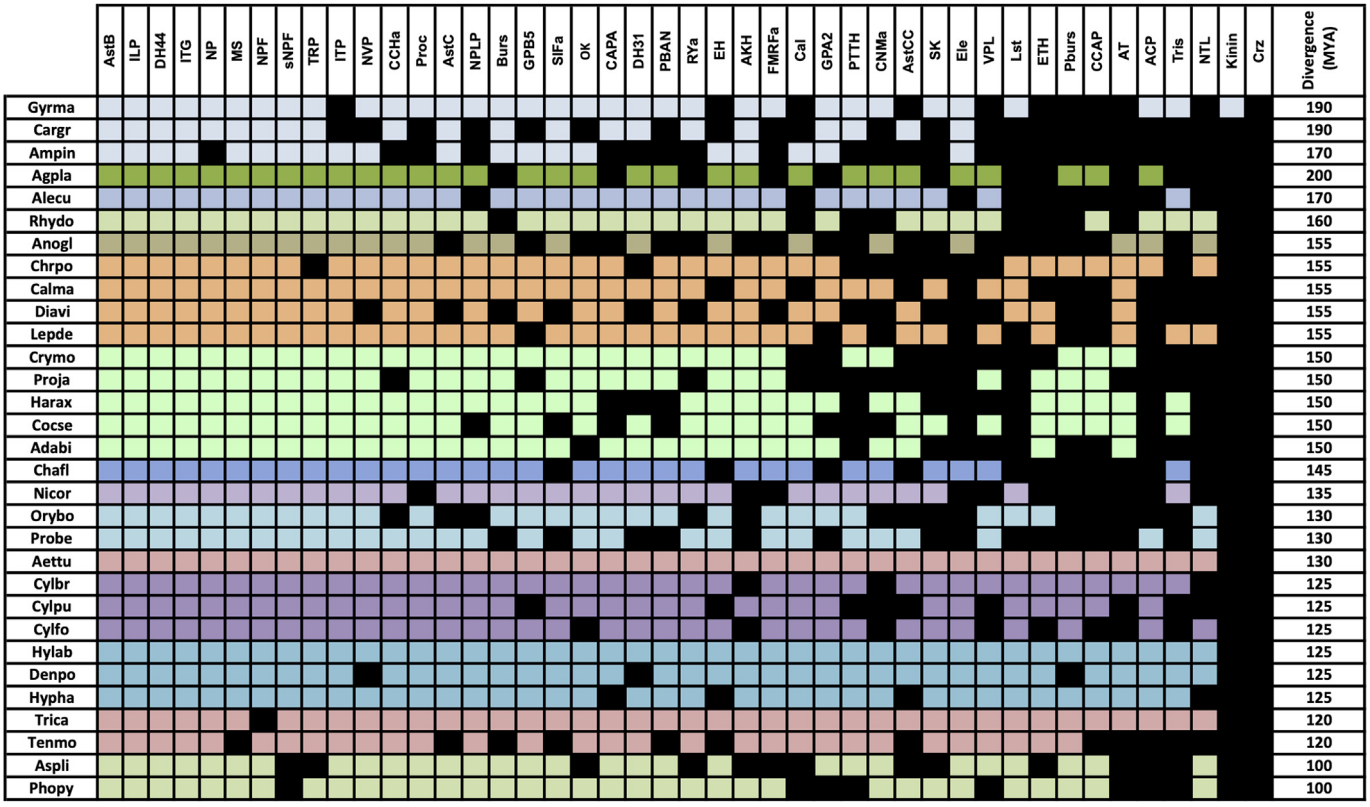


Fig. 7. A heat map of the neuropeptide profiles of each of the species under study grouped in descending order from oldest to youngest in terms of divergence in MYA. Lengthwise from left to right the neuropeptides are arranged from most prevalent to least prevalent. Missing neuropeptides are displayed as black boxes.

species (see Fig. 7). The absence of neuropeptides could also be due to loss of their functionality or the presence of other neuropeptides that are more necessary than others.

With regards to habitat, we have categorized the species based on whether they are aquatic or terrestrial. The ancient Adephaga members are aquatic while most if not all Polyphaga members are terrestrial (Fig. 8). Terrestrial species would require better water regulation especially after activities that result in water loss such as myoactive

