Insect Biochemistry and Molecular Biology 101 (2018) 94–107

Contents lists available at ScienceDirect

Insect Biochemistry and Molecular Biology

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

Coordinated RNA-Seq and peptidomics identify neuropeptides and G-protein coupled receptors (GPCRs) in the large pine weevil *Hylobius abietis*, a major forestry pest

Aniruddha A. Pandit[a,1], Lapo Ragionieri[b,1], Richard Marley[a], Joseph G.C. Yeoh[a], Daegan J.G. Inward[c], Shireen-Anne Davies[a], Reinhard Predel[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] Functional Peptidomics Group, Institute for Zoology, Department of Biology, University of Cologne, Zuelpicher Str. 47b, D-50674 Cologne, Germany
[c] Forest Research, Alice Holt Lodge, Farnham, Surrey GU10 4LH, UK

ARTICLE INFO

Keywords:
Insect pest
Neuropeptides
Transcriptomics
Peptidomics
Mass spectrometry
GPCR

ABSTRACT

Hylobius abietis (Linnaeus), or large pine weevil (Coleoptera, Curculionidae), is a pest of European coniferous forests. In order to gain understanding of the functional physiology of this species, we have assembled a *de novo* transcriptome of *H. abietis,* from sequence data obtained by Next Generation Sequencing. In particular, we have identified genes encoding neuropeptides, peptide hormones and their putative G-protein coupled receptors (GPCRs) to gain insights into neuropeptide-modulated processes. The transcriptome was assembled *de novo* from pooled paired-end, sequence reads obtained from RNA from whole adults, gut and central nervous system tissue samples. Data analysis was performed on the transcripts obtained from the assembly including, annotation, gene ontology and functional assignment as well as transcriptome completeness assessment and KEGG pathway analysis. Pipelines were created using Bioinformatics tools and techniques for prediction and identification of neuropeptides and neuropeptide receptors. Peptidomic analysis was also carried out using a combination of MALDI-TOF as well as Q-Exactive Orbitrap mass spectrometry to confirm the identified neuropeptide. 41 putative neuropeptide families were identified in *H. abietis,* including Adipokinetic hormone (AKH), CAPA and DH31. Neuropeptide F, which has not been yet identified in the model beetle *T. castaneum,* was identified. Additionally, 24 putative neuropeptide and 9 leucine-rich repeat containing G protein coupled receptor-encoding transcripts were determined using both alignment as well as non-alignment methods. This information, submitted to the NCBI sequence read archive repository (SRA accession: SRP133355), can now be used to inform understanding of neuropeptide-modulated physiology and behaviour in *H. abietis*; and to develop specific neuropeptide-based tools for *H. abietis* control.

1. Introduction

The large pine weevil, *Hylobius abietis,* (Linnaeus Coleoptera Curculionidae), is a notorious pest of coniferous trees of genus *Pinus.* Widespread in European forests and across the Palaeartic, the impact of *H. abietis* has been exacerbated by large-scale silviculture and the resultant clear-cutting of forests.

In managed forests, larval *H. abietis* feed and develop in the bark of root-stumps of felled conifers, whilst adult feeding can significantly impact the fitness of young replanted seedlings. Infestation can result in the girdling of bark around the stem, separating the root system from the higher foliage (Leather et al., 1999), and resulting in the death of affected trees. Moreover, regular clear felling practiced by forestry management and/or logging companies produce a huge numbers of fresh stumps, which can be exploited for oviposition by migrating *H. abietis,* leading to widespread infestation.

Coleoptera is the most-species rich Order of Class Insecta and is also a very diverse group consisting of many harmful pests as well as a small percentage of beneficial ones (Riley and Goyer, 1986; Teshler et al., 2004; Nash et al., 2008). Due to their agricultural and economical importance, there has been a progressive increase in the study of beetles at the gene level, as evidenced by the genome sequencing of several

Abbreviations: GPCRs, G-protein coupled receptors; PK, Pyrokinin; CAPA, Capability; ACP, AKH/Corazonin related peptide

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

[1] Aniruddha A. Pandit and Lapo Ragionieri contributed equally.

https://doi.org/10.1016/j.ibmb.2018.08.003
Received 27 March 2018; Received in revised form 30 July 2018; Accepted 24 August 2018
Available online 27 August 2018
0965-1748/

beetle species including the red flour beetle, *T. castaneum* (Richards et al., 2008), the mountain pine beetle, *Dendroctonus ponderosae* (Keeling et al., 2013), the burying beetle, *Nicrophorus vespilloides* (Cunningham et al., 2015), the coffee borer beetle, *Hypothenemus hampei* (Vega et al., 2015), the scarab beetle, *Oryctes borbonicus* (Meyer et al., 2016) and the Asian long horned beetle, *Anoplophora glabripennis* (McKenna et al., 2016). Recent advances in Next Generation Sequencing technologies as well as improvements in software tools for *de novo* (without genome reference) assembly, has resulted in assembly of pest beetle transcriptomes such as the Colorado potato beetle, *Leptinotarsa decemlineata* (Kumar et al., 2014), Japanese pine sawyer beetle, *Monochamus alternatus* (Lin et al., 2015) and the Seed Beetle, *Callosobruchus maculatus* (Sayadi et al., 2016).

Insect neuropeptides play an important role in regulating homeostatic processes, including growth, development, reproduction, water and ion homeostasis and stress tolerance (Nässel and Winther, 2010; Terhzaz et al., 2015). Despite acceleration in the identification of neuropeptides in various species including insects (Christie, 2008, 2014, 2015; Gard et al., 2009; Ma et al., 2010; Predel et al., 2010), the complete 'neuropeptidome' of only one beetle, the red flour beetle *T. castaneum,* has been studied so far (Li et al., 2008).

G-protein coupled receptors (GPCRs), one of the biggest protein families responsible for a wide range of regulatory processes, have been used as potential targets for drug discovery by pharmaceutical industries in the process of developing new drugs (Garland, 2013) and have been proposed as potential insecticide targets (Audsley and Down, 2015). By targeting neuropeptides and their corresponding GPCRs, potential biocides could block or over-stimulate normal activity of these proteins, producing a fatal or deleterious effect on pests. Along with neuropeptide GPCR prediction, we also performed leucine-rich repeat containing GPCRs (or LGR) identification. Three types of LGRs are known namely Type A, B and C, which are differentiated by their structural features (Van Hiel et al., 2012). LGRs are known to play significant roles in various physiological activities including development and reproduction.

By identifying the neuropeptide and the receptor sequences of *H. abietis,* novel, effective, integrated pest management strategies may be developed against this pest. To further this goal and enable targeted future investigations we have performed *de novo* transcriptome assembly of RNA sequence data from various tissue samples of *H. abietis,* followed by functional annotation and ontology. Neuropeptide and neuropeptide receptor prediction was also performed using database searches, protein family profile searches as well as GPCR prediction using hidden Markov model algorithms.

Transcriptome-predicted precursor sequences and putative mature neuropeptides were subsequently analyzed and confirmed using a combination of direct tissue profiling by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) and Q Exactive Orbitrap MS/MS.

This allowed us to identify genes of interest, which encode predicted neuropeptides and peptide hormones, and their putative GPCRs. This study will provide the first annotated transcriptome of *H. abietis* for the benefit of any future studies carried out in the field of insect pest management or general entomology.

2. Materials and methods

2.1. Insect rearing, sample preparation and RNA extraction

Adult wild *Hylobius abietis* were collected from a two-year old Sitka spruce clear-fell in a forest (Tywi North) located approximately 10 km east of Tregaron, central Wales, UK (52.2384° N, −3.7764° W). They were held in plastic containers containing an abundant supply of pine bark allowing the weevils to feed *ad libitum*. Due to the wild caught nature of the *H. abietis* it was not possible to confirm the age of specimens. The captured weevils were kept in 12:12 light:dark cycle at 18 °C and 50% relative humidity.

As the aim of the project was transcript discovery, with emphasis on neuropeptides and receptors, we sampled three transcriptomes. Whole-body mRNA captures the widest range of transcripts, and provides a baseline reference. To maximize the chances of finding relatively low abundance neuropeptide & receptor transcripts (of particular interest to this study), we additionally dissected the CNS and gut tissues. Three biological replicates of each were initially sampled, with equal numbers of males and females contributing to each sample. Insects were anaesthetized on ice and dissections were performed in ice cold Schneider's *Drosophila* Medium (1×) (Life technologies) with large blunt-nosed forceps to break through the tough cuticle. The gut sample consisted of the entire alimentary canal from mouth to anus, extracted in sections due to the nature of the dissection. The guts of three female and three male *H. abietis* were isolated, opened flat, carefully cleaned of all partly digested bark. They were then pooled together prior to lysis. Care was taken to not scrape the inner lumen of the alimentary canal, but some damage occurring during this process cannot be ruled out. Similarly the CNS sample was pooled from a combination of approximately, 10 female and 10 male heads and ventral nerve cords per replicate. Both gut and CNS tissues were homogenised with a pestle and mortar in 350 µl QIAzol without centrifugation and stored at −80 °C until RNA isolation.

For the whole insect sample, one male and one female were homogenised in 350 µl of QIAzol with a Polytron homogeniser with a 5 mm head, which was cleaned thoroughly between uses to prevent sample cross contamination, for at least 60 s until the carapace was broken up into small fragments. The sample was quick spun for 1 min in a desktop accuSpin Micro R centrifuge until all the heavy particulate matter present (fragments of the cuticle) had been removed and the supernatant could be cleanly extracted. Supernatant was stored at −80 °C until RNA isolation took place.

Total RNA was extracted with a Qiagen MiRNeasy Mini Kit (Qiagen), Sample RNA concentration and quality were then checked on an Agilent 2100 Bioanalyzer and the most robust samples (one sample per tissue) were sent forward for RNA sequencing at Glasgow Polyomics (http://www.polyomics.gla.ac.uk/).

2.2. cDNA library preparation and sequencing

RNA was processed using Illumina's TruSeq Stranded mRNA Library Prep Kit according to the manufacturer's protocols (Cat. No.: RS-122-2103, Illumina, San Diego, CA). Libraries were pooled in an equimolar manner, combined at 4 nM and prepared for sequencing as per the Illumina recommendations. Sequencing was then performed using the NextSeq® 500/550 Kit v2 (300 cycles) (Cat. No.: FC-404-2004, Illumina, San Diego, CA) with an average of 50 million reads per sample.

2.3. Transcriptome assembly

Using the Illumina NextSeq sequencing platform, a total of 188 million paired-end reads were generated from the three tissue samples. As per the protocol of Grabherr et al. (2011), the reads were concatenated into a single dataset to be used to generate a single reference transcriptome assembly. An overall 'left reads' file and an overall 'right reads' file were generated for the *de novo* assembly. 119 million paired-end reads were retained and were subsequently used to perform the assembly. These reads were initially assessed for quality using the software FastQC available at https://www.bioinformatics.babraham.ac.uk/projects/fastqc/. FastQC is run to ensure that the quality of the reads is satisfactory for further analysis. Sequencing adapters were removed with the command-line software Cutadapt (Martin, 2011) using the default parameters. Additionally, Trimmomatic (Bolger et al., 2014) was used for removal of low-quality reads (i.e. those with a Phred quality score of less than 20), duplication, artifacts and/or contamination, which may influence the overall efficiency of the eventual

assembly.

Digital normalisation was performed on the raw RNA sequence data by running the script 'insilico_read_normalization.pl' provided in the Trinity software suite. This allowed for greater efficiency in the *de novo* assembly by reducing the total number of reads without affecting the overall quality of assembly of the transcriptome (Brown et al., 2012).

Following normalisation, *de novo* transcriptome assembly was performed using Trinity (version 2.2.0), which is software for short-read assembly using the protocol set by Haas et al. (2013). Composed of three main modules - Inchworm, Chrysalis and Butterfly, Trinity joins overlapping reads to construct longer contigs. Assembly was performed with a default k-mer size of 25. All resulting overlapped contigs were then clustered into components using de Bruijn graphs. From these graphs, full-length unigene transcripts and their alternatively spliced forms were determined and generated.

