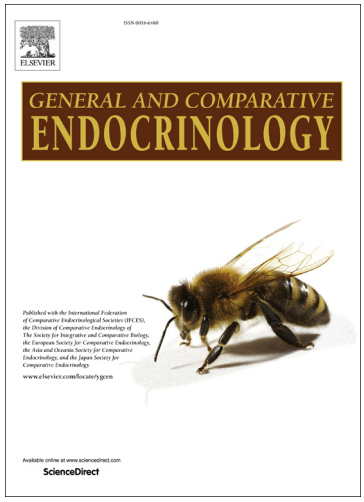
Accepted Manuscript



Characterization of the neuropeptidome of a Southern Ocean decapod, the Antarctic shrimp *Chorismus antarcticus*: focusing on a new decapod ITP-like peptide belonging to the CHH peptide family

Jean-Yves Toullec, Erwan Corre, Perrine Mandon, Marcelo Gonzalez-Aravena, Céline Ollivaux, Chi-Ying Lee

PII: S0016-6480(17)30407-0
DOI: http://dx.doi.org/10.1016/j.ygcen.2017.07.015
Reference: YGCEN 12699

To appear in: *General and Comparative Endocrinology*

Received Date: 24 May 2017
Revised Date: 6 July 2017
Accepted Date: 16 July 2017

Please cite this article as: Toullec, J-Y., Corre, E., Mandon, P., Gonzalez-Aravena, M., Ollivaux, C., Lee, C-Y., Characterization of the neuropeptidome of a Southern Ocean decapod, the Antarctic shrimp *Chorismus antarcticus*: focusing on a new decapod ITP-like peptide belonging to the CHH peptide family, *General and Comparative Endocrinology* (2017), doi: http://dx.doi.org/10.1016/j.ygcen.2017.07.015

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

Characterization of the neuropeptidome of a Southern Ocean decapod, the Antarctic shrimp *Chorismus antarcticus*: focusing on a new decapod ITP-like peptide belonging to the CHH peptide family.

Jean-Yves Toullec[1*], Erwan Corre[2], Perrine Mandon[1,3], Marcelo Gonzalez-Aravena[4], Céline Ollivaux[5] and Chi-Ying Lee[6]

[1] Sorbonne Universités, UPMC Université Paris 06, UMR 7144 CNRS, Equipe ABICE, Station Biologique de Roscoff, 29682 Roscoff, France

[2] Sorbonne Universités, UPMC Université Paris 06, FR 2424 CNRS, ABiMS, Station Biologique de Roscoff, Roscoff, France.

[3] Muséum National d'Histoire Naturelle, Sorbonne Universités, Institut de Systématique, Evolution, Biodiversité, (ISYEB – UMR 7205 – CNRS, MNHN, UPMC-Paris 06, EPHE), 57 rue Cuvier, CP26, 75005 Paris, France

[4] Laboratorio de Biorecursos Antárticos, Departamento Científico, Instituto Antártico Chileno, Punta Arenas, Chile

[5] Sorbonne Universités, UPMC Université Paris 06, UMR 8227 CNRS, Equipe PCE, Station Biologique de Roscoff, 29682 Roscoff, France

[6] Department of Biology, National Changhua University of Education, Changhua 50058, Taiwan

* Corresponding author

E-mail address: jean-yves.toullec@sb-roscoff.fr

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Abstract

As part of the study of the resilience of Antarctic crustaceans to global warming, the shrimp *Chorismus antarcticus* was subjected to an analysis of global approach using the Next Generation Sequencing Illumina Hi-Seq platform. With this data a detailed study into the principal neuropeptides and neurohormones of this species have been undertaken. Total RNAs from whole animals were enriched with eyestalk extracts to ensure maximum sequencing depth of the different neurohormones and neuropeptides mainly expressed into the X organ-sinus gland complex, which is a major endocrine organ of their synthesis. Apart from the information that can provide the availability of the transcriptome of a polar crustacean, the study of neuropeptides of a caridean shrimp will partially fill the limited data available for this taxon. Illumina sequencing was used to produce a transcriptome of the polar shrimp. Analysis of the Trinity assembled contigs produced 55 pre-pro-peptides, coding for 111 neuropeptides belonging to the following families: adipokinetic-corazonin-like peptide, Allatostatins (A, B et C), Bursicon (α), CCHamide, Crustacean Hyperglycemic Hormones (CHH), Crustacean Cardioactive Peptide (CCAP), Corazonin, Crustacean Female Sex Hormone (CSFH), Diuretic Hormones 31 and 45 (DH), Eclosion Hormone (EH), FLRFamide, GSEFLamide, Intocin, Ion Transport Peptide-like (ITP-like), Leucokinin, Molt-inhibiting Hormone, Myosuppresin, Neuroparsin, Neuropeptide F (NPF), Orcokinin, Orcomyotropin, Pigment Dispersing Hormone (PDH), Pyrokinin, Red Pigment Concentrating Hormone (RPCH), SIFamide, small Neuropeptide F (sNPF), sulfakinin and finally Tachykinin Related peptides. Among the new peptides highlighted in this study, the focus was placed on the peptides of the CHH family and more particularly on a new ITP-like in order to confirm its belonging to a new group of peptides of the family. A phylogeny made from more than 200 sequences of peptides, included new sequences from new species besides *Chorismus antarcticus*, confirms the peculiarity of this new set of peptides gathered under the name ITP-like.