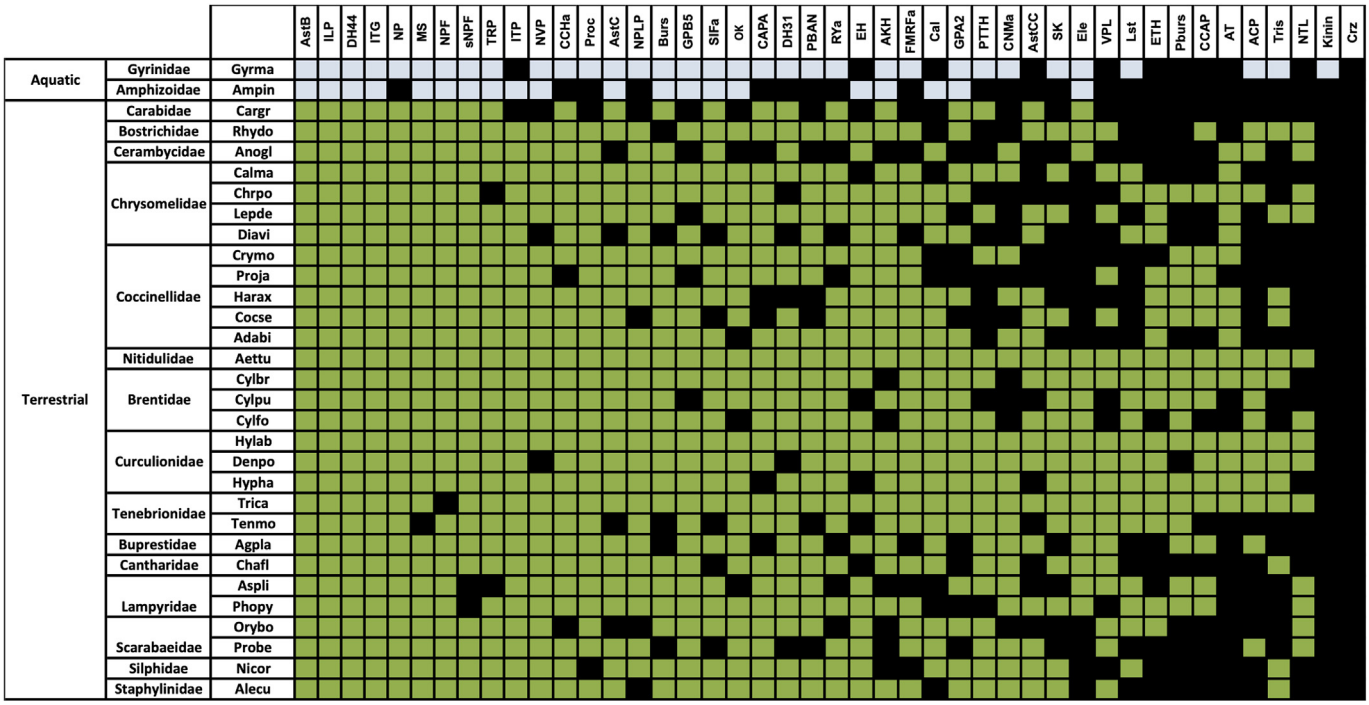


Fig. 8. A heat map of the neuropeptide profiles of each of the species under study grouped in according to their habitat, either aquatic or terrestrial. Lengthwise from left to right the neuropeptides are arranged from most prevalent to least prevalent. Missing neuropeptides are displayed as black boxes.

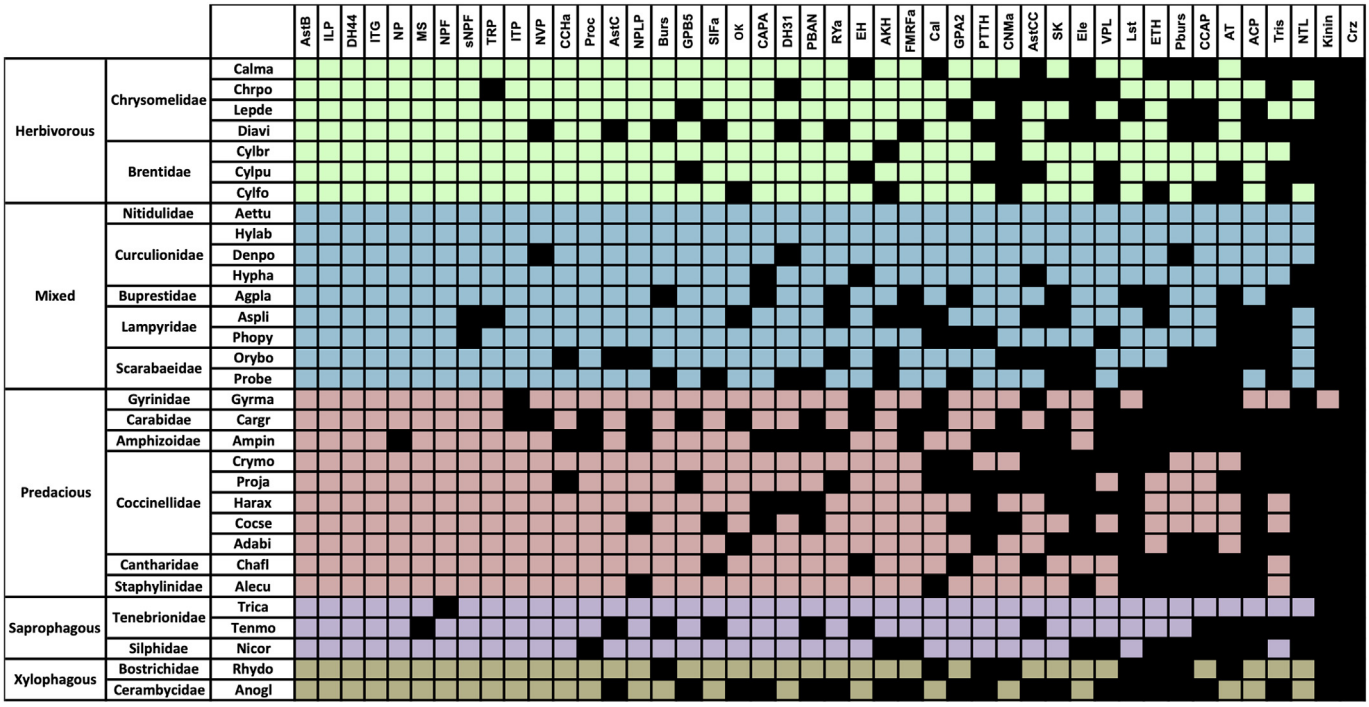


Fig. 9. A heat map of the neuropeptide profiles of each of the species under study grouped in according to their dietary behaviour including herbivorous, mixed, predacious, saprophagous and xylophagous. Lengthwise from left to right the neuropeptides are arranged from most prevalent to least prevalent. Missing neuropeptides are displayed as black boxes.

behaviour, eclosion, and most importantly, excretion. The presence or absence of certain neuropeptides such as diuretic hormones and peptides, among others involved in these functionalities, would be of great significance. The lack of a substantial number of aquatic species is a major factor that does not allow an accurate and definitive trend analysis on the effect this parameter has on the presence or absence of neuropeptides. Similarly with the terrestrial species, there could be trends based on their particular habitats but that is beyond the scope of this study.