2.4. Functional annotation and ontology

To better comprehend *de novo* transcriptome data, it is essential to perform functional annotation. Blast2GO version 4.1.9 (Conesa et al., 2005), a popular bioinformatics platform, was used to perform BLAST (Altschul et al., 1990) searches, mapping and annotation of the assembled transcripts.

All assembled transcripts (or putative genes as they will be referred to henceforth) were searched against the NCBI non-redundant (nr) protein database (ftp://ftp.ncbi.nih.gov.blast.db) and the SwissProt database as part of functional annotation. The genes were further searched against the Kyoto Encyclopedia of Genes and Genomes (KEGG), Gene Ontology (GO) and Interpro databases using BLASTx with an e-value cut-off set of 1e-5.

The BUSCO v3 (Benchmarking Universal Single-Copy Orthologs) library was used to assess the completeness of the assembled transcriptome (Simao et al., 2015) by comparing each putative gene obtained from the assembly to near-universally-distributed single-copy genes in the insect set obtained from the OrthoDB v9 available at http://www.orthodb.org (Zdobnov et al., 2016).

2.5. Neuropeptide prediction

All neuropeptide sequences, used for alignment and comparison to the transcript data from *H. abietis,* were downloaded from DINeR, available from http://www.neurostresspep.eu/diner/infosearch (Yeoh et al., 2017), unless stated otherwise. Sequence alignments were generated using Clustal Omega (http://www.ebi.ac.uk/Tools/msa/clustalo/) and sequence logos were generated using Weblogo 3.4 (Crooks et al., 2004).

The transcript data from the assembly was used to run a BLASTx search against the entire set of neuropeptides downloaded from DINeR. Parallel to this, using the software GetORFs (Cock et al., 2013) that is available at http://emboss.sourceforge.net/apps/cvs/emboss/apps/getorf.html, Open Reading Frames or ORFs were extracted from the transcript data. These ORFs were then compared to mature neuropeptide sequences from DINeR as well as the precursor sequences from *T. castaneum* (Li et al., 2008) to identify putative genes. Once all the sequences were collated from the above methods, each sequence was run through the SignalP software (Petersen et al., 2011) to predict signal peptide cleavage. Mono and dibasic cleavage sites are identified by combinations of K and R based on rules set out by Veenstra (2000). After neuropeptide identification, the NCBI database was queried using BLASTx and BLASTp to identify homologous sequences in other insects.

2.6. MALDI-TOF mass spectrometry

Sample preparation for direct tissue profiling of CNS (brain, thoracic and abdominal ganglia), frontal ganglion, retrocerebral complex and abdominal nerves (MS1) was performed using the individual tissue samples from single specimens as described in (Schachtner et al., 2010) but using modified insect saline solution at pH 7.4 (7.50 g/l NaCl, 0.20 g/l KCl, 0.20 g/l $CaCl_2$, 0.10 g/l, $NaHCO_3$). As matrix we used 10-mg/ml α-cyano-4-hydroxycinnamic acid (α-CHCA; ethanol/acetonitrile/water, 60/36/4) or 10 mg/ml 2,5-dihydroxybenzoic acid (DHB; water/acetonitrile/formic acid, 79/20/1); subsequent analyses were performed using the ultrafleXtreme MALDI-TOF/TOF mass spectrometer (Bruker Daltonik GmbH, Bremen, Germany) under manual control in reflectron positive-ion mode.

Before pipetting the α-CHCA matrix onto the tissues, the stock solution was dissolved in 50% methanol/water in a ratio of 2:1. For external calibration we used the Bruker peptide standard for the lower mass range (m/z 600–4000) and a mixture of bovine insulin, glucagon and ubiquitin for the higher mass range (m/z 3000–10,000). Peptide ion fragmentations of *H. abietis* peptides were mostly performed on the MALDI TOF/TOF ABI 4800 Proteomics Analyzer (Applied Biosystems Framingham, MA) in gas off mode using α-CHCA as matrix.

Mass spectra obtained with the ultrafleXtreme were analyzed using flexAnalysis 3.4 software package (Bruker Daltonik) and MSMS spectra obtained with the ABI 4800 Proteomics Analyzer were analyzed using Data Explorer v. 4.3 (Applied Biosystems).

2.7. Q-Exactive Orbitrap mass spectrometry

CNS and retrocerebral complex of single specimens were dissected in insect saline and transferred in 30 µl extraction solution (50% methanol, 49% H_2O, 1% formic acid). Subsequently, tissues were disintegrated in an ultrasonic bath for 5 min (Transonic 660/H, Elma Schmidbauer GmbH, Hechingen, Germany) and three times with an ultrasonic-probe for 5 s (BandelinSono HD 200, Bandelin electronic GmBH, Berlin, Germany). The samples were then centrifuged for 20 min at 15,000 rpm at 4 °C and the remaining methanol was evaporated from the supernatants in a vacuum concentrator before the extracts were stored at −20 °C. Directly before Q-Exactive Orbitrap analyses, samples were desalted using self-packed Stage Tip C18 Spin columns as described in (Rappsilber et al., 2007).

Subsequently, peptides were separated on an EASY nanoLC II system (Thermo Fisher Scientific, Waltham, MA, USA) using in-house packed RP-C18-columns 50 cm (Reprosil 1.9 µm, pore diameter 60A°; Maisch GmbH, Ammerbuch-Entringen, Germany) and a binary buffer system (A: 0.1% FA; B: 80% ACN, 0.1% FA). Running conditions were as follows: linear gradient from 7 to 23% B in 46 min, from 23 to 54% B in 2 min and final washing from 54 to 85% B in 6 min (45 °C, flow rate 250 nl/min).

Finally the gradients were re-equilibrated for 10 min at 5% B. The HPLC was coupled with a Q-Exactive Plus (Thermo Scientific, Bremen, Germany). The Orbitrap mass spectrometer was run in positive mode and spectra were obtained in a data dependent mode with full MS1 acquisition at 70,000 resolution with automatic gain control target (AGC target) at 3e6, maximum injection time (maximum IT) at 80 ms and scan range between 400 and 6000 m/z. The top 10 most intense peaks were selected for HCD fragmentation and measured in mass analyzer with a resolution of 35,000, AGC target at 1 e6, maximum IT at 120 ms and dynamic exclusion set at 25 s.

2.8. Q-Exactive data analysis

Raw data were analyzed using PEAKS 8.0 (PEAKS Studio 8.0, BSI, ON, Canada). MS/MS spectra were matched against an internal database containing all the identified neuropeptide precursors obtained from the transcriptome assembly of *H. abietis* (Supplementary Table 1). Searches were started using parent error mass tolerance of 0.1 Da and fragment mass error tolerance of 0.2 Da in non-enzyme mode. The false discovery rate (FDR) was determined by a decoy database search using a threshold below 1% as implemented in PEAKS (Zhang et al., 2012). A maximum of six post-translational modifications (PTMs) for each

peptide was allowed. Conversion of C-terminal glycine (amidation) was selected as fixed post-translational modification (PTM) while oxidation at methionine, N-terminal acetylation, pyroglutamate from glutamine, pyroglutamate from glutamic acid, sulfation, and disulfide bridges were included as variable PTMs. Each spectrum having a peptide score (−10lgP) greater than 20 was manually checked for final validation which is equivalent to a P-value of 1%.

2.9. Neuropeptide receptor prediction

A combination of Bioinformatics techniques was used to create a data analysis pipeline to identify neuropeptide GPCRs from the assembled transcript data. This pipeline included alignment methods (BLAST, Pfam Database search) and non-alignment techniques (TMHMM).

Using BLAST2GO, sequence alignment was performed on the nucleotide sequences by performing BLASTx against the NCBI non-redundant (nr) database as well as the Swissprot database with an e-value cutoff of 1e-5.

ORFs were extracted from the transcript data using GetORF and saved as FASTA files to be used in the further analyses. This file was then used for two separate analyses.

In the first analysis, the GPCRHMM server [available at http://gpcrhmm.sbc.su.se/] was used to predict GPCRs from the amino acid sequences in the FASTA file. The prediction was performed using a Hidden Markov Model used to profile GPCRs. This software lists all possible GPCRs using HMM profiles and generic trans-membrane detectors of known GPCRs and non-GPCRs (Wistrand et al., 2006). All possible GPCRs were filtered out and the amino acid sequences for these were then BLASTed against the nr and SwissProt database using BLASTp. From the results of these BLAST searches, putative GPCRs were noted and saved.

The final part of the pipeline involves searching for putative GPCRs by comparing the ORFs in the FASTA file created above against Pfam, a database of protein families and domains (Finn et al., 2009). Profile searches of each of the ORFs were performed against the Pfam-A HMM database to identify appropriate Transmembrane 7 (tm7) protein families. These searches were done using the tool PfamScan (Li et al., 2015), available from ftp://ftp.ebi.ac.uk/pub/databases/Pfam/Tools/,

The putative GPCR encoding sequences, which were identified by the two methods above and were further verified using BLASTp searches against the nr and SwissProt databases.

The conserved domain database or CDD (Marchler-Bauer et al., 2016) was searched using CD-Search tool (Marchler-Bauer and Bryant, 2004) available at NCBI (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) and used to identify and highlight the seven transmembrane units in each of the putative GPCR encoding sequences found above.

Multiple sequence alignment was carried out with the online T-Coffee software suite (http://tcoffee.crg.cat/apps/tcoffee/index.html), using the PSI/TM-Coffee server. This particular server performs multiple sequence alignment and highlights all the found transmembrane helices. This is done by homology extension against the UniRef database (Chang et al., 2012).

The amino acid sequences of the neuropeptide GPCRs from *H. abietis* along with published neuropeptide GPCR sequences of the well-studied *Tribolium castaneum* (representing Coleoptera); *Drosophila melanogaster* (representing Diptera) and *Bombyx mori* (representing Lepidoptera) and *Anoplophora glabripennis* (representing Coleoptera where certain receptors were not found in *T. castaneum*) were grouped together in a FASTA file. This file was then imported into CLC Genomics Workbench 9 (https://www.qiagenbioinformatics.com/) and alignments were created using the default settings. The alignments were further used to construct phylogenetic trees using the Neighbour Joining method with 1000-fold bootstrap resampling using the tree creation module of CLC Genomics Workbench 9.

Table 1
Summary statistics of sequencing and *de novo* transcriptome assembly data of Hylobius abietis.

Read Processing	
Raw Reads	188016997
Post-Trimming	173904861
Post Normalisation	11998056
Trinity *de novo* Assembly	
Total trinity 'genes'	106862
Total trinity transcripts	163805
Percent GC	37.05
Contig N50	1496
Median contig length	544
Average contig	967.57
Total assembled bases	158493471

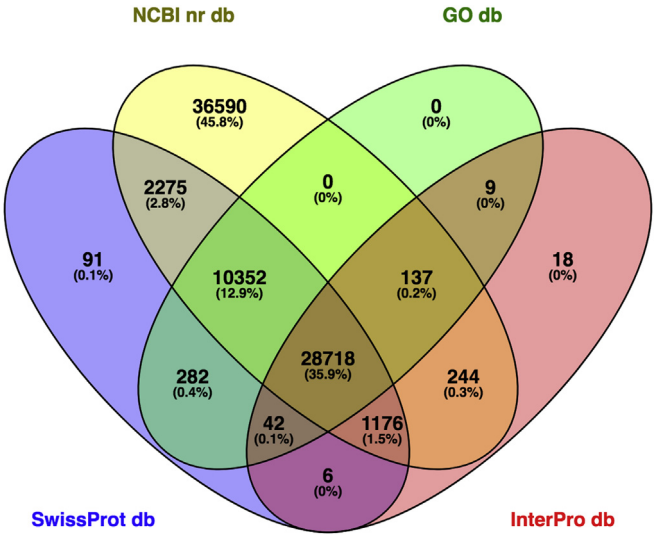

Fig. 1. The results of the annotation of the sequenced transcripts against the nr database, the SwissProt database, the InterPro database and the Gene Ontology (GO) database and the overlap between the databases.