Keywords: Crustacea, Neuropeptides, CHH, ITP-like, Transcriptomics, Antarctica

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

1. Introduction

The scarcity of representatives of crustacean decapods in the Antarctic Ocean is one of the most surprising and enigmatic observations in the study of biodiversity (Gorny, 1999; Thatje and Arntz, 2004). This diversity is summed up by a dozen species of benthic caridean shrimps among which is the Antarctic shrimp *Chorismus antarcticus.* This small hippolytid shrimp (Pfeffer, 1887) only occurs on the continental shelf in depths shallower than 700m (Arntz and Gorny, 1991; Basher et al., 2014). The presence of this shrimp on the bottom of the continental shelf suggests that, like other benthic invertebrates, it would be strictly stenothermal and therefore would possess a limited capacity to respond to a potential warming of waters (Peck, 2004; Peck et al., 2010; Portner et al., 2007).

As part of an ongoing study of the resilience of Antarctic crustaceans such as krill to global warming (Cascella et al., 2015), *C. antarcticus* seemed another good model because of its different life mode and its close phylogenetic position in relative to euphausiids. So, a similar global approach was taken in this study using the Next Generation Sequencing Illumina Hi-Seq platform. With this data a detailed study into the principal neuropeptides and neurohormones of this species have been undertaken. As with the ice krill *Euphausia crystallorophias* (Toullec et al., 2013), total RNAs from whole animals were enriched with eyestalk extracts to maximize sequencing depth of the different neurohormones and neuropeptides mainly expressed into the X organ-sinus gland complex, which is the major endocrine organs of their synthesis. Apart from the information that can provide the availability of the transcriptome of a polar crustacean, the study of neuropeptides of a caridean shrimp will partially fill the limited data available for this taxon. Indeed, paradoxically, few neuropeptides sequences are available outside of economically important species such as *Macrobrachium sp.* and there is not, to our knowledge, another transcriptomic analysis focusing on these neuropeptides except again on *M. rosenbergii* (Suwansa-Ard et al., 2015). Moreover, the characterization of an ITP-like sequence within this decapod species has represented the opportunity not only to deepen the reality of the existence of this new family of peptides in this taxon but also to make a point on the phylogeny of the CHH family in Euarthropods by incorporating a maximum of new sequences resulting from studies of recent peptidomes.

ACCEPTED MANUSCRIPT

2. Materials and methods

This project (IPEV- 1039) was approved by IPEV (Institut Paul Emile Victor, the French Polar Institute) review committee and was declared to and approved by the Terres Australes et Antarctiques Françaises in 2009 according the Annex I of the Madrid Protocol and the French Decree No 2005-403. No endangered or protected species were used.

2.1. Biological material, RNA extraction and Illumina sequencing

The shrimps *Chorismus antarcticus* were trawl-fished during the 2011 summer from the continental plateau in the immediate vicinity of the French station Dumont d'Urville (DDU) in Terre Adélie, at the foot of the Astrolabe glacier (66°40'S-140°01'E). The sampling depth was around 80 meters. The animals were frozen in liquid nitrogen immediately after returning to the station and then stored at -80 ° C until the RNAs were extracted. Two whole animals were used for RNA extractions from the thorax and abdomen. Due to the size of the animals (5-6 cm), extractions were carried out separately on the thorax and abdomen. In addition, 20 eyestalks were partially dissected to remove the pigmented regions and then snap frozen in liquid nitrogen until extraction. RNAs were extracted from these tissues using the SV Total RNA Isolation System (Promega, Madison, WI, USA). The RNAs extracted from the thorax and abdomen were mixed with a ratio (w/w) of 3 to 2; and to the mixture the RNAs extracted from 20 eyestalks were added. The pooled and eyestalk-enriched RNAs sample was used for sequencing conducted by the McGill University and Génome Québec Innovation Centre (Montréal, Québec, Canada) following the manufacturer's instructions (Illumina, San Diego, CA).