Evolutionarily, beetles have been either primarily predacious or herbivorous i.e. plant feeding. Over time, their dietary behaviour has evolved into xylophagous (feeding on plants and plant products i.e. grain, cereal, flowers etc.) or even saprophagous (feeding on animal carcass, animal material, flour derived from plant products among others.). Some of the phylogenetically more recent beetles show a mixed diet i.e. different stages consume different type of food. An example would be the two Lampyridaens, *Photinus pyralis* and *Aspisoma lineatum* whose larvae feed on snails and slugs while adults feed primarily on nectar. Fig. 9 shows the trends of the various groups of species based on their diet. Herbivore, polymorphic and saprophagous species show more neuropeptides than the predacious and xylophagous ones but this could be attributed to the number of species for each category available for this study.

Beetles are found in a multitude of sizes as mentioned before. Size may play a factor on the presence or absence of neuropeptides, wherein differently sized species may have differing osmoregulatory or growth requirements. With the species in this study, although there is no clear trend in terms of divergence and age, we can observe that the smallest beetles and the relatively middle sized ones showed more neuropeptides than the larger ones. The exception to this is the largest, *Rhyzopertha dominica,* in which we identified 34 of the total neuropeptides under study (Fig. 10).

4. Discussion

In this study, using Bioinformatics pipelines, we have assembled and analysed the sequence data of 29 different members of the order Coleoptera (and studied the pre-assembled genomic data for two) and further identified putative neuropeptide precursors and mature peptides to create neuropeptide profiles for each of them. The species include the sub orders, Polyphaga and Adephaga and representative species from their underlying families and sub-families thus accounting for a wide diversity of species. It should be noted that the focus of this study was the identification of neuropeptides with known homologs rather than novel neuropeptide discovery. Interestingly, peptide families range from always found (AstB, ILP, DH44 and ITG), through those found in all but one species studied (NP, MS, NPF), through a large majority found in a subset of species, to kinin (now found in just two species) and Corazonin (not found). Although absence of a neuropeptide in a single species could reflect poor sequence coverage, the high BUSCO threshold used here makes it very unlikely that reported absence across multiple species is entirely due to such sequencing shortfalls. The emerging picture therefore, is that while nearly all peptide families have uses and have been known to perform or be involved in multiple functionalities, and presumably offer adaptive advantages (in the form of carrying out each other's roles, thereby diminishing the need for numerous neuropeptides), the subset of essential peptide families is much smaller than we had previously supposed. This could lead us to speculate that these factors could gradually contribute to a reduction in the number of peptides involved in neuropeptide signaling.

We have also looked at the evolutionary development and trends in terms of date of origin i.e. divergence, habitat, dietary behaviour and size parameters to gain a better understanding of gene conservation/differentiation of neuropeptide families across this comprehensive set of Coleopteran species. To our surprise, although there were some suggestions of relationships, there was no clear-cut association between size, habitat or diet. As new genomes are sequenced, it will be interesting to survey the peptide families across all the Insecta.