3. Results

3.1. De novo assembly

Initial assembly produced 428410 contigs with an average length of 584.8 bp and an N50 of 955. Removal of short contigs (< 300 bp) reduced this to 217308 contigs with an average length of 918.43 bp and an N50 of 1411. Contigs with low expression were removed by applying an FPKM (Fragments Per Kilobase Million) value cut-off of 1. This filtration produced 163805 contigs with an average length of 967.57 bp and an N50 of 1496 bp. This filtered transcriptome was used in subsequent stages of annotation and analysis. The summary of the assembly is shown in Table 1.

3.2. Assessment of completeness

An additional assessment of the quality of the *H. abietis* transcriptome was performed using Benchmarking Universal Single-Copy Orthologs or BUSCO libraries. The library used was the Insecta orthologous genes library, which consists of 1658 single-copy, annotated and conserved insect orthologs. A total of 1603 (96.7%) complete BUSCOs were found consisting of 800 (48.3%) complete and single-copy hits and 803 (48.4%) duplicate hits. 37 (2.2%) of hits were only partially recovered or fragmented BUSCO hits. 18 (1.1%) BUSCO groups were missing. When the same BUSCO analysis was run on the assemblies for the individual tissues the number of complete and duplicated hits were

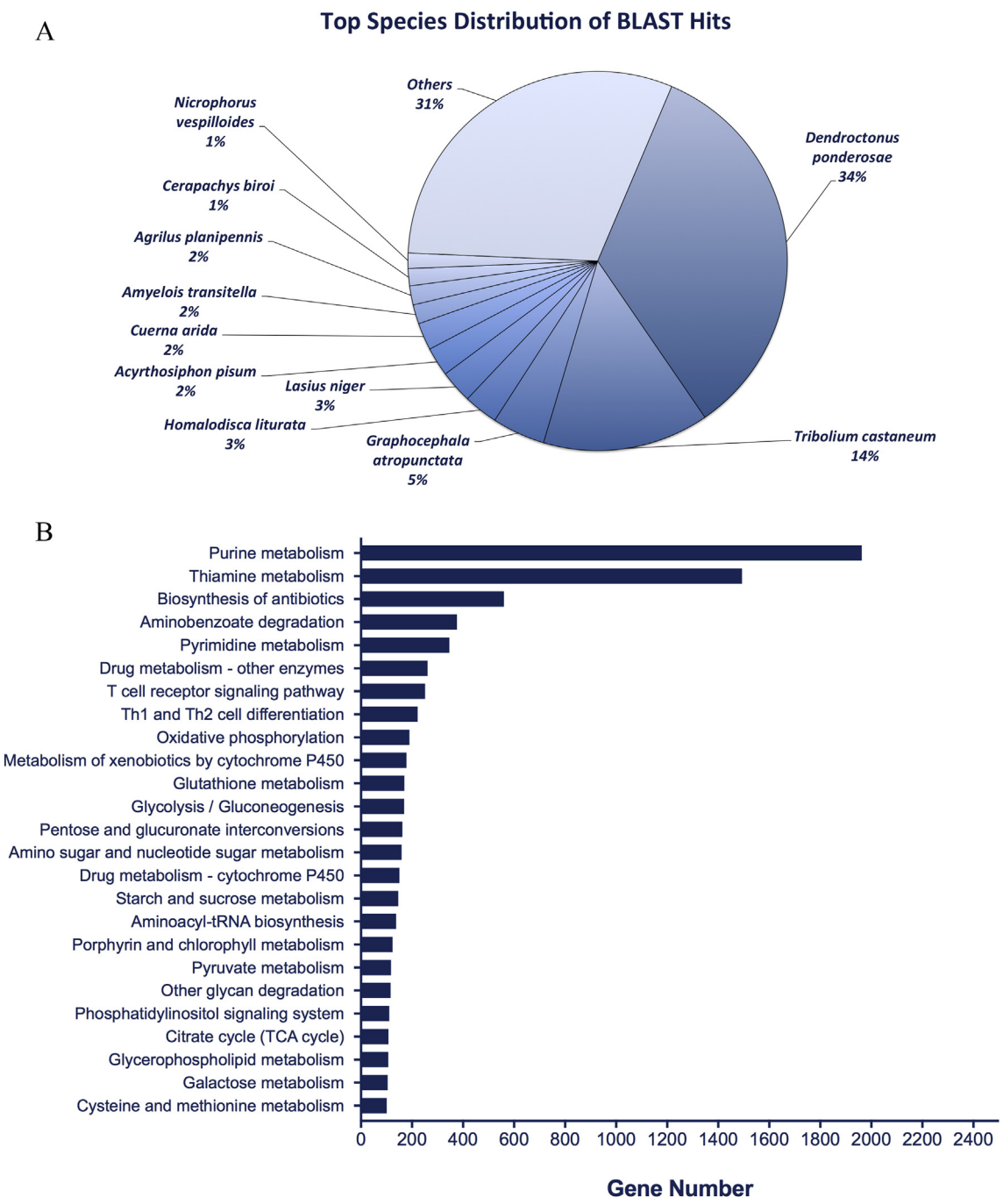

Fig. 2. Summary of the data analysis performed on the transcriptome assembly of the *Hylobius abietis* including: (A) Top Species distribution of BLAST Hits against genes in the NCBI nr (non-redundant) database, (B) Top 25 KEGG Pathways distribution with number of genes per pathway.

98.2% (CNS), 97.4% (Gut) and 97.3% (Whole Insect) with 44.2%, 42.6% and 31.1% duplicates respectively.

3.3. Functional annotation and further analysis

Functional annotation was carried out by alignment of the transcripts against databases including NCBI non-redundant (nr) protein database, SwissProt, InterPro, the GO database and the KEGG database. All BLAST alignments were carried out with an e-value cutoff of 1e-5. The results of these alignments indicate that out of a total 163805 transcripts, 79492 (48.5%) were annotated against the nr database while 42942 (26.2%) were annotated against the SwissProt database. A total of 30350 (18.5%) transcripts were annotated against the InterPro database, while 39540 (24.1%) were annotated against the GO database. The overlap between the databases is illustrated in Figure (1). Although there was some additional transcript annotation from the other resources, NCBI nr was clearly the best single performer.

Further analysis of the transcripts included that of the Top species distribution and E-value distribution. The maximum number of blast hits was found against *Dendroctonus ponderosae* (Coleoptera, Curculionidae) also a major insect pest of pine forests, with a total of about 34% of genes showing similarity with those of *H. abietis*. Next was the well-studied red-flour beetle, *T. castaneum* (Coleoptera, Tenebrionidae) with 14% similarity to *H. abietis* genes. These and all other top-species distribution are illustrated in Fig. 2(A). Around 48.5% of the BLAST hits with the nr database had E-values equal to zero. The majority of hits ranged from 1e-6 to 1e-

Table 2
Neuropeptide precursors identified from H. abietis. The result of the homology search is also shown. ORF: open reading frame, 1 Not full length, 2 Sequence obtained from transcriptome of A. grandis (Run: SRR610743).

Neuropeptide	Abbreviation	Transcript name	ORF(aa)	Homology search with known protein		
				Species	E-value	Accession No.
Adipokinetic hormone	AKH	TRINITY_DN79477_c0_g1_i1	73	*Tribolium castaneum*	8.00E-15	NP_001107818.1
AKH/Corazonin-related peptide	ACP	TRINITY_DN45440_c0_g1_i1	87	*Anoplophora glabripennis*	3.00E-08	XP_018564238.1
Allatostatin B	AstB	TRINITY_DN95192_c0_g1_i2	200	*Tribolium castaneum*	9.00E-40	XP_008191728.1
Allatostatin C	AstC	TRINITY_DN89390_c0_g1_i1	116	*Leptinotarsa decemlineata*	8.00E-15	AIW62334.1
Allatostatin CC	AstCC	TRINITY_DN90454_c4_g2_i5	144	*Drosophila melanogaster*	3.00E-07	NP_609483.2
Allatotropin	AT	TRINITY_DN84904_c0_g1_i1	156	*Dendroctonus ponderosae*	3.00E-28	XP_019758395.1
Bursicon	Burs	TRINITY_DN29559_c0_g1_i1	177	*Tribolium castaneum*	4.00E-92	NP_001107779.1
Calcitonin-B-1	CalB-1	TRINITY_DN80855_c0_g7_i1	145 1	*Tribolium castaneum*	1.00E-05	XP_015834921.1
Calcitonin-B-2	CalB-2	TRINITY_DN80855_c0_g8_i2	133 1	*Tribolium castaneum*	6.00E-10	XP_015834921.1
Capability	CAPA	TRINITY_DN91637_c0_g2_i2	118	*Anthonomus grandis*	1.00E-39	SRA0599592
CCHamide-1a	CCH-1a	TRINITY_DN90167_c0_g1_i1	156	*Tribolium castaneum*	2.00E-10	XP_008201341.1
CCHamide-1b	CCH-1b	TRINITY_DN90167_c0_g1_i2	147	*Tribolium castaneum*	2.00E-13	XP_008201341.1
CCHamide-2	CCH-2	TRINITY_DN93068_c0_g1_i2	120	*Tribolium castaneum*	3.00E-18	XP_008190391.1
CNMamide	CNMa	TRINITY_DN90746_c0_g1_i4	139	*Tribolium castaneum*	8.00E-04	XP_015833687.1
Crustacean Cardio-Active Peptide	CCAP	TRINITY_DN73446_c0_g1_i2	148	*Tribolium castaneum*	2.00E-54	XP_008201233.2
Diuretic Hormone 31	DH31	TRINITY_DN96510_c0_g3_i3	122	*Tribolium castaneum*	4.00E-27	EEZ99367.2
Diuretic Hormone 44	DH44	TRINITY_DN94346_c0_g3_i1	162	*Tribolium castaneum*	9.00E-21	EFA12665.2
Ecdysis-triggering hormone	ETH	TRINITY_DN55094_c0_g1_i1	111 1	*Tribolium castaneum*	1.00E-13	EFA07492.2
Eclosion hormone	EH	TRINITY_DN11597_c0_g1_i1	78	*Tribolium castaneum*	4.00E-23	XP_008190384.1
Elevenin	Ele	TRINITY_DN82774_c0_g1_i7	116	*Dendroctonus ponderosae*	9.00E-18	XP_019770585.1
FMRFamide	FMRFa	TRINITY_DN86238_c0_g2_i2	137	*Tribolium castaneum*	2.00E-10	XP_008191572.1
Glycoprotein hormone alpha 2	GPA2	TRINITY_DN81517_c0_g1_i2	123	*Tribolium castaneum*	4.00E-52	NP_001164244.1
Glycoprotein hormone beta 5	GPB5	TRINITY_DN85603_c0_g1_i2	164	*Tribolium castaneum*	2.00E-69	NP_001280517.1
Insulin-like Peptide	ILP	TRINITY_DN79383_c0_g3_i1	126	*Tribolium castaneum*	2.00E-06	EFA02796.2
Ion transport peptide a	ITP-a	TRINITY_DN78410_c0_g1_i2	127	*Tribolium castaneum*	4.00E-42	NP_001076808.1
Ion transport peptide b	ITP-b	TRINITY_DN78410_c0_g1_i1	139	*Tribolium castaneum*	3.00E-44	XP_008195073.1
ITG-like peptide	ITG	TRINITY_DN92590_c0_g7_i1	212	*Agrilus planipennis*	4.00E-118	XP_018325057.1
Myosuppressin	MS	TRINITY_DN71981_c0_g1_i3	95	*Tribolium castaneum*	4.00E-11	EFA12055.1
Natalisin	NTL	TRINITY_DN66744_c0_g1_i1	142	*Tribolium castaneum*	2.00E-25	XP_008200697.1
Neuroparsin	NP	TRINITY_DN87374_c0_g1_i1	105	*Tribolium castaneum*	1.00E-17	XP_008193973.1
Neuropeptide F (long transcript)	NPF-1	TRINITY_DN97392_c0_g1_i4	133	*Dendroctonus ponderosae*	2.00E-65	XP_019762446.1
Neuropeptide F (short transcript)	NPF-2	TRINITY_DN97392_c0_g1_i8	96	*Dendroctonus ponderosae*	1.00E-34	XP_019762446.1
Neuropeptide-like precursor 1 isoform a	NPLP1-a	TRINITY_DN100327_c0_g2_i2	362	*Tribolium castaneum*	5.00E-46	KYB25699.1
Neuropeptide-like precursor 1 isoform b	NPLP1-b	TRINITY_DN100327_c0_g2_i5	316	*Tribolium castaneum*	2.00E-59	EFA09268.2
NVP-like	NVP	TRINITY_DN96596_c0_g2_i4	282	*Tribolium castaneum*	5.00E-48	XP_008196925.1
Orcokinin B	OK-B	TRINITY_DN81862_c0_g3_i1	211 1	*Blatella germanica*	1.00E-34	AKR13996.1
Partner of bursicon	Pburs	TRINITY_DN72229_c0_g2_i1	136	*Tribolium castaneum*	4.00E-67	NP_001107780.1
Proctolin	Proc	TRINITY_DN93986_c1_g4_i3	88	*Dendroctonus ponderosae*	3.00E-30	XP_019763695.1
Prothoracicotropic hormone	PTTH	TRINITY_DN16009_c0_g1_i1	114 1	*Tribolium castaneum*	2.00E-15	AKN79607.1
Pyrokinin-1	PK-1	TRINITY_DN98133_c0_g2_i5	154	*Dendroctonus ponderosae*	5.00E-59	XP_019770936.1
Pyrokinin-2	PK-2	TRINITY_DN79095_c0_g1_i1	155	*Aethina tumida*	1.00E-22	XP_019871287.1
RYamide	RYa	TRINITY_DN72700_c0_g1_i3	120	*Tribolium castaneum*	7.00E-21	NP_001280530.1
Short neuropeptide F	sNPF	TRINITY_DN91247_c0_g2_i2	103	*Tribolium castaneum*	8.00E-05	DAA34847.1
SIFamide	SIFa	TRINITY_DN66935_c0_g1_i1	79	*Tribolium castaneum*	1.00E-20	XP_001814498.1
Sulfakinin	SK	TRINITY_DN80110_c0_g2_i1	114	*Tribolium castaneum*	3.00E-17	XP_008194373.1
Tachykinin-related peptide	TRP	TRINITY_DN96824_c1_g1_i2	271	*Tribolium castaneum*	4.00E-64	KYB25859.1
Trissin	Tris	TRINITY_DN61822_c0_g1_i2	113	*Dendroctonus ponderosae*	2.00E-52	XP_019766470.1
Inotocin (Vasopressin-Like)	VPL	TRINITY_DN75890_c0_g3_i1	149	*Tribolium castaneum*	4.00E-63	NP_001078831.1