2.2. RNA-Seq datasets

The cDNA library was sequenced to produce 100bp paired-end reads. Raw reads were filtered from low-quality sequences, low-complexity sequences and trimmed using FASTX toolkit (http://hannonlab.cshl.edu/fastx_toolkit/index.html). The reads were trimmed and filtered using a quality threshold of 25 (base calling) and a minimal size of 60bp. Only reads in which more than 75% of nucleotides had a minimal quality threshold of 20 were retained. Reads were then cleaned from adapter ends using cutadapt (version 1.01(Martin, 2011)). Finally, the cleaning process was checked using fastQC (version 0.10.01 http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc/).

ACCEPTED MANUSCRIPT

The assembly resulting from all the cleaned reads was performed using Trinity (release 2013-02-25; (Grabherr et al., 2011)), a genome-independent transcriptome assembler. Finally, reads were remapped to the full transcriptome using Bowtie (version 0.12.8;(Langmead et al., 2009)) and relative abundances were estimated using RSEM (version 1.2.0 ;(Li and Dewey, 2011)) to get the FPKM (Fragments per kilobase of exon per million fragments mapped) values for the identification of low coverage contigs (FPKM<1) and rare isoforms (<1%) that were excluded later from the analysis (both software programs were launched through the Trinity package Wrapper filter_fasta_by_rsem_values.pl). Peptide prediction was performed using Transdecoder (Haas et al., 2013). Similarity search (blastp of the Transdecoder predicted peptides) was performed against the uniprot-swissprot database (release 2013-09). Peptide signal prediction was performed using signalP v4.0 (Petersen et al., 2011). Transmembrane peptides detection was performed using TMHMM v2.0c (Krogh et al., 2001). Protein domain search was performed using hmmscan from the hmmer v.3.1b1 suite against the Pfam-A database ((Finn et al., 2014) release 27.0). Finally Transcriptome functional annotation was performed using the Trinotate pipeline (http://trinotate.github.io) described in Haas et al. (2013).

2.3. *Chorismus antarcticus* peptide selection

A local database of annotated peptides, with their corresponding sequences was developed. The peptides were chosen based on the most highly characterized neuropeptide and neurohormone sequences in the Arthropods (Christie, 2016a, b; Christie and Pascual, 2016; Christie et al., 2017), with particular reference to the *Daphnia pulex* genome (Christie et al., 2011; Dircksen et al., 2011). In the first instance, relevant Blast2Go annotations from the Trinity assembly were identified. Each identified contig was then Blast searched independently at the NCBI website to confirm the annotation. The contigs were then translated and the putative coding sequences delineated. These sequences were then subjected to a Blastp search and subsequently aligned with orthologous sequences from arthropods. A second approach consisted in a direct tBlastn searching on local database with orthologous peptide sequences already characterized in other decapod transcriptomes. All of the Blast search data and alignments were performed in CLC Main Workbench 7. The signal peptides were identified using SignalP.

ACCEPTED MANUSCRIPT

2.4. CHH family peptides

Most of sequences of CHH family members come from databases. However, the copmplied data has been increased for caridea taxon by the search for orthologous within the transcriptomes resulting from the work of Havird and Santos (Genomic Resources Development Consortium et al., 2014)(3 species)(*Antecaridina lauensis*; *Halocaridinides trigonopthalma*; *Metabetaeus lohena*) and Mandon et al. (8 species; Unpublished results)(*Atyopsis moluccensis*; *Athanas nitescens*; *Rimicaris exoculata*; *Oplophorus gracilirostris*; *Crangon crangon*; *Caridion steveni*; *Periclimenes brevicarpalis*; *Heterocarpus sp*.). The sequences found in these last eight species have been submitted to Genbank and therefore have an accession number.

The transcriptomes of *Neocaridina denticulata* and *Palaemon carinicauda* were reassembled from reads deposed in SRA as PRJNA240382 and PRJNA240382 respectively (Mandon et al., unpublished data).