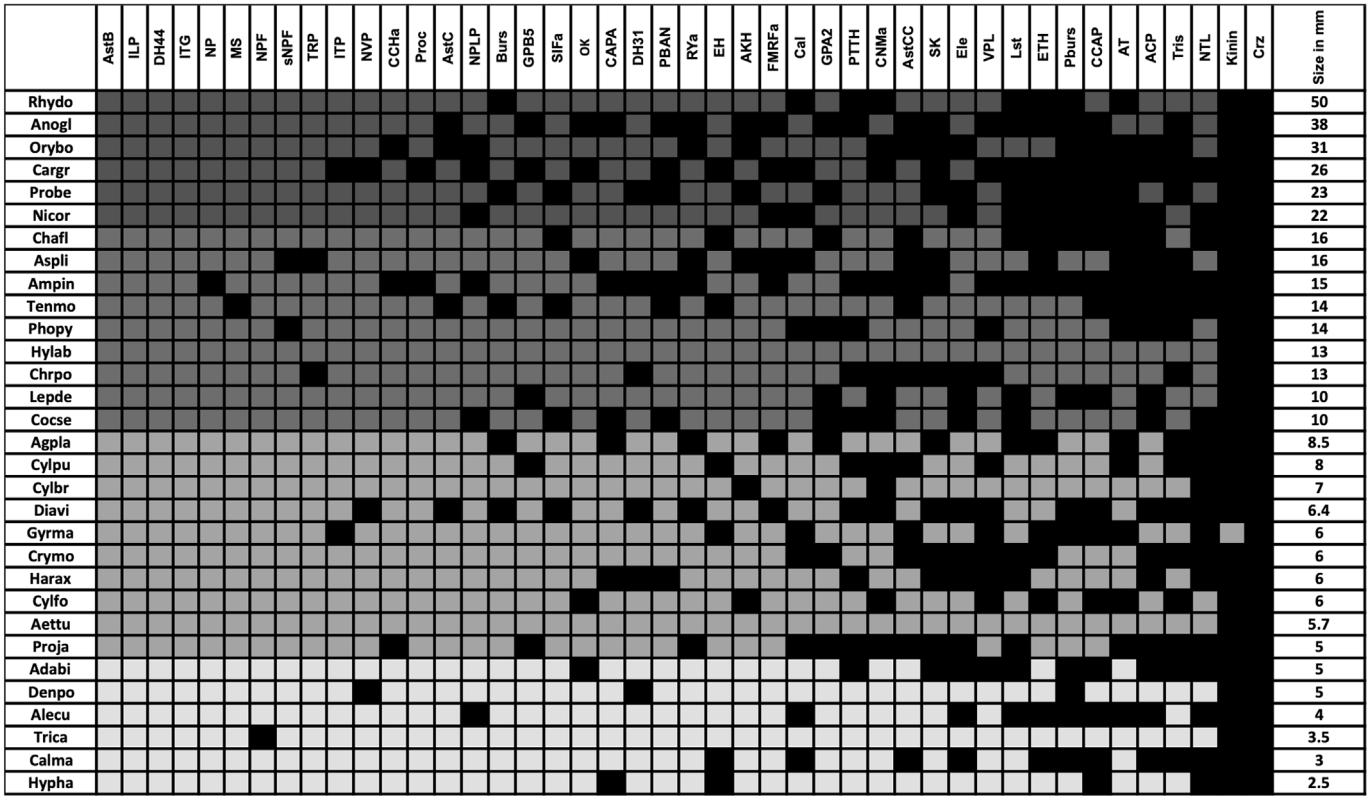


Fig. 10. A heat map of the neuropeptide profiles of each of the species under study grouped in according to their adult size grouped in descending order from largest to smallest. Lengthwise from left to right the neuropeptides are arranged from most prevalent to least prevalent. Missing neuropeptides are displayed as black boxes.

Funding

This work was supported by the European Union's Horizon 2020 Research and Innovation programme (SD/JATD/GS) under grant agreement No. 634361.

Conflicts of interest

None declared.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ibmb.2019.103227.

References

Afgan, E., Baker, D., Van den Beek, M., Blankenberg, D., Bouvier, D., Čech, M., Chilton, J., Clements, D., Coraor, N., Eberhard, C., Grüning, B., 2016. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. Nucleic Acids Res. 44 (W1), W3–W10.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215 (3), 403–410.

Aikins, M.J., Schooley, D.A., Begum, K., Detheux, M., Beeman, R.W., Park, Y., 2008. Vasopressin-like peptide and its receptor function in an indirect diuretic signaling pathway in the red flour beetle. Insect Biochem. Mol. 38 (7), 740–748.

Alfa, R.W., Park, S., Skelly, K.R., Poffenberger, G., Jain, N., Gu, X., Kockel, L., Wang, J., Liu, Y., Powers, A.C., Kim, S.K., 2015. Suppression of insulin production and secretion by a decretin hormone. Cell Metab. 21 (2), 323–334.

Altstein, M., Nässel, D.R., 2010. Neuropeptide signaling in insects. In: In: Geary, T.G., Maule, A.G. (Eds.), Neuropeptide Systems as Targets for Parasite and Pest Control. ADV EXP MED BIOL, vol 692 Springer, Boston, MA.

Andrews, S., 2016. FastQC: a Quality Control Tool for High Throughput Sequence Data. pp. 2010.

Asahina, K., Watanabe, K., Duistermars, B.J., Hoopfer, E., González, C.R., Eyjólfsdóttir, E.A., Perona, P., Anderson, D.J., 2014. Tachykinin-expressing neurons control male-specific aggressive arousal in Drosophila. Cell 156 (1–2), 221–235.

Audsley, N., McIntosh, C., Phillips, J.E., 1992. Isolation of a neuropeptide from locust corpus cardiacum, which influences ileal transport. J. Exp. Biol. 173 (1), 261–274.

Bendena, W.G., Tobe, S.S., 2012. Families of allatoregulator sequences: a 2011 perspective 1 1 This review is part of a virtual symposium on recent advances in understanding a variety of complex regulatory processes in insect physiology and endocrinology, including development, metabolism, cold hardiness, food intake and digestion, and diuresis, through the use of omics technologies in the postgenomic era. Can. J. Zool. 90 (4), 521–544.

Beutel, R.G., Leschen, R.A.B., 2005. Handbook of zoology. In: A Natural History of the Phyla of the Animal Kingdom 4.

Bharucha, K.N., Tarr, P., Zipursky, S.L., 2008. A glucagon-like endocrine pathway in Drosophila modulates both lipid and carbohydrate homeostasis. J. Exp. Biol. 211 (19), 3103–3110.

Birse, R.T., Söderberg, J.A., Luo, J., Winther, Å.M., Nässel, D.R., 2011. Regulation of insulin-producing cells in the adult Drosophila brain via the tachykinin peptide receptor DTKR. J. Exp. Biol. 214 (24), 4201–4208.

Blackburn, M.B., Wagner, R.M., Kochansky, J.P., Harrison, D.J., Thomas-Laemont, P., Raina, A.K., 1995. The identification of two myoinhibitory peptides, with sequence similarities to the galanins, isolated from the ventral nerve cord of *Manduca sexta*. Regul. Pept. 57 (3), 213–219.

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30 (15), 2114–2120.

Broeck, J.V., 2001. Neuropeptides and their precursors in the fruitfly, Drosophila melanogaster. Peptides 22 (2), 241–254.

Brown, C.T., Howe, A., Zhang, Q., Pyrkosz, A.B., Brom, T.H., 2012. A Reference-free Algorithm for Computational Normalization of Shotgun Sequencing Data. arXiv preprint arXiv:1203.4802.