Table 3
Neuropeptide products of H. abietis identified using Mass Spectrometry analyses. + Peptide confirmed by MSMS; (+) MSMS missing (MALDI; fingerprint data only) or MSMS partial sequence only (Orbitrap).

Neuropeptide	Abbreviation	Peptide Sequence	[M+H]+, m/z	MALDI-TOF	Orbitrap
Adipokinetic hormone	AKH	pQVNFSPNW-NH2	973.44	+*	+
AKH/Corazonin-related peptide	ACP	pQVTFSRNWEPSG-NH2	1389.65	+	(+)
Allatostatin B	AstB-1	GWKTDLPMW-NH2	1132.56	+	+
	AstB-2	GWSNLHSGW-NH2	1042.49	+	+
	AstB-3	AWQRLQGGW-NH2	1100.57	+	+
	AstB-4	NWKSFSDGW-NH2	1125.51	+	(+)
	AstB-5	SNKWEKFRGSW-NH2	1423.72	+	+
	AstB-6 [Q]	QPAWNNLKGLW-NH2	1325.68	(+)	(+)
	AstB-6 [pQ]	pQPAWNNLKGLW-NH2	1308.68	+	–
Allatostatin C	AstC [pQ]	pQSRYRLcYFNPVScF-OH	1863.83	+	–
	AstC [Q]	QSRYRLcYFNPVScF-OH	1880.87	(+)	–
Allatostatin CC	AstCC	GHQSSLTNGQNKARIYWRcYFNAVTcF-OH	3162.47	–	(+)
Allatotropin	AT-1	IGENLFTARGY-NH2	1239.65	(+)	+
	AT-2	SSKTEMALLSARRF-NH2	1595.87	(+)	+
Capability	PVK	SAPQFLLNPRV-NH2	1240.72	+	+
	CAPA-PP-1	NDKMWKIRPLENLQ-OH	1784.94	+	–
	CAPA-tryptoPK	SDPKAADLWFGPRI-NH2	1571.83	+	(+)
CCHamide-1	CCH-1a	SRGScLSYGHAcWGAH-NH2	1688.71	(+)	(+)
	ext. CCH-1a	RYEIKKSRGScLSYGHAcWGAH-NH2	2506.2	(+)	+
	ext. CCH-1a	GRYEIKKSRGScLSYGHAcWGAH-NH2	2563.22	(+)	+
	CCH-1b	ScLSYGHAcWGAH-NH2	1388.58	+	–
CCHamide-2	CCH-2	KRGcANFGHScYGGV-NH2	1552.69	–	+
Crustacean cardioactive peptide	CCAP	PFcNAFTGc-NH2	956.38	(+)	–
Diuretic hormone 31	DH31	GIDLGLGRGFSGSQAAKHLMGLAAANFAGGP-NH2	2940.53	(+)	+
Elevenin	Ele	KVINcRPLPQccRKYVFSPGcRGVAA-OH	2861.46	(+)	–
Extended FMRFamides	FMRFa-1	NGDKFLRF-NH2	995.54	+	(+)
	FMRFa-2	AGRQEKNDHFIRF-NH2	1616.84	(+)	+
	FMRFa-3	DKSSFLRF-NH2	998.54	+	–
	FMRFa-4	HENFLRF-NH2	961.5	+	(+)
	ext. FMRFa-4	ETVSSSEALKRNSNFMRF-NH2	2149.11	(+)	+
	FMRFa-5	NSNFMRF-NH2	914.43	+	+
Myosuppressin	MS [pQ]	pQDVDHVFLRF-NH2	1257.64	+	+
	MS [Q]	QDVDHVFLRF-NH2	1274.66	+	+
Natalisin	Nat-1	SASQDDMGPFWANR-NH2	1580.69	–	+
	Nat-2	DPTYTQDRSFAEEPHWILLRKYSEYEPFFTSR-NH2	4007.93	–	(+)
	Nat-3	DYGTDAEFDPAFFAAR-NH2	1791.79	–	–
Neuropeptide F	NPFa51-65 [pQ]	pQELDNLYSPRSRPRF-NH2	1859.95	–	+
	NPFa51-65	QELDNLYSPRSRPRF-NH2	1876.97	–	+
	NPFb14-28	MKLDQLYSSIARPRF-NH2	1824	–	+
Neuropeptide-like precursor	NPLP-PP1	KNGQPPLMPLPLENQ-OH	1675.88	–	+
	NPLP-PP2	GIEALARNGDLH-OH	1265.66	–	+
	NPLP-PP2-3	GIEALARNGDLHRRQNYQTLLDSLDF-OH	3015.54	(+)	+
	NPLP-1	NVASLARAYNFPVGSTAGF-NH2	1941	+	+
	NPLP-2	SIGSLAKNGDLPYSY-NH2	1583.81	+	+
	NPLP-3	NIQSLARDGAL-NH2	1156.64	+	+
	NPLP-4	SADYLDVSQ-NH2	996.46	(+)	+
	NPLP-PP4	NIQSVKAQQRN-OH	1285.7	(+)	+
	NPLP-5	HVAALARLGWLPSYRSVRRFNRS-NH2	2711.52	(+)	+
	NPLP-PP5	YLLQPAVDNILLRKVFRHPRMTL-OH	2793.61	(+)	(+)
	NPLP2-PP1	FLGRIPQMGKPRTTAIPKSRRYH-OH	2710.52	(+)	(+)
NVP-like	NVP	DQPNHNGPEENIGRMSYFNNKPTSAI-OH	2930.34	+	+
Proctolin	Proc	RYLPT-OH	649.38	(+)	–
Pyrokinin 1 precursor	tryptoPK	FDAPNDRGPKEVSPMWFGPRI-NH2	2415.2	+	+
	PK1-1	NPGDAKVYRNDKEQSSLGLL DVLRESPLVVVAVNEASKQHNFTPRL-NH2	5102.71	–	+
	PK1-2	SNNPFAPRL-NH2	1014.55	+	+
	PK1-3	NNYNPFSPRL-NH2	1220.62	+	+
Pyrokinin 2 precursor	tryptoPK2	LVPSNNADNTIDSNGQHQNQEKYLSSMWFGPRM-NH2	3777.75	+	+
	PK2-1	GNTHNFTPRL-NH2	1155.6	+	+
	PK2-2	SAAFSPRL-NH2	847.48	+	+
	PK2-3	GRHFSPYSPRL-NH2	1315.7	+	+
	PK2-3+4	SAAFSPRLGRGRHFSPYSPRL-NH2	2358.27	(+)	+
RYamide	RYa-1	MPGSTHAFQQMVRY-NH2	1651.78	(+)	+
	RYa-2	ADVFYLGPRY-NH2	1199.62	(+)	+
Short neuropeptide F	sNPF	AGRSPQLRLRF-NH2	1299.8	+	+
SIFamide	SIFa	TYRKPPFNGSIF-NH2	1425.76	+	+
Sulfakinin	SK-1 (pQ)	pQQSNSDDYGHLRF-NH2	1548.68	+	+
	SK-1 (Q)	QQSNSDDYGHLRF-NH2	1565.71	(+)	+
	SK-2	DEQFDDYGHMRF-NH2	1558.64	+	+ -

(*continued on next page*)

Table 3 (*continued*)

Neuropeptide	Abbreviation	Peptide Sequence	[M + H] +, m/z	MALDI-TOF	Orbitrap
Tachykinin-related peptides	TRP-1	VPSGFTGVR-NH2	918.526	–	+
	TRP-2 + 6	APSGFFGMR-NH2	968.48	+	+
	TRP-3 + 5	APMGFVGMR-NH2	964.49	+	+
	TRP-PP-1	PWNYPSYTYPEEMY-OH	1839.75	(+)	–
	TRP-4	AQMGFFGMR-NH2	1043.49	(+)	+
	TRP-PP-2	YLIEPGYQGD-OH	1154.53	(+)	
	TRP-7	RPSGFFGMR-NH2	1053.54	(+)	+
	TRP-PP-3	DFENPAGYFEDN	1417.55	(+)	
	TRP-8	YPYEFRGKFVGVR-NH2	1616.87	+	+
Inotocin (Vasopressin-Like)	Inotocin (VPL)	cLITNcPRG-NH2	973.48	(+)	–

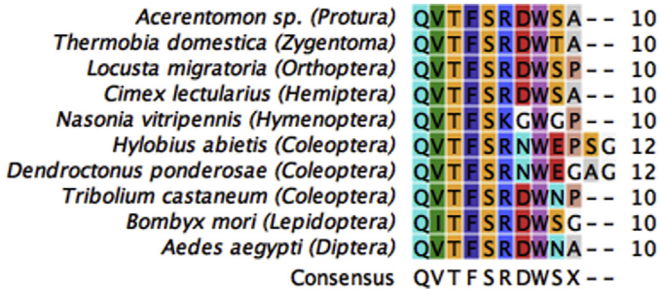

Fig. 3. Multiple sequence alignment of unusual AKH/corazonin-related peptide (ACP) of *Hylobius abietis* with other insect species from various orders. All peptide sequences except those of Protura and Zygentoma have been biochemically confirmed using Mass Spectrometry (Siegert, 1999; Li et al., 2008; Roller et al., 2008; Hauser et al., 2010; Predel et al., 2010; Derst et al., 2016; Predel et al., 2017).