The sequences from the crab *Metograpsus thukuhar* and the polar isopod *Glyptonotus antarcticus* were extracted from unpublished transcriptomes kindly provided by Pr C.Y. Lee and Dr M. Gonzalez-Aravena respectively.

2.5. Phylogenetic analyze of CHH family peptides

The alignments were performed manually with CLC Main Workbench 7 software (Quiagen) with the complete sequences of the peptides of the CHH family from various Arthropoda. After removal of N-terminal and C-terminal unconserved residues, the dataset contained 202 taxa and 71 characters. ITP-like sequences of Chelicerata were used as outgroup.

Phylogenetic reconstructions were carried out on amino acid sequences using Bayesian inference and maximum likelihood. Bayesian analyses were performed with MrBayes 3.2.5 with four chains of 10^6 generations; trees were sampled every 100 generations and burn-in value set to 20% of the sampled trees. We checked that standard deviation of the split frequencies fell below 0.01 to insure convergence in tree search. Protein sequences were analyzed with a mixed amino acid model (Ronquist and Huelsenbeck, 2003). Maximum likelihood reconstruction was carried out with the LG+I+G substitution model (Whelan and Goldman, 2001) determined as the best-fit model of protein evolution by ProtTest 1.3 (Abascal et al., 2005) http://darwin.uvigo.es/software/prottest_server.html, following Akaike Information Criterion. Rate heterogeneity was set at four categories. The gamma distribution

ACCEPTED MANUSCRIPT

parameters and the proportion of invariable sites were estimated from the datasets. Tree reconstructions were performed using PhyML 3.0 (Guindon et al., 2010; Guindon and Gascuel, 2003) from SeaView version 4 (Gouy et al., 2010) and validate with 1000 bootstrap replicates.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

3. Results and discussion

3.1. RNAseq assemblies

A total of 102,119,756 paired-end raw reads with read lengths of 100 bp were generated. After data cleaning to remove adapters and poor quality parts, 100,923,981, high quality paired reads were obtained. Reads were deposited in Sequence Read Archive (SRA) under the references SRR5138508; SRR5138509.

Based on these high-quality reads, contigs were assembled into a first assembly of 275,284 transcripts (corresponding to 185,677 Trinity « genes ») with lengths ranging from 201 to 24,080 bp, an average length of 1008.5 bp, and a mean length of 418 bp. 91.2% of the cleaned reads were remapped successfully to the full transcriptome indicating a strong support of the assembled transcriptome by the reads. Lowly expressed transcripts (FPKM <1) and rare isoforms (< 1%) were excluded from the initial assembly leading to a filtered assembly of 62852 transcripts (corresponding to 40,302 Trinity « genes »), with lengths ranging from 201 to 18,966 bp, an average length of 1382.9 bp, and a mean length of 807 bp.

3.2 Peptides families identified

Most of the sought peptides on the basis of their supposed presence in insects or crustaceans were found in the transcriptome of *C. antarcticus*. The majority of the precursors are full length. Thus, 55 peptide precursors were obtained (Table 1). They code for 111 different mature peptides (Table 2). Many precursor-related peptides (PRP) were present as well but they are not listed here to focus on the known peptide families. The main neuropeptide and peptidic hormone families are described alphabetically below.

3.2.1 Adipokinetic hormone-corazonin-like peptide (ACP)

Two ACP transcripts were found coding for two putative precursors of the ACP respectively with 97 and 100 residues (Figure 1A), unlike the lobster, the crayfish and the prawn where one alone transcript has been found (Christie et al., 2015; Christie et al., 2017; Suwansa-Ard et al., 2016; Veenstra, 2015). The deduced precursor sequences were different except for the mature ACP itself (pQITFSRSWVPQa) (Figure 1A), which remains identical. It is conserved within the decapoda in which it has been characterized until now.