Christie, A.E., 2008. Neuropeptide discovery in Ixodoidea: an in silico investigation using publicly accessible expressed sequence tags. Gen. Comp. Endocr. 157 (2), 174–185.

Christie, A.E., 2015. In silico prediction of a neuropeptidome for the eusocial insect *Mastotermes darwiniensis*. Gen. Comp. Endocr. 224, 69–83.

Coast, G.M., Holman, G.M., Nachman, R.J., 1990. The diuretic activity of a series of cephalomyotropic neuropeptides, the achetakinins, on isolated Malpighian tubules of the house cricket, *Acheta domesticus*. J. Insect Physiol. 36 (7), 481–488.

Coast, G.M., Webster, S.G., Schegg, K.M., Tobe, S.S., Schooley, D.A., 2001. The Drosophila melanogaster homologue of an insect calcitonin-like diuretic peptide stimulates V-ATPase activity in fruit fly Malpighian tubules. J. Exp. Biol. 204 (10), 1795–1804.

Crooks, G.E., Hon, G., Chandonia, J.M., Brenner, S.E., 2004. WebLogo: a sequence logo generator. Genome Res. 14 (6), 1188–1190.

Crowson, R.A., 2013. The Biology of the Coleoptera. Academic Press.

Dai, L., Dewey, E.M., Zitnan, D., Luo, C.W., Honegger, H.W., Adams, M.E., 2008. Identification, developmental expression, and functions of bursicon in the tobacco hawkmoth, Manduca sexta. J. Comp. Neurol. 506 (5), 759–774.

Derst, C., Dircksen, H., Meusemann, K., Zhou, X., Liu, S., Predel, R., 2016. Evolution of