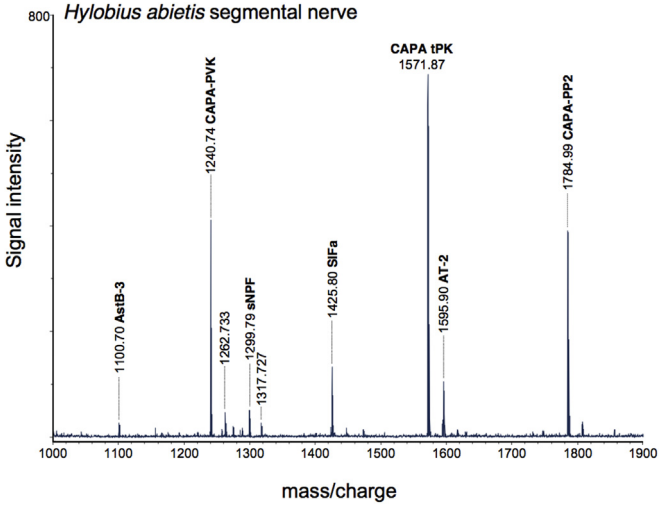

Fig. 4. Direct tissue profiling by MALDI-TOF mass spectrometry (m/z 600–1900) of an abdominal segmental nerve containing axons of abdominal neurosecretory cells (Va cells). Ion signals mass identical with predicted CAPA-peptides are labeled; the identity of these peptides was confirmed by MSMS. In addition, ion signals likely representing AstB, SIFa, sNPF, and AT are labeled.

100. It should be noted the percentage of blast hits with non-insect species, including microorganisms and plant species was very low < 0.5%.

To understand the functional properties and interactions of the gene sequences in the *H. abietis* transcriptome, Blast2Go was used to map each of the annotated transcripts against the Kyoto Encyclopedia of Genes and Genomes (KEGG) database for a pathway-based analysis. This resulted in a total of 11767 genes being associated with at least one KEGG pathway out of a total of 136 pathways. Fig. 2(B) shows the top 25 KEGG pathways with the number of genes in each pathway.

3.4. Identification of neuropeptides

The neuropeptide prediction results are shown in Table 2. Neuropeptides that were not found in the search include Allatostatin A, Allatostatin CCC, Corazonin, Kinin, Sex peptide and Pigment-dispersing factor, none of which have been identified in any beetle species to date.

Using a combination of MALDI direct tissue profiling and Orbitrap mass spectrometry we identified more than 70 peptides from 26 precursor genes, with 62 putative bioactive neuropeptides. Most of these peptides were further sequence-confirmed by MS/MS fragmentation (Table 3). With a few exceptions, the neuropeptides and their precursors identified in *H. abietis* are similar to those found in *T. castaneum* (Table 2). A comparison of the putative neuropeptide precursors identified in *H. abietis* with the ones found in *T. castaneum* and the closely related *D. ponderosae* is illustrated in Supplementary Table 2.

The neuropeptides from the insect allatostatins families B, which plays an important role in ecdysis (Kim et al., 2006; Yamanaka et al., 2010) and C, which is myoinhibitory in nature (Price et al., 2002; Matthews et al., 2007) are found in *Hylobius* and *Tribolium* but not found in *Dendroctonus*. The neuropeptides, Elevenin, involved in melanisation (Uchiyama et al., 2018) and Neuropeptide F (discussed below) are found in *Hylobius* and *Dendroctonus*, but not found in *Tribolium*. Precursors for the multifunctional FMRFamide (Sedra and Lange, 2014; Suggs et al., 2016; Kim et al., 2006) peptides, along with the visceral and heart muscle inhibitory neuropeptides Myosuppressins (Orchard et al., 2001; Nässel and Winther, 2010) were also identified as were the diuretic hormones DH31 and DH44 which are involved in fluid secretion and muscle contractions in various organs (Te Brugge et al., 2008; LaJeunesse et al., 2010; Orchard, 2009). Among other neuropeptides identified were the glycoprotein hormone bursicon and partner of bursicon that induce cuticle tanning after metamorphosis (Luo et al., 2005; Zitnan and Adams, 2012). Precursors with neuropeptides that clearly differ from those of related insects are discussed in detail.

AKH/corazonin-related peptide or ACP is a structural intermediate between the neuropeptides AKH and corazonin discovered by Hansen et al. (2010) in various species including *A. gambiae* (Diptera), *B. mori* (Lepidoptera) as well as *T. castaneum* (Coleoptera). The putative precursor identified in the *H. abietis* transcriptome encodes a single ACP (see Supplementary Table 1), confirmed by mass spectrometry (see Table 3). Notably, the sequence of *Hylobius*-ACP contains two additional amino acids at the C-terminus. A BLAST search revealed a similar extended C-terminus of ACP only in another weevil, *Dendroctonus ponderosae* (Fig. 3).

One long (NPF_a) and one short (NPF_b) splice variant of the *npf* gene were identified, consistent with NPF precursors found in other insects (Xu et al., 2016; Veenstra, 2014; Derst et al., 2016). In *T. castaneum*, only the short splice variant of the *npf* gene has been identified as well as a putative receptor (Richards et al., 2008). A later study by Huang et al. (2011) identified the ligand but the GenBank record (AAJJ01002624.1) has since been removed. NPF was reported from another beetle species, *Dendroctonus ponderosae* (Nässel and Wegener, 2011). Mass spectrometry

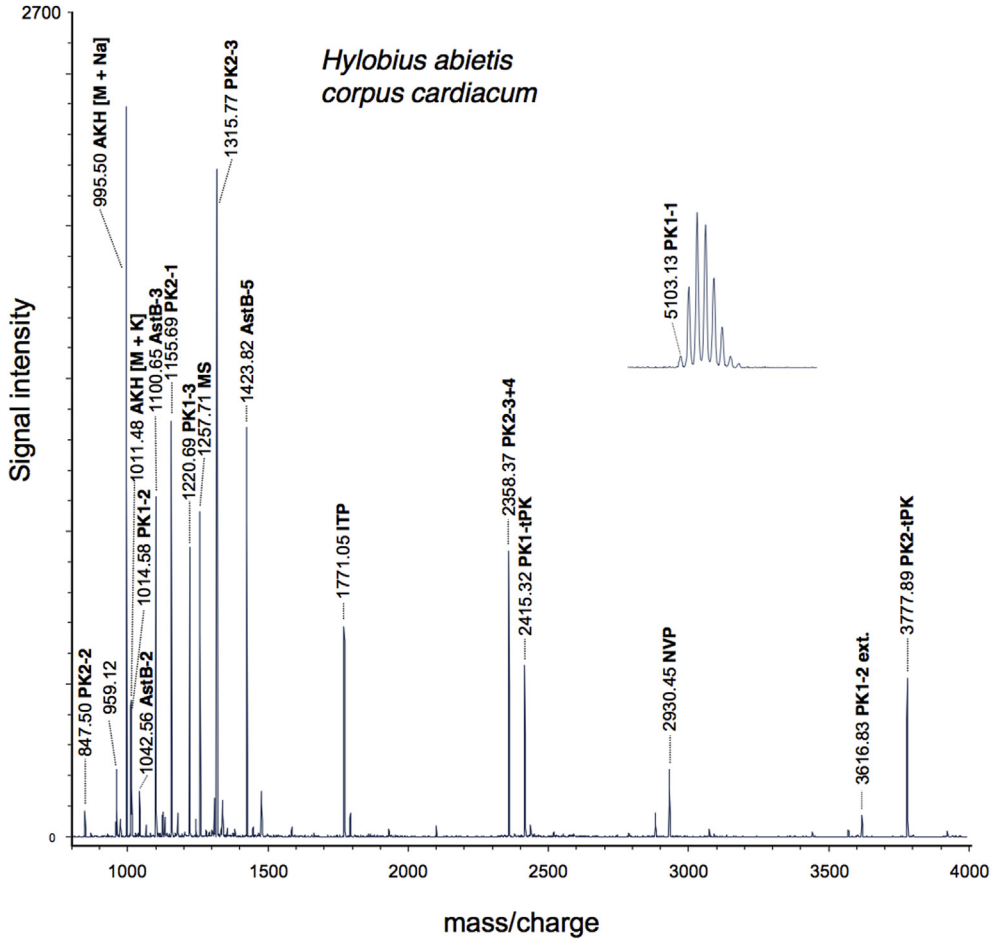

Fig. 5. Direct tissue profiling by MALDI-TOF mass spectrometry (m/z 800–4000 Da) of a single *corpus cardiacum* from *H. abietis*. Ion signals showing mass-identity with predicted pyrokinins of both PK precursors are marked. PK1-1 is the only PK, which was not detected as expected. The presence of an N-terminal extended form is indicated by mass-match (see inset). The identity of further peptide signals, AKH, AstB, MS, ITP, and a product of the NVP-precursor (DQPNHNGPEENIGRMS-YFNNKPTSAI) were confirmed by MSMS.

analyses were not able to identify the complete NPF of *H. abietis*, but truncated NPFs of both splicing variants. These peptides were apparently cleaved at Leu residues. However, the biological significance of these differences is not clear.

Pyrokinin and *capa* genes likely result from a gene duplication of a common ancestor gene at the origin of hexapods. This ancestor gene encoded already the neuropeptides periviscerokinins (PVKs), pyrokinins (PKs), and tryptoPKs (Derst et al., 2016), each activating specific receptors (Cazzamali et al., 2005; Park et al., 2002; Iversen et al., 2002; Terhzaz et al., 2012). Following gene duplication, capa and pk genes underwent substantial and often lineage-specific differentiations in insects. These differentiations included the allocation of PVKs, PKs, and tryptoPKs to a single or both of the genes as well as different transcription and processing events. A recent comparison of tenebrionid beetles (*Tribolium, Zophobas, Tenebrio*) and the weevil *Anthonomus grandis* (Curculionidae) demonstrated remarkable differences in the processing of CAPA peptides, even within the order Coleoptera (Neupert et al., 2017). Although the precursor sequences of *A. grandis* were incomplete, that study also suggested the presence of two *pk* genes for this species. The CAPA and PK precursors of *H. abietis* are complete and in agreement with the suggested condition in *A. grandis*. The CAPA precursor contains only a single PVK and a tryptoPK. Both peptides were detected, by mass spectrometry, in abdominal segmental nerves, which contain the axons of neurosecretory *capa* neurons in abdominal ganglia (see Fig. 4). Two PK precursors (PK1 and PK2) were identified in the transcriptome of *H. abietis,* and the detection of all predicted PKs (except PK1-1) confirmed the processing of both precursors. The PKs but not the CAPA peptides were identified in the retrocerebral complex, with peptides of the PK2 precursor showing always a higher abundance (Fig. 5).

The two PK precursors of *H. abietis* are very similar to each other regarding the order of putative bioactive peptides; i.e. a single tryptoPK followed by three PKs. The sequences of the complete precursors show, however, a large number of amino acid substitutions, which contradicts a possible recent gene duplication. Interestingly, a single PK (PK2-3) shows a Tyr in the typically conserved C-terminal FXPRL-NH_2 motif (YSPRL-NH_2). Similar sequences are known from basal hexapods (Derst et al., 2016) but also from cockroaches and termites (Predel et al., 1999; Veenstra, 2014). At least for cockroaches, it was shown that replacement of Phe in the active core sequence by Tyr does not decrease bioactivity in visceral muscle assays (Predel and Nachman, 2001).