3.2.2 Allatostatin family (AST)

The allatostatins are neuropeptides implicated in the inhibition of the synthesis of juvenile hormone by the *corpora allata* in insects. However this family is widely distributed throughout the animal kingdom (Bendena et al., 1999), including the crustaceans. In the latter, these peptides appear to target, in the absence of juvenile hormone, methyl farnesoate and

ACCEPTED MANUSCRIPT

farnesoic acid in the crustacean equivalent of the *corpora allata*, the mandibular organ. Three types of peptides belonging to the allatostatins have been defined (Figure 1B, C, D):

- Allatostatin A (AST-A or FGL amide)

The members of this first family are characterised by a C-terminus with the structure: F/Y-X-F-G-L-amide. A single complete precursor was characterised in the *C. antarcticus* database with a putative precursor sequence of 616aa that contains a signal peptide of 27 residues. 32 sequences containing the AST-A signature were present in this precursor distributed in 25 different peptides (Figure 1B). Each of the sequences appeared to be of a unique origin, which is in contrast to analyses in *Macrobrachium rosenbergii* (Yin et al., 2006) and *Procambarus clarkii* (Yasuda-Kamatani and Yasuda, 2006) where two AST-A genes are present numerous times, indicating multiple gene duplication events. Most of them have 8 residues (21/25). The number of AST-A-like sequences in the precursor is in line with the mean of the observations made in the crustaceans (Christie et al., 2015; Christie et al., 2008; Yasuda-Kamatani and Yasuda, 2006; Yin et al., 2006).

- Allatostatin B (AST-B or $X_nW(X_6)$Wamide)

The pre-pro-peptide is full length with a 345aa sequence that contains a 25aa signal peptide. 10 different forms can be identified that place *C. antarcticus* between *Carcinus maenas* (Ma et al., 2009b; Stay and Tobe, 2007) and *Cancer borealis* (Szabo et al., 2011)) which own 13 and 9 peptides respectively (Figure 1C).

- Allatostatin C (AST-C or X_nCX_6CF)

Three isoforms from three different genes have been detected (Figure 1D). They were named according to *Daphnia pulex* AST-C designation (Dircksen et al., 2011). These isoforms possess the disulfide bridge characteristic of allatostatin-C but also the signature motif -AVSCF for two of them and the motif -PISCF for the third one. The sequence (SYWKQ**C**AFN**AVSCF**a), which is particularly well conserved in both crustaceans and insects, has been found (AST-C1). An isoform with the –PISCF motif (pQIRYHQ**C**YFN**PISCF**) has been found too (AST-C3). That confirms the hypothesis according to which these two forms might be present within decapods as their presence in *Homarus americanus* (Christie et al., 2015; Stemmler et al., 2010) and *Cancer borealis* (Ma et al., 2009a) tended to suggest. The third sequence (AST-C2) characterized in the transcriptome of *C. antarcticus* is the longest one and finish with the same motif than AST-

ACCEPTED MANUSCRIPT

C1 but potentially without amidation. This is the first time that this sequence is highlighted apart from insects and *Daphnia* and recently *H. americanus*.

3.2.3 CCHamide

Two transcripts of different length were identified as coding for CCHamide precursors. The shortest (137aa) codes for a CCHa1 of 13 residues whose sequence is well preserved compared to that obtained in lobster or crayfish with a single residue change (Figure 2D). The second, longer (221aa), carries a CCHa2 of 19 residues corresponding to the long form also found in the two species of Astacidae cited before. Most of the variations among the potential orthologous sequences of this second form are restricted to the first five N-terminal residues.

3.2.4 Crustacean hyperglycemic hormone family (CHH)

As molecular investigation techniques gain in performance, the diversity of the peptides of the CHH family becomes more complex, confirming the important role of this family in the physiology of arthropods. Thus, no less than eight different sequences have been extracted from the transcriptome of *C. antarcticus*.

- CHH stricto sensu

The two non-spliced (CHH1L) and spliced (CHH1) isoforms were encoded in 5 and 4 transcripts respectively, whereas the other CHH isoforms have been found only in a single transcript (Figure 2E). CHH1 and CHH2 had conventional structures for crustacean CHHs. They were 71 aa long and were close to the sequence level. The CHH3 was more divergent even though the characteristics of the family were respected. As a proof, it preferentially blasted with the gill form of *M. rosenbergii* that has always been placed at a particular position in the trees built from CHH isoforms. The CHH4 was coded by a partial precursor and the first four residues of the mature peptide were missing. However, beside the fact the six cysteines were present at the correct place, the sequence was longer than the other CHHs in particular with 16 residues (instead of 15) between the first two cysteines, as MIH. However it was not a glycine, which is a MIH signature, but a tyrosine. Such a sequence was found in the CHH of another shrimp, *Pandalopsis japonicus* (AFG16934.1)(Jeon et al., 2012), attesting to the reality of the assembled transcript.