neuropeptides in non-pterygote hexapods. BMC Evol. Biol. 16 (1), 51.
Dickerson, M., McCormick, J., Mispelon, M., Paisley, K., Nichols, R., 2012. Structure–activity and immunochemical data provide evidence of developmental-and tissue-specific myosuppressin signaling. Peptides 36 (2), 272–279.
Dow, J.A., 2009. Insights into the Malpighian tubule from functional genomics. J. Exp. Biol. 212 (3), 435–445.
Duve, H., Johnsen, A.H., Sewell, J.C., Scott, A.G., Orchard, I., Rehfeld, J.F., Thorpe, A., 1992. Isolation, structure, and activity of-Phe-Met-Arg-Phe-NH2 neuropeptides (designated calliFMRFamides) from the blowfly *Calliphora vomitoria*. P. Natl. Acad. Sci. U.S.A. 89 (6), 2326–2330.
Duve, H., East, P.D., Thorpe, A., 1999. Regulation of lepidopteran foregut movement by allatostatins and allatotropin from the frontal ganglion. J. Comp. Neurol. 413 (3), 405–416.
Ewer, J., 2005. Behavioral actions of neuropeptides in invertebrates: insights from Drosophila. Horm. Behav. 48 (4), 418–429.
Gade, G., 1997. The explosion of structural information on insect neuropeptides. In: Fortschritte der Chemie organischer Naturstoffe/Progress in the Chemistry of Organic Natural Products. Springer, Vienna, pp. 1–128.
Gäde, G., 2009. Peptides of the adipokinetic hormone/red pigment-concentrating hormone family. Ann. NY Acad. Sci. 1163 (1), 125–136.
Gäde, G., Auerswald, L., 2003. Mode of action of neuropeptides from the adipokinetic hormone family. Gen. Comp. Endocr. 132 (1), 10–20.
Gilbert, L.I., Rybczynski, R., Warren, J.T., 2002. Control and biochemical nature of the ecdysteroidogenic pathway. Annu. Rev. Entomol. 47 (1), 883–916.
Girardie, J., Girardie, A., Huet, J.C., Pernollet, J.C., 1989. Amino acid sequence of locust neuroparsins. FEBS Lett. 245 (1–2), 4–8.
Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. 29 (7), 644–652.
Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., Couger, M.B., Eccles, D., Li, B., Lieber, M., MacManes, M.D., 2013. De novo transcript sequence reconstruction from RNA-Seq: reference generation and analysis with Trinity. Nat. Protoc. 8 (8).
Halberg, K.A., Terhzaz, S., Cabrero, P., Davies, S.A., Dow, J.A., 2015. Tracing the evolutionary origins of insect renal function. Nat. Commun. 6, 6800.
Hauser, F., Grimmelikhuijzen, C.J., 2014. Evolution of the AKH/corazonin/ACP/GnRH receptor superfamily and their ligands in the Protostomia. Gen. Comp. Endocr. 209, 35–49.
Hewes, R.S., Taghert, P.H., 2001. Neuropeptides and neuropeptide receptors in the Drosophila melanogaster genome. Genome Res. 11 (6), 1126–1142.
Hofer, S., Dircksen, H., Tollbäck, P., Homberg, U., 2005. Novel insect orcokinins: characterization and neuronal distribution in the brains of selected dicondylian insects. J. Comp. Neurol. 490 (1), 57–71.
Hofer, S., Homberg, U., 2006. Evidence for a role of orcokinin-related peptides in the circadian clock controlling locomotor activity of the cockroach Leucophaea maderae. J. Exp. Biol. 209 (14), 2794–2803.
Holman, G.M., Cook, B.J., Nachman, R.J., 1986. Isolation, primary structure and synthesis of two neuropeptides from Leucophaea maderae: members of a new family of cephalomyotropins. Comp. Biochem. Phys. C 84 (2), 205–211.
Holman, G.M., Cook, B.J., Nachman, R.J., 1987. Isolation, primary structure and synthesis of leucokinins VII and VIII: the final members of this new family of cephalomyotropic peptides isolated from head extracts of Leucophaea maderae. Comp. Biochem. Phys. C 88 (1), 31–34.
Hummon, A.B., Richmond, T.A., Verleyen, P., Baggerman, G., Huybrechts, J., Ewing, M.A., Vierstraete, E., Rodriguez-Zas, S.L., Schoofs, L., Robinson, G.E., Sweedler, J.V., 2006. From the genome to the proteome: uncovering peptides in the Apis brain. Science 314 (5799), 647–649.
Hunt, T., Bergsten, J., Levkanicova, Z., Papadopoulou, A., John, O.S., Wild, R., Hammond, P.M., Ahrens, D., Balke, M., Caterino, M.S., Gómez-Zurita, J., 2007. A comprehensive phylogeny of beetles reveals the evolutionary origins of a superradiation. Science 318 (5858), 1913–1916.
Ida, T., Takahashi, T., Tominaga, H., Sato, T., Kume, K., Ozaki, M., Hiraguchi, T., Maeda, T., Shiotani, H., Terajima, S., Sano, H., 2011a. Identification of the novel bioactive peptides dRYamide-1 and dRYamide-2, ligands for a neuropeptide Y-like receptor in Drosophila. Biochem. Bioph. Res. Co. 410 (4), 872–877.
Ida, T., Takahashi, T., Tominaga, H., Sato, T., Kume, K., Yoshizawa-Kumagaye, K., Nishio, H., Kato, J., Murakami, N., Miyazato, M., Kangawa, K., 2011b. Identification of the endogenous cysteine-rich peptide trissin, a ligand for an orphan G protein-coupled receptor in Drosophila. Biochem. Bioph. Res. Co. 414 (1), 44–48.
Jiang, H., Lkhagva, A., Daubnerova, I., Chae, H.S., Simo, L., Jung, S.H., Yoon, Y.K., Lee, N.R., Seong, J.Y., Zitnan, D., Park, Y., Kim, Y.J., 2013. Natalisin, a tachykinin-like signaling system, regulates sexual activity and fecundity in insects. P. Natl. Acad. Sci. U.S.A. 110 (37), E3526–E3534.
Jung, S.H., Lee, J.H., Chae, H.S., Seong, J.Y., Park, Y., Park, Z.Y., Kim, Y.J., 2014. Identification of a novel insect neuropeptide, CNMa and its receptor. FEBS Lett. 588 (12), 2037–2041.
Kahsai, L., Martin, J.R., Winther, Å.M., 2010. Neuropeptides in the Drosophila central complex in modulation of locomotor behavior. J. Exp. Biol. 213 (13), 2256–2265.
Keeling, C.I., Yuen, M.M., Liao, N.Y., Docking, T.R., Chan, S.K., Taylor, G.A., Palmquist, D.L., Jackman, S.D., Nguyen, A., Li, M., Henderson, H., 2013. Draft genome of the mountain pine beetle, *Dendroctonus ponderosae* Hopkins, a major forest pest. Genome Biol. 14 (3), R27.
Kim, Y.J., Žitňan, D., Galizia, C.G., Cho, K.H., Adams, M.E., 2006. A command chemical triggers an innate behavior by sequential activation of multiple peptidergic ensembles. Curr. Biol. 16 (14), 1395–1407.
King, D.S., Meredith, J., Wang, Y.J., Phillips, J.E., 1999. Biological actions of synthetic locust ion transport peptide (ITP). Insect Biochem. Mol. 29 (1), 11–18.