Involved in control of reproduction and feeding (Lee at al., 2004), sNPF is characterized by an RLRFamide C-terminal with a size range between 8 and 19 amino acids (Broeck, 2001; Mertens et al., 2002). We identified a putative precursor encoding sNPF with the predicted sequence AGRSPQLRLRFamide in the *H. abietis* transcriptome. It was reported from several species, including *D. melanogaster* (Predel et al., 2004), that cleavage at the internal Arg results in a more prominent shorter sequence. In *H. abietis*, we did not find any ion signals in mass spectra that indicate a similar cleavage of sNPF. Similar to the NPFs, the biological significance of these differences has not been evident.

Table 4
Putative GPCRs identified in H. abietis. The result of the homology search is also shown.

Neuropeptide Receptor	Abbreviation	GPCR Class/Type	Transcript ORF ID	Homology search with known protein		
				Species	Accession No.	E-value
Adipokinetic hormone	AKH	Class A	TRINITY_DN90130_c0_g2_i1	*Dendroctonus ponderosae*	XP_019759039.1	0.00E+00
AKH/Corazonin-related peptide	ACP	Class A	TRINITY_DN90531_c0_g1_i1	*Tribolium castaneum*	NP_001280549.1	2.00E-159
Allatostatin C	AstC	Class A	TRINITY_DN83466_c0_g1_i1	*Dendroctonus ponderosae*	XP_019756512.1	0.00E+00
Allatotropin	AT	Class A	TRINITY_DN87881_c0_g1_i3	*Dendroctonus ponderosae*	XP_019771021.1	0.00E+00
Capability	CAPA	Class A	TRINITY_DN73886_c0_g1_i1	*Dendroctonus ponderosae*	XP_019756920.1	0.00E+00
CCHamide	CCHa	Class A	TRINITY_DN101990_c1_-g2_i1	*Dendroctonus ponderosae*	XP_019758999.1	0.00E+00
Crustacean Cardio-Active Peptide	CCAP	Class A	TRINITY_DN138776_c0_-g1_i1	*Anoplophora glabripennis*	XP_018569571.1	2.00E-71
Ecdysis-triggering hormone	ETH	Class A	TRINITY_DN92264_c1_g1_i1	*Dendroctonus ponderosae*	XP_019769415.1	0.00E+00
FMRFamide	FMRFa	Class A	TRINITY_DN83380_c0_g1_i1	*Dendroctonus ponderosae*	XP_019768282.1	0.00E+00
Myosuppressin	MS	Class A	TRINITY_DN89923_c0_g1_i1	*Dendroctonus ponderosae*	XP_019756536.1	0.00E+00
Neuropeptide F	NPF	Class A	TRINITY_DN143038_c0_-g1_i1	*Anoplophora glabripennis*	XP_018564523.1	1.00E-71
Pyrokinin	PK	Class A	TRINITY_DN71715_c0_g1_i1	*Dendroctonus ponderosae*	XP_019763370.1	0.00E+00
RYamide	Rya	Class A	TRINITY_DN80530_c0_g1_i2	*Dendroctonus ponderosae*	XP_019769489.1	0.00E+00
Sex peptide	SP	Class A	TRINITY_DN85632_c0_g2_i1	*Dendroctonus ponderosae*	XP_019758277.1	0.00E+00
Short neuropeptide F	sNPF	Class A	TRINITY_DN92830_c0_g3_i2	*Dendroctonus ponderosae*	XP_019761868.1	0.00E+00
SIFamide	SIFa	Class A	TRINITY_DN138746_c0_-g1_i1	*Dendroctonus ponderosae*	XP_019765870.1	2.00E-90
Sulfakinin	SK	Class A	TRINITY_DN84486_c0_g1_i1	*Dendroctonus ponderosae*	XP_019756917.1	0.00E+00
Tachykinin-related peptide	TRP	Class A	TRINITY_DN80463_c0_g1_i1	*Dendroctonus ponderosae*	XP_019768744.1	0.00E+00
Trissin	Trissin	Class A	TRINITY_DN84408_c0_g1_i1	*Dendroctonus ponderosae*	XP_019773394.1	0.00E+00
Inotocin (Vasopressin-Like)	VPL	Class A	TRINITY_DN78584_c0_g2_i1	*Tribolium castaneum*	ABX00684.1	0.00E+00
Calcitonin	Cal	Class B	TRINITY_DN85987_c0_g1_i1	*Dendroctonus ponderosae*	XP_019756394.1	0.00E+00
Diuretic Hormone 31	DH31	Class B	TRINITY_DN89716_c0_g1_i2	*Dendroctonus ponderosae*	XP_019763866.1	0.00E+00
Diuretic Hormone 44	DH44	Class B	TRINITY_DN98561_c0_g2_i9	*Tribolium castaneum*	XP_008198347.1	0.00E+00
Pigment-dispersing factor	PDF	Class B	TRINITY_DN88487_c0_g4_i1	*Dendroctonus ponderosae*	XP_019773389.1	0.00E+00
Leucine-rich Repeat-containing GPCR - CG(R)	LGR	Type A	TRINITY_DN94186_c0_g2_i2	*Dendroctonus ponderosae*	XP_019771060.1	0.00E+00
Leucine-rich Repeat-containing GPCR - FSH	LGR	Type A	TRINITY_DN96282_c1_g2_i1	*Dendroctonus ponderosae*	XP_019769811.1	0.00E+00
Leucine-rich Repeat-containing GPCR - LGR4-like	LGR	Type B	TRINITY_DN87527_c0_g1_i1	*Dendroctonus ponderosae*	XP_019753406.1	0.00E+00
Leucine-rich Repeat-containing GPCR - LGR5-like	LGR	Type B	TRINITY_DN93127_c0_g6_i1	*Dendroctonus ponderosae*	XP_019760595.1	0.00E+00
Leucine-rich Repeat-containing GPCR - Burs-like	LGR	Type B	TRINITY_DN94105_c0_g1_i1	*Tribolium castaneum*	EFA02891.2	0.00E+00
Leucine-rich Repeat-containing GPCR - LGR5-like	LGR	Type B	TRINITY_DN98008_c0_g1_i2	*Dendroctonus ponderosae*	XP_019754543.1	0.00E+00
Leucine-rich Repeat-containing GPCR - Insulin-like	LGR	Type C	TRINITY_DN97208_c0_g2_i3	*Dendroctonus ponderosae*	XP_019755840.1	0.00E+00
Leucine-rich Repeat-containing GPCR - Insulin-like	LGR	Type C	TRINITY_DN101696_c0_-g1_i1	*Dendroctonus ponderosae*	XP_019765260.1	0.00E+00
Leucine-rich Repeat-containing GPCR - Relaxin 2-like	LGR	Type C	TRINITY_DN84670_c0_g1_i2	*Dendroctonus ponderosae*	XP_019760354.1	0.00E+00

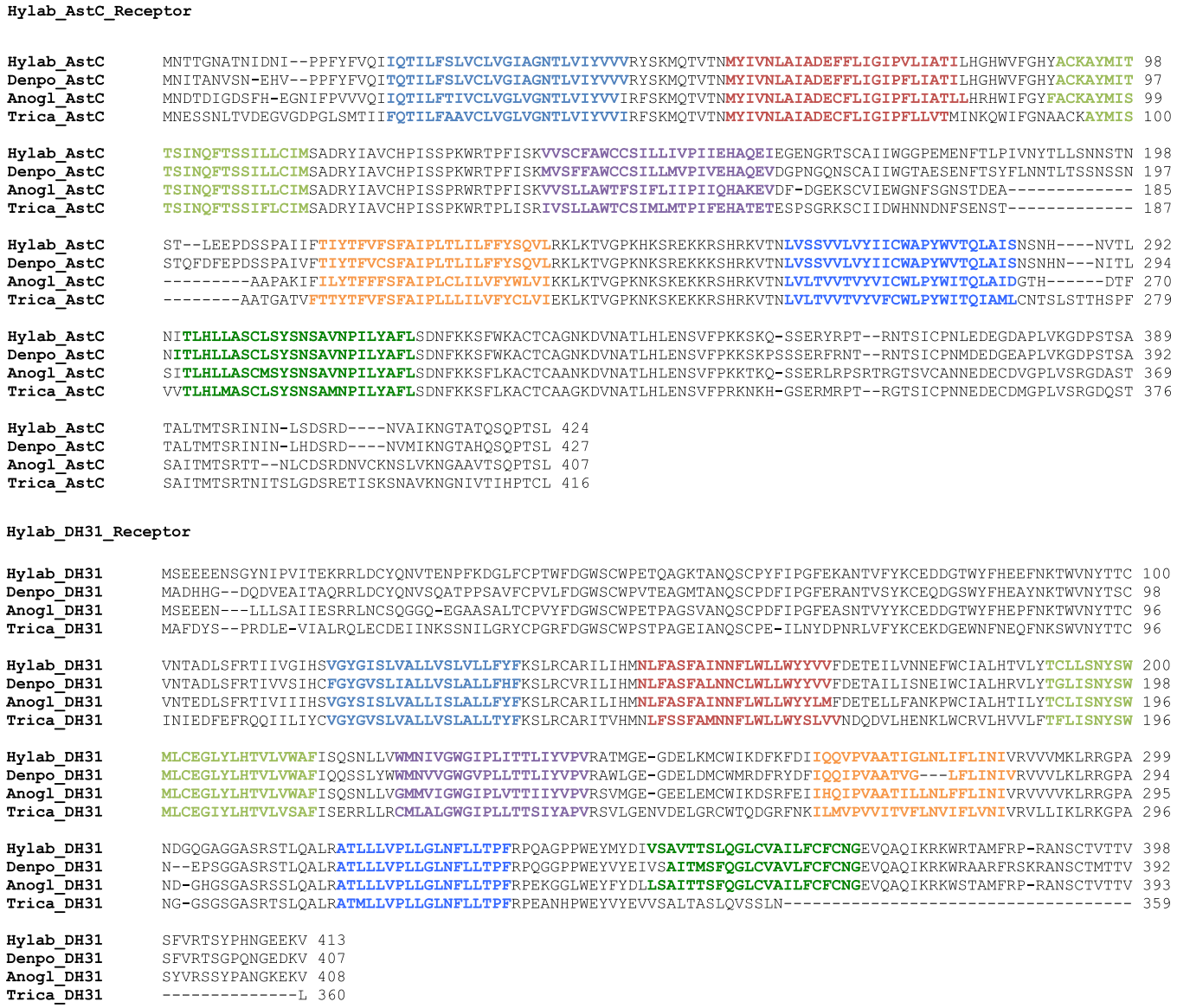

Fig. 6. Multiple sequence alignment of the identified putative neuropeptide GPCR sequences for A) AstC, from Class A or Rhodopsin-like receptor Class And B) DH31, from Class B or Secretin-like receptor family. The alignment displays the location of each of the 7 transmembrane helices found in *H. abietis as well as* in the aligned sequences of other insect species. See Supplementary Table 3 for all the identified putative neuropeptide GPCRs.

3.5. Identification of G protein-coupled receptors (GPCRs) for neuropeptides

Neuropeptide receptors belong to either GPCR Class A (Rhodopsin-like receptor family) or Class B (Secretin-like receptor family). Using the GPCR prediction pipeline described earlier, a total of 25 putative neuropeptide GPCRs were identified in *H. abietis*. Out of these 21 putative receptor sequences belonged to Class A (represented by 7tm_1 or PF0001 in the Pfam database), while the remaining four belonged to Class B (represented by 7tm_2 or PF00002 in the Pfam database) (Table 4).