Ko, K.I., Root, C.M., Lindsay, S.A., Zaninovich, O.A., Shepherd, A.K., Wasserman, S.A., Kim, S.M., Wang, J.W., 2015. Starvation promotes concerted modulation of appetitive olfactory behavior via parallel neuromodulatory circuits. elife 4, e08298.
Lawrence, J.F., Newton Jr., A.F., 1982. Evolution and classification of beetles. Annu. Rev. Ecol. Syst. 13 (1), 261–290.
Leinonen, R., Sugawara, H., Shumway, M., International Nucleotide Sequence Database Collaboration, 2010. The sequence read archive. Nucleic Acids Res. 39 (Suppl. l_1), D19–D21.
Leschen, R.A.B., Beutel, R.G., Lawrence, J.F., Slipinski, A., 2009. Handbook of Zoology, vol. 2 Coleoptera, Beetles.
Leschen, R.A.B., Beutel, R.G. (Eds.), 2014. Handbook of Zoology, Coleoptera, vol. 3 Morphology and Systematics (Phytophaga) Walter de Gruyter, Berlin.
Li, B., Predel, R., Neupert, S., Hauser, F., Tanaka, Y., Cazzamali, G., Williamson, M., Arakane, Y., Verleyen, P., Schoofs, L., Schachtner, J., 2008. Genomics, transcriptomics, and peptidomics of neuropeptides and protein hormones in the red flour beetle *Tribolium castaneum*. Genome Res. 18 (1), 113–122.
Luo, C.W., Dewey, E.M., Sudo, S., Ewer, J., Hsu, S.Y., Honegger, H.W., Hsueh, A.J., 2005. Bursicon, the insect cuticle-hardening hormone, is a heterodimeric cystine knot protein that activates G protein-coupled receptor LGR2. P. Natl. Acad. Sci. U.S.A. 102 (8), 2820–2825.
Martin, M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J. 17 (1), 10.
Masood, M., Orchard, I., 2014. Molecular characterization and possible biological roles of allatotropin in *Rhodnius prolixus*. Peptides 53, 159–171.
Mckenna, D.D., Farrell, B.D., Caterino, M.S., Farnum, C.W., Hawks, D.C., Maddison, D.R., Seago, A.E., Short, A.E., Newton, A.F., Thayer, M.K., 2015. Phylogeny and evolution of *S taphyliniformia* and *S carabaeiformia*: forest litter as a stepping-stone for diversification of nonphytophagous beetles. Syst. Entomol. 40 (1), 35–60.
Mendive, F.M., Van Loy, T., Claeysen, S., Poels, J., Williamson, M., Hauser, F., Grimmelikhuijzen, C.J., Vassart, G., Vanden Broeck, J., 2005. Drosophila molting neurohormone bursicon is a heterodimer and the natural agonist of the orphan receptor DLGR2. FEBS Lett. 579 (10), 2171–2176.
Mertens, I., Meeusen, T., Huybrechts, R., De Loof, A., Schoofs, L., 2002. Characterization of the short neuropeptide F receptor from Drosophila melanogaster. Biochem. Bioph. Res. Co. 297 (5), 1140–1148.
Mirabeau, O., Joly, J.S., 2013. Molecular evolution of peptidergic signaling systems in bilaterians. P. Natl. Acad. Sci. U.S.A. 110 (22), E2028–E2037.
Mizoguchi, A., Okamoto, N., 2013. Insulin-like and IGF-like peptides in the silkmoth *Bombyx mori*: discovery, structure, secretion, and function. Front. Physiol. 4, 217.
Nässel, D.R., Winther, Å.M., 2010. Drosophila neuropeptides in regulation of physiology and behavior. Prog. Neurobiol. 92 (1), 42–104.
Orchard, I., Lange, A.B., Bendena, W.G., 2001. FMRFamide-related peptides: a multifunctional family of structurally related neuropeptides in insects. Adv. Insect Physiol. 28, 267–329.
Paluzzi, J.P., Russell, W.K., Nachman, R.J., Orchard, I., 2008. Isolation, cloning, and expression mapping of a gene encoding an antidiuretic hormone and other CAPA-related peptides in the disease vector, *Rhodnius prolixus*. Endocrinology 149 (9), 4638–4646.
Paluzzi, J.P., Vanderveken, M., O'Donnell, M.J., 2014. The heterodimeric glycoprotein hormone, GPA2/GPB5, regulates ion transport across the hindgut of the adult mosquito, Aedes aegypti. PLoS One 9 (1), e86386.
Pandit, A.A., Ragionieri, L., Marley, R., Yeoh, J.G., Inward, D.J., Davies, S.A., et al., 2018. Coordinated RNA-Seq and peptidomics identify neuropeptides and G-protein coupled receptors (GPCRs) in the large pine weevil *Hylobius abietis*, a major forestry pest. Insect Biochem. Mol. 101, 94–107.
Pascual, N., Castresana, J., Valero, M.L., Andreu, D., Bellés, X., 2004. Orcokinins in insects and other invertebrates. Insect Biochem. Mol. 34 (11), 1141–1146.
Peabody, N.C., Diao, F., Luan, H., Wang, H., Dewey, E.M., Honegger, H.W., White, B.H., 2008. Bursicon functions within the Drosophila CNS to modulate wing expansion behavior, hormone secretion, and cell death. J. Neurosci. 28 (53), 14379–14391.
Petersen, T.N., Brunak, S., von Heijne, G., Nielsen, H., 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat. Methods 8 (10), 785–786.
Predel, R., Neupert, S., Garczynski, S.F., Crim, J.W., Brown, M.R., Russell, W.K., Kahnt, J., Russell, D.H., Nachman, R.J., 2010. Neuropeptidomics of the mosquito Aedes aegypti. J. Proteome Res. 9 (4), 2006–2015.
Raina, A.K., Menn, J.J., 1993. Pheromone biosynthesis activating neuropeptide: from discovery to current status. Arch. Insect. Biochem. 22 (1-2), 141–151.
Riehle, M.A., Garczynski, S.F., Crim, J.W., Hill, C.A., Brown, M.R., 2002. Neuropeptides and peptide hormones in Anopheles gambiae. Science 298 (5591), 172–175.
Ren, G.R., Hauser, F., Rewitz, K.F., Kondo, S., Engelbrecht, A.F., Didriksen, A.K., Schjøtt, S.R., Sembach, F.E., Li, S., Søgaard, K.C., Søndergaard, L., 2015. CCHamide-2 is an orexigenic brain-gut peptide in Drosophila. PLoS One 10 (7), e0133017.
Rewitz, K.F., Yamanaka, N., Gilbert, L.I., O'Connor, M.B., 2009. The insect neuropeptide PTTH activates receptor tyrosine kinase torso to initiate metamorphosis. Science 326 (5958), 1403–1405.
Richards, S., Gibbs, R.A., Weinstock, G.M., Brown, S.J., Denell, R., Beeman, R.W., Gibbs, R., Beeman, R.W., Brown, S.J., Bucher, G., Friedrich, M., 2008. The genome of the model beetle and pest Tribolium castaneum. Nature 452, 949–955.
Rudwall, A.J., Sliwowska, J., Nässel, D.R., 2000. Allatotropin-like neuropeptide in the cockroach abdominal nervous system: myotropic actions, sexually dimorphic distribution and colocalization with serotonin. J. Comp. Neurol. 428 (1), 159–173.
Sano, H., Nakamura, A., Texada, M.J., Truman, J.W., Ishimoto, H., Kamikouchi, A., Nibu, Y., Kume, K., Ida, T., Kojima, M., 2015. The nutrient-responsive hormone CCHamide-2 controls growth by regulating insulin-like peptides in the brain of Drosophila

melanogaster. PLoS Genet. 11 (5), e1005209.
Schoofs, L., Holman, G.M., Hayes, T.K., Nachman, R.J., De Loof, A., 1991. Isolation, identification and synthesis of locustamyoinhibiting peptide (LOM-MIP), a novel biologically active neuropeptide from *Locusta migratoria*. Regul. Pept. 36 (1), 111–119.
Schoofs, L., Broeck, J.V., De Loof, A., 1993. The myotropic peptides of *Locusta migratoria*: structures, distribution, functions and receptors. Insect Biochem. Mol. 23 (8), 859–881.
Sedra, L., Lange, A.B., 2014. The female reproductive system of the kissing bug, Rhodnius prolixus: arrangements of muscles, distribution and myoactivity of two endogenous FMRFamide-like peptides. Peptides 53, 140–147.
Shull, V.L., Vogler, A.P., Baker, M.D., Maddison, D.R., Hammond, P.M., 2001. Sequence alignment of 18S ribosomal RNA and the basal relationships of adephagan beetles: evidence for monophyly of aquatic families and the placement of Trachypachidae. Syst. Biol. 50 (6), 945–969.
Sievers, F., Higgins, D.G., 2018. Clustal Omega for making accurate alignments of many protein sequences. Protein Sci. 27 (1), 135–145.
Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., Zdobnov, E.M., 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31 (19), 3210–3212.
Southey, B.R., Amare, A., Zimmerman, T.A., Rodriguez-Zas, S.L., Sweedler, J.V., 2006. NeuroPred: a tool to predict cleavage sites in neuropeptide precursors and provide the masses of the resulting peptides. Nucleic Acids Res. 34 (Suppl. l_2), W267–W272.
Stangier, J., Hilbich, C., Beyreuther, K., Keller, R., 1987. Unusual cardioactive peptide (CCAP) from pericardial organs of the shore crab *Carcinus maenas*. P. Natl. Acad. Sci. U.S.A. 84 (2), 575–579.
Starratt, A.N., Brown, B.E., 1975. Structure of the pentapeptide proctolin, a proposed neurotransmitter in insects. Life Sci. 17 (8), 1253–1256.
Suggs, J.M., Jones, T.H., Murphree, S.C., Hillyer, J.F., 2016. CCAP and FMRFamide-like peptides accelerate the contraction rate of the antennal accessory pulsatile organs (auxiliary hearts) of mosquitoes. J. Exp. Biol. 219 (15), 2388–2395.
Terhzaz, S., Rosay, P., Goodwin, S.F., Veenstra, J.A., 2007. The neuropeptide SIFamide modulates sexual behavior in Drosophila. Biochem. Bioph. Res. Co. 352 (2), 305–310.
Veenstra, J.A., 2000. Mono-and dibasic proteolytic cleavage sites in insect neuroendocrine peptide precursors. Arch. Insect. Biochem. 43 (2), 49–63 Published in Collaboration with the Entomological Society of America.
Veenstra, J.A., 2009. Allatostatin C and its paralog allatostatin double C: the arthropod somatostatins. Insect Biochem. Mol. 39 (3), 161–170.
Veenstra, J.A., 2010. Neurohormones and neuropeptides encoded by the genome of *Lottia gigantea*, with reference to other mollusks and insects. Gen. Comp. Endocr. 167 (1), 86–103.
Veenstra, J.A., 2014. The contribution of the genomes of a termite and a locust to our understanding of insect neuropeptides and neurohormones. Front. Physiol. 5, 454.
Veenstra, J.A., 2019. Coleoptera genome and transcriptome sequences reveal numerous differences in neuropeptide signaling between species (No. e27561v1). PEERJ.
Wegener, C., Herbert, Z., Eckert, M., Predel, R., 2002. The periviscerokinin (PVK) peptide family in insects: evidence for the inclusion of CAP2b as a PVK family member. Peptides 23 (4), 605–611.
Xu, G., Gu, G.X., Teng, Z.W., Wu, S.F., Huang, J., Song, Q.S., Ye, G.Y., Fang, Q., 2016. Identification and expression profiles of neuropeptides and their G protein-coupled receptors in the rice stem borer *Chilo suppressalis*. Sci. Rep.-UK 6, 28976.
Yeoh, J.G., Pandit, A.A., Zandawala, M., Nässel, D.R., Davies, S.A., Dow, J.A., 2017. DINeR: database for insect neuropeptide research. Insect Biochem. Mol. 86, 9–19.
Zandawala, M., 2012. Calcitonin-like diuretic hormones in insects. Insect Biochem. Mol. 42 (10), 816–825.
Zandawala, M., Haddad, A.S., Hamoudi, Z., Orchard, I., 2015. Identification and characterization of the adipokinetic hormone/corazonin-related peptide signaling system in Rhodnius prolixus. FEBS J. 282 (18), 3603–3617.
Zdobnov, E.M., Tegenfeldt, F., Kuznetsov, D., Waterhouse, R.M., Simão, F.A., Ioannidis, P., Seppey, M., Loetscher, A., Kriventseva, E.V., 2016. OrthoDB v9. 1: cataloging evolutionary and functional annotations for animal, fungal, plant, archaeal, bacterial and viral orthologs. Nucleic Acids Res. 45 (D1), D744–D749.
Zhang, Q., Nachman, R.J., Kaczmarek, K., Zabrocki, J., Denlinger, D.L., 2011. Disruption of insect diapause using agonists and an antagonist of diapause hormone. P. Natl. Acad. Sci. U.S.A. 108 (41), 16922–16926.