Identified neuropeptide receptors belonging to Class A included those for Adipokinetic hormone (AKH), Allatostatin C (AstC), Capability/CAP2b (CAPA) and Pyrokinin (PK). On performing Conserved Domain searches on identified neuropeptide GPCR transcript ORFS, 18 (85.7%) of the 21 had the complete set of seven transmembrane helices while three (14.3%) had less than 4.

Among the Class B neuropeptide receptors, Calcitonin (Cal), Diuretic Hormones 31 (DH31) and 44 (DH44) as well as Pigment-dispersing factor (PDF) receptors were identified. Conserved Domain searches showed all four receptors to have a complete set of seven transmembrane helices.

Several LGRs were also identified. These included Type A LGRs including Follicle stimulating hormone receptor (FSH) and the chorionadotropin receptor (CG) as well as Type B LGRs including Bursicon-like as well as LGR4 and 5-like receptors. Finally, relaxin 2-like and insulin-like receptors, which are of Type C, were also identified. Details of these can be found in Table 4 and Supplementary Table 3.

Fig. 6 (a and b) illustrates the output from PSI/TM-Coffee software where the sequence of a receptor from Class A (Allatostatin C) of *H. abietis* was aligned with other allatostatin C receptor sequences from other species belonging to the same order i.e. *D. ponderosae, A. glabripennis* and *T. castaneum* as well as the sequence of a receptor from Class B (Diuretic Hormone 33).

Phylogenetic trees constructed from predicted amino acid sequences for the neuropeptide GPCRs of both Class A and Class B are shown in Fig. 7 (a and b respectively). The evolutionary similarities of *H. abietis* neuropeptide GPCRs and other model species including *T. castaneum, D. melanogaster* and *B. mori,* the closely related *D. ponderosae* as well as *A.*

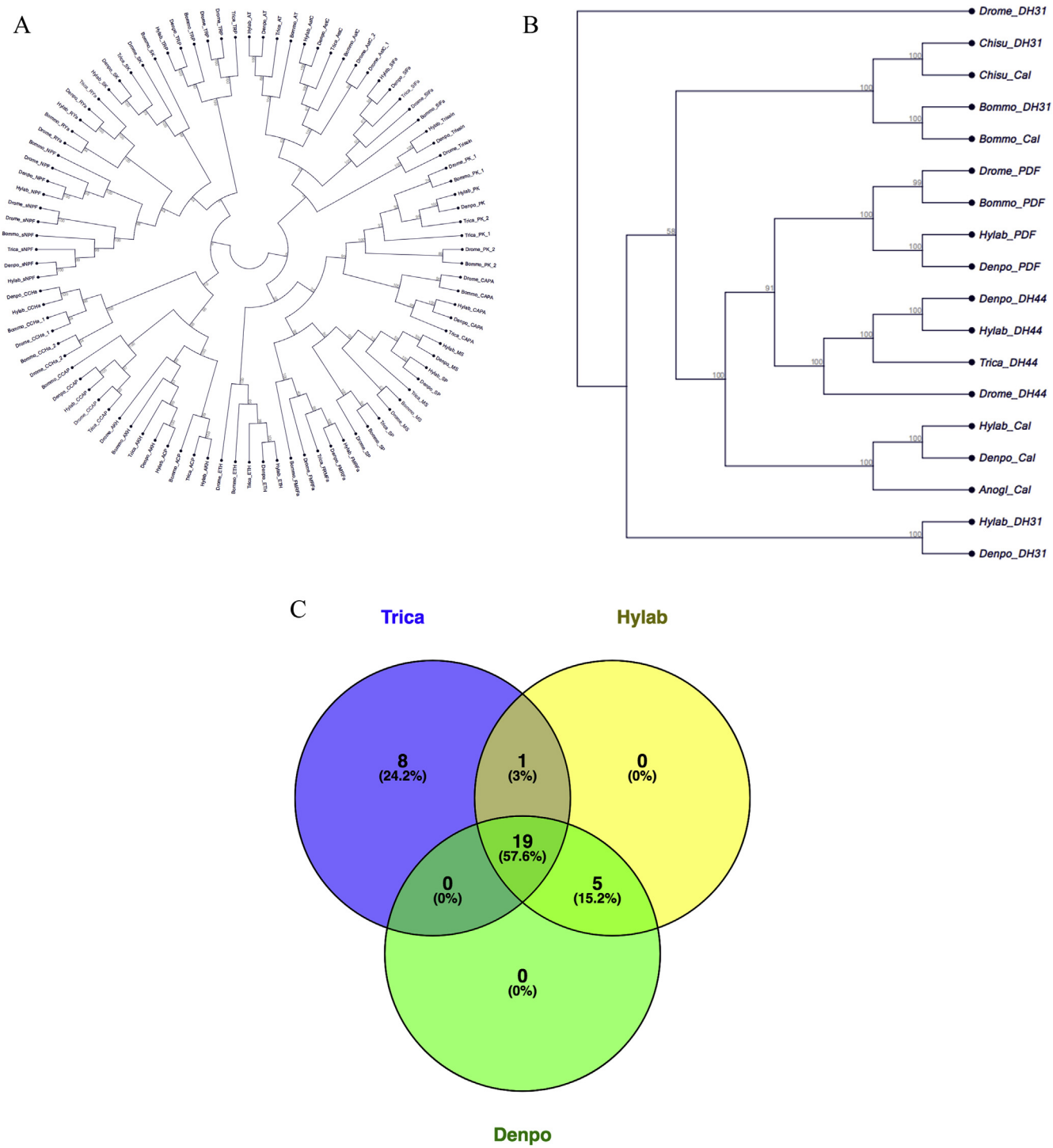

Fig. 7. Phylogenetic trees of the Class A (a) and Class B (b) neuropeptide GPCRs from *H. abietis* (Hylab), *T. castaneum* (Trica), *D. melanogaster* (Drome), *B. mori* (Bommo), *D. ponderosae* (Denpo) and *A. glabripennis* (Anogl) and (c) Illustration of the overlap of neuropeptide GPCRs putatively identified in H. abietis that are common with those that have been established already in *T. castaneum* i.e. the model Coleopteran and *D. ponderosae*, the closest studied beetle to *H. abietis*.

glabripennis can be clearly observed. Fig. 7(c) illustrates the number of neuropeptide GPCR sequences putatively identified in *Hylobius* in comparison to the ones pre-identified in *T. castaneum* and *D. ponderosae*.

4. Discussion

The neuropeptide profile of *H. abietis* from a combined approach of transcriptomics and peptidomics features the most complete repertoire of insect neuropeptides identified in a beetle to date and without the need for sequencing the whole genome. 41 different insect neuropeptide families were identified from 52 neuropeptide precursor sequences. Allatostatin A, Allatostatin CCC, Corazonin, Kinin, Sex Peptide and Pigment Dispersing Factor mature peptides were not identified in *H. abietis*, consistent with *T. castaneum*. Additionally, both long and short Neuropeptide F isoforms from *H. abietis* were identified although Neuropeptide F mature peptides are not found in *T. castaneum* (Li et al. (2008). Additionally, Veenstra (2014) did not identify the elevenin precursor in *T. castaneum* but elevenin was identified in this study.

Furthermore, the allatotropin precursor in H. *abietis*, *A. glabripennis* and *D. ponderosae* contains an additional putative mature peptide, which is absent in *T. castaneum*.

Among the alignment methods used in the prediction of GPCRs, neuropeptide receptor sequence similarity performed using BLAST software as well as Pfam database searches displayed more results than the non-alignment methods i.e. using TMHMM software. This may be due to the fact that there is an increased availability of insect peptide sequences and protein family information. The GPCRs identified in *H. abietis* shared sequence homology or high similarity with those of the mountain pine beetle, *D. ponderosae* (34%) and the red flour beetle, *T. castaneum* (14%). The relatively low number of predicted neuropeptide receptors compared to a high number of neuropeptides could be due to lack of extensive information on insect neuropeptide GPCRs, so that a high number of 'orphan' receptors hinders positive and accurate identification of more neuropeptide receptors.

The transcriptome generated by this study has been made available publicly (SRA accession: SRP133355), with the intent that it act as a resource for *H. abietis* in other areas including gene expression profiling, gene discovery or simple comparative genomics as well as neuropeptidomic studies for other non-model pest species. Although research in these areas is at early stage, the neuropeptide and receptor profiles of various pest insects are being increasingly studied, assisted by a reduction in time and costs, due, in part, to advances in next generation sequencing methods, transcriptomics and genomic analyses.

Funding

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

Conflicts of interest

None declared.

Acknowledgements

We would like to thank Saurav Ghimire for help with the images and Dr. Lucy Alford for useful discussions.

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.ibmb.2018.08.003.

References

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.
Audsley, N., Down, R.E., 2015. G protein coupled receptors as targets for next generation pesticides. INSECT BIOCHEM MOLEC 67, 27–37.
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.
Cazzamali, G., Torp, M., Hauser, F., Williamson, M., Grimmelikhuijzen, C.J., 2005. The Drosophila gene CG9918 codes for a pyrokinin-1 receptor. Biochem Bioph Res Co 335 (1), 14–19.
Chang, J.M., Di Tommaso, P., Taly, J.F., Notredame, C., 2012. Accurate multiple sequence alignment of transmembrane proteins with PSI-Coffee. BMC Bioinf. 13 (4), S1.
Christie, A.E., 2008. Neuropeptide discovery in Ixodoidea: an in silico investigation using publicly accessible expressed sequence tags. Gen. Comp. Endocrinol. 157 (2), 174–185.
Christie, A.E., 2014. Prediction of the first neuropeptides from a member of the Remipedia (Arthropoda, Crustacea). Gen. Comp. Endocrinol. 201, 74–86.
Christie, A.E., 2015. In silico prediction of a neuropeptidome for the eusocial insect *Mastotermes darwiniensis*. Gen. Comp. Endocrinol. 224, 69–83.
Cock, P.J., Grüning, B.A., Paszkiewicz, K., Pritchard, L., 2013. Galaxy tools and workflows for sequence analysis with applications in molecular plant pathology. PEERJ 1 e167.
Conesa, A., Götz, S., García-Gómez, J.M., Terol, J., Talón, M., Robles, M., 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21 (18), 3674–3676.
Crooks, G.E., Hon, G., Chandonia, J.M., Brenner, S.E., 2004. WebLogo: a sequence logo generator. Genome Res. 14 (6), 1188–1190.
Cunningham, C.B., Ji, L., Wiberg, R.A.W., Shelton, J., McKinney, E.C., Parker, D.J., Meagher, R.B., Benowitz, K.M., Roy-Zokan, E.M., Ritchie, M.G., Brown, S.J., 2015. The genome and methylome of a beetle with complex social behavior, *Nicrophorus vespilloides* (Coleoptera: Silphidae). Genome Biol Evol 7 (12), 3383–3396.
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.
Finn, R.D., Mistry, J., Tate, J., Coggill, P., Heger, A., Pollington, J.E., Gavin, O.L., Gunasekaran, P., Ceric, G., Forslund, K., Holm, L., 2009. The Pfam protein families' database. Nucleic Acids Res. 38 (Suppl. l_1), D211–D222.
Gard, A.L., Lenz, P.H., Shaw, J.R., Christie, A.E., 2009. Identification of putative peptide paracrines/hormones in the water flea *Daphnia pulex* (Crustacea; Branchiopoda; Cladocera) using transcriptomics and immunohistochemistry. Gen. Comp. Endocrinol. 160 (3), 271–287.
Garland, S.L., 2013. Are GPCRs still a source of new targets? J. Biomol. Screen 18 (9), 947–966.
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).
Hansen, K.K., Stafflinger, E., Schneider, M., Hauser, F., Cazzamali, G., Williamson, M., Kollmann, M., Schachtner, J., Grimmelikhuijzen, C.J., 2010. Discovery of a novel insect neuropeptide signaling system closely related to the insect adipokinetic hormone and corazonin hormonal systems. J. Biol. Chem. 285 (14), 10736–10747.
Hauser, F., Neupert, S., Williamson, M., Predel, R., Tanaka, Y., Grimmelikhuijzen, C.J., 2010. Genomics and peptidomics of neuropeptides and protein hormones present in the parasitic wasp *Nasonia vitripennis*. J. Proteome Res. 9 (10), 5296–5310.
Huang, Y., Crim, J.W., Nuss, A.B., Brown, M.R., 2011. Neuropeptide F and the corn earworm, *Helicoverpa zea*: a midgut peptide revisited. Peptides 32 (3), 483–492.
Iversen, A., Cazzamali, G., Williamson, M., Hauser, F., Grimmelikhuijzen, C.J., 2002. Molecular cloning and functional expression of a Drosophila receptor for the neuropeptides capa-1 and-2. Biochem Bioph Res Co 299 (4), 628–633.
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.
Kumar, A., Congiu, L., Lindström, L., Piiroinen, S., Vidotto, M., Grapputo, A., 2014. Sequencing, de novo assembly and annotation of the Colorado potato beetle, *Leptinotarsa decemlineata*, transcriptome. PLoS One 9 (1), e86012.
LaJeunesse, D.R., Johnson, B., Presnell, J.S., Catignas, K.K., Zapotoczny, G., 2010. Peristalsis in the junction region of the Drosophila larval midgut is modulated by DH31 expressing enteroendocrine cells. BMC Phys 10 (1), 14.
Leather, S.R., Day, K.R., Salisbury, A.N., 1999. The biology and ecology of the large pine weevil, *Hylobius abietis* (Coleoptera: Curculionidae): a problem of dispersal? B ENTOMOL RES 89 (1), 3–16.
Lee, K.S., You, K.H., Choo, J.K., Han, Y.M., Yu, K., 2004. Drosophila short neuropeptide F regulates food intake and body size. J. Biol. Chem. 279 (49), 50781–50789.
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.
Li, W., Cowley, A., Uludag, M., Gur, T., McWilliam, H., Squizzato, S., Park, Y.M., Buso, N., Lopez, R., 2015. The EMBL-EBI bioinformatics web and programmatic tools framework. Nucleic Acids Res. 43 (W1), W580–W584.
Lin, T., Cai, Z., Wu, H., 2015. Transcriptome analysis of the Japanese pine sawyer beetle, *Monochamus alternatus* (Coleoptera: Cerambycidae) by high-throughput Illumina sequencing. J. Asia Pac. Entomol. 18 (3), 439–445.
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 USA 102 (8), 2820–2825.
Ma, M., Gard, A.L., Xiang, F., Wang, J., Davoodian, N., Lenz, P.H., Malecha, S.R., Christie, A.E., Li, L., 2010. Combining in silico transcriptome mining and biological mass spectrometry for neuropeptide discovery in the Pacific white shrimp *Litopenaeus vannamei*. Peptides 31 (1), 27–43.
Martin, M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.J 17 (1), 10.
Marchler-Bauer, A., Bo, Y., Han, L., He, J., Lanczycki, C.J., Lu, S., Chitsaz, F., Derbyshire, M.K., Geer, R.C., Gonzales, N.R., Gwadz, M., 2016. CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. Nucleic Acids Res. 45 (D1), D200–D203.
Marchler-Bauer, A., Bryant, S.H., 2004. CD-Search: protein domain annotations on the fly. Nucleic Acids Res. 32 (Suppl. l_2), W327–W331.

Matthews, H.J., Audsley, N., Weaver, R.J., 2007. Interactions between allatostatins and allatotropin on spontaneous contractions of the foregut of larval *Lacanobia oleracea*. J. Insect Physiol. 53 (1), 75–83.
McKenna, D.D., Scully, E.D., Pauchet, Y., Hoover, K., Kirsch, R., Geib, S.M., Mitchell, R.F., Waterhouse, R.M., Ahn, S.J., Arsala, D., Benoit, J.B., 2016. Genome of the Asian longhorned beetle (*Anoplophora glabripennis*), a globally significant invasive species, reveals key functional and evolutionary innovations at the beetle–plant interface. Genome Biol. 17 (1), 227.
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.
Meyer, J.M., Markov, G.V., Baskaran, P., Herrmann, M., Sommer, R.J., Rödelsperger, C., 2016. Draft genome of the scarab beetle *Oryctes borbonicus* on La Réunion Island. Genome Biol Evol 8 (7), 2093–2105.
Nässel, D.R., Wegener, C., 2011. A comparative review of short and long neuropeptide F signaling in invertebrates: any similarities to vertebrate neuropeptide Y signaling? Peptides 32 (6), 1335–1355.
Nässel, D.R., Winther, Å.M., 2010. Drosophila neuropeptides in regulation of physiology and behavior. Prog. Neurobiol. 92 (1), 42–104.
Nash, M.A., Thomson, L.J., Hoffmann, A.A., 2008. Effect of remnant vegetation, pesticides, and farm management on abundance of the beneficial predator Notonomus gravis (Chaudoir) (Coleoptera: Carabidae). Biol. Control 46 (2), 83–93.
Neupert, S., Marciniak, P., Köhler, R., Nachman, R.J., Suh, C.P.C., Predel, R., 2017. Different processing of CAPA and pyrokinin precursors in the giant mealworm beetle *Zophobas atratus* (Tenebrionidae) and the boll weevil *Anthonomus grandis grandis* (Curculionidae). Gen. Comp. Endocrinol. 258, 53–59.
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.
Orchard, I., 2009. Peptides and serotonin control feeding-related events in *Rhodnius prolixus*. Front. Biosci. 1, 250–262.
Park, Y., Kim, Y.J., Adams, M.E., 2002. Identification of G protein-coupled receptors for Drosophila PRXamide peptides, CCAP, corazonin, and AKH supports a theory of ligand-receptor coevolution. P Natl Acad Sci USA 99 (17), 11423–11428.
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., Kellner, R., Nachman, R.J., Holman, G.M., Rapus, J., Gäde, G., 1999. Differential distribution of pyrokinin-isoforms in cerebral and abdominal neurohemal organs of the American cockroach. Insect Biochem. Molec. 29 (2), 139–144.
Predel, R., Nachman, R.J., 2001. Efficacy of native FXPRLamides (pyrokinins) and synthetic analogs on visceral muscles of the American cockroach. J. Insect Physiol. 47 (3), 287–293.
Predel, R., Neupert, S., Derst, C., Reinhardt, K., Wegener, C., 2017. Neuropeptidomics of the bed bug Cimex lectularius. J. Proteome. Res. 17 (1), 440–454.
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.
Predel, R., Wegener, C., Russell, W.K., Tichy, S.E., Russell, D.H., Nachman, R.J., 2004. Peptidomics of CNS-associated neurohemal systems of adult *Drosophila melanogaster*: a mass spectrometric survey of peptides from individual flies. J. Comp. Neurol. 474 (3), 379–392.
Price, M.D., Merte, J., Nichols, R., Koladich, P.M., Tobe, S.S., Bendena, W.G., 2002. Drosophila melanogaster flatline encodes a myotropin orthologue to *Manduca sexta* allatostatin. Peptides 23 (4), 787–794.
Rappsilber, J., Mann, M., Ishihama, Y., 2007. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. Nat. Protoc. 2 (8), 1896–1906.
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 2008.
Riley, M.A., Goyer, R.A., 1986. Impact of beneficial insects on *Ips* spp. (Coleoptera: Scolytidue) Bark Beetles in Felled Loblolly and Slash Pines in Louisiana. Environ. Entomol. 15 (6), 1220–1224.
Roller, L., Yamanaka, N., Watanabe, K., Daubnerová, I., Žitňan, D., Kataoka, H., Tanaka, Y., 2008. The unique evolution of neuropeptide genes in the silkworm *Bombyx mori*. Insect Biochem Molec 38 (12), 1147–1157.
Sayadi, A., Immonen, E., Bayram, H., Arnqvist, G., 2016. The de novo transcriptome and its functional annotation in the seed beetle *Callosobruchus maculatus*. PLoS One 11 (7), e0158565.
Schachtner, J., Wegener, C., Neupert, S., Predel, R., 2010. Direct peptide profiling of brain tissue by MALDI-TOF mass spectrometry. In: PEPTIDOMICS. Humana Press, pp. 129–135.
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.
Siegert, K.J., 1999. Locust corpora cardiaca contain an inactive adipokinetic hormone. Febs. Lett. 447 (2-3), 237–240.
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.
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.
Te Brugge, V.A., Schooley, D.A., Orchard, I., 2008. Amino acid sequence and biological activity of a calcitonin-like diuretic hormone (DH31) from *Rhodnius prolixus*. J. Exp. Biol. 211 (3), 382–390.
Terhzaz, S., Cabrero, P., Robben, J.H., Radford, J.C., Hudson, B.D., Milligan, G., Dow, J.A., Davies, S.A., 2012. Mechanism and function of Drosophila capa GPCR: a desiccation stress-responsive receptor with functional homology to human neuromedinU receptor. PLoS One 7 (1), e29897.
Terhzaz, S., Teets, N.M., Cabrero, P., Henderson, L., Ritchie, M.G., Nachman, R.J., Dow, J.A., Denlinger, D.L., Davies, S.A., 2015. Insect capa neuropeptides impact desiccation and cold tolerance. P NATL ACAD SCI USA 112 (9), 2882–2887.
Teshler, M.P., Dernovici, S.A., Ditommaso, A., Coderre, D., Watson, A.K., 2004. A novel device for the collection, storage, transport, and delivery of beneficial insects, and its application to *Ophraella communa* (Coleoptera: Chrysomelidae). Biocontrol. Sci. Tech. 14 (4), 347–357.
Uchiyama, H., Maehara, S., Ohta, H., Seki, T., Tanaka, Y., 2018. Elevenin regulates the body color through a G protein-coupled receptor NlA42 in the brown planthopper *Nilaparvata lugens*. Gen. Comp. Endocrinol. 258, 33–38.
Van Hiel, M.B., Vandersmissen, H.P., Van Loy, T., Broeck, J.V., 2012. An evolutionary comparison of leucine-rich repeat containing G protein-coupled receptors reveals a novel LGR subtype. Peptides 34 (1), 193–200.
Veenstra, J.A., 2000. Mono-and dibasic proteolytic cleavage sites in insect neuroendocrine peptide precursors. Arch Insect Biochem 43 (2), 49–63.
Veenstra, J.A., 2014. The contribution of the genomes of a termite and a locust to our understanding of insect neuropeptides and neurohormones. Front. Physiol. 5.
Vega, F.E., Brown, S.M., Chen, H., Shen, E., Nair, M.B., Ceja-Navarro, J.A., Brodie, E.L., Infante, F., Dowd, P.F., Pain, A., 2015. Draft genome of the most devastating insect pest of coffee worldwide: the coffee berry borer. *Hypothenemus hampei*. SCI REP 5.
Wistrand, M., Käll, L., Sonnhammer, E.L., 2006. A general model of G protein-coupled receptor sequences and its application to detect remote homologs. Protein Sci. 15 (3), 509–521.
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. 6, 28976.
Yamanaka, N., Hua, Y.J., Roller, L., Spalovská-Valachová, I., Mizoguchi, A., Kataoka, H., Tanaka, Y., 2010. Bombyx prothoracicostatic peptides activate the sex peptide receptor to regulate ecdysteroid biosynthesis. P Natl Acad Sci USA 107 (5), 2060–2065.
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 molec 86, 9–19.
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, J., Xin, L., Shan, B., Chen, W., Xie, M., Yuen, D., Zhang, W., Zhang, Z., Lajoie, G.A., Ma, B., 2012. PEAKS DB: de novo sequencing assisted database search for sensitive and accurate peptide identification. Mol. Cell. Proteomics 11 (4), M111–M010587.
Zitnan, D., Adams, M.E., 2012. Neuroendocrine regulation of ecdysis. In: INSECT ENDOCRINOLOGY, pp. 253–309.