- Molt inhibiting hormone/Vitellogenesis inhibiting hormone (MIH/VIH)

Three transcripts were identified as encoding two different putative full-length MIH precursors. They both have a 32aa signal peptide and the characteristic Gly_{12}. The FPKM

ACCEPTED MANUSCRIPT

values of the two isoforms appear to show that MIH/VIH2 is more strongly expressed than MIH/VIH1, suggesting a different function or a different regulation (Figure 2D).

- *Ion transport peptide like (ITP-like)*

Three transcripts potentially encoded one precursor belonging to CHH peptide family (Figure 3). Indeed, the six conserved cysteines were in place and the sequence could be aligned with the other members of the family. However, there were significant differences too (Figure 4). There were no PRP sequence and dibasic cleavage site that are characteristics of the CHHs and ITPs, but there was no Gly_{12}, which is the MIH/VIH signature. The sequence is longer than classical CHH family peptide with 84aa. However, the blast hits clearly pointed out an ITP membership clustering with insects and *Daphnia* ITPs and with *P. clarkii* ITP, the first similar isoform evidenced in decapods (Manfrin et al., 2015). If ITPs had been detected in *Daphnia*, it seemed until now that there was exclusion between CHHs and ITPs since no ITP had been detected until recently in decapods. Manfrin et al. (2015) have raised the problem for the first time with the demonstration of such a peptide in the crayfish. The characterization of a peptide clearly related to the Prc-ITP tends to confirm its existence and to invalidate reciprocal exclusion. Moreover, the FPKM values were far from being negligible. They highlight an important expression and thus a functional implication of this form, which globally was more expressed than the CHH. It is also interesting to note that this category of peptides can exist with several isoforms, as seems to attest the identification of two sequences in the crab *Metograpsus thukuhar* as well as in the shrimp *Antecaridina lauensis*. In order to better understand the phylogenetic position of this new type of peptide of the CHH family, we have therefore sought of similar peptides in the available or unpublished databases graciously made available, in particular in carideans and a crab. The results obtained will be discussed in detail in another chapter of this publication.

3.2.5 Crustacean female sex hormone (CFSH) like

The new peptide hormone recently discovered in the crab *Callinectes sapidus* (Zmora and Chung, 2014) has also been characterized in *C. antarcticus*. Like *Procambarus clarkii*, three isoforms were obtained from the transcriptomic data (Veenstra, 2015). The observation of the alignment seems to attest to the existence of at least two types of isoforms possessing either strictly the 8 cysteine residues involved in the creation of the 4 di-sulphide bridges characteristic of the family or 2 additional cysteines at N-terminal extremity (Figure 5 A).

ACCEPTED MANUSCRIPT

3.2.6 Neuroparsin (NP)

The neuroparsins were originally discovered in the locust *Locusta migratoria* due to their inhibitory effect on vitellogenesis via the neurosecretory cells of the *pars intercerebralis-corpora cardiaca* complex (Girardie et al., 1987; Moreau et al., 1988). They are fairly large peptides, often over 100aa and possess at least 12 cysteines, making them one of the most cysteine-rich neurohormone families. With six disulfide bridges, these peptides structurally resemble the insulin-like growth factor binding proteins (IGFBP) of vertebrates. Three full-length transcripts potentially coding for NP precursors were characterized within the assembly (Figure 6B). The sizes are 97, 99 and 100 residues respectively, with signal peptides counting 22, 25 and 26 aa. The three mature isoforms showed the same number of cysteine residues (12), with one ending the C-ter sequence, suggesting the presence of disulfide bridges. Like the cysteines, most of the glycine residues are well conserved too. The number of isoforms seems quite variable from one species to another or according to the depth of the transcriptomes obtained, including four isoforms in lobster (Christie et al., 2017), three in the crayfish (Veenstra, 2015) or two in the *Macrobrachium* prawn (Suwansa-Ard et al., 2015).

3.2.7 Neuropeptide F (NPF)

The naming of Neuropeptide F originates from the consensus C-terminal sequence found in all family members (-R-X-R-Famide) (Maule et al., 1991). Members of this neuropeptide family are highly conserved throughout the animal kingdom, in particular in mammals where they are called neuropeptide Y (NPY) (Nassel and Wegener, 2011). In *C. antarcticus*, five precursors were identified (Figure 6C, D). The first of these encoded a putative 100aa protein, including a 29aa signal peptide. Cha-NPF1 was encoded immediately after the signal peptide and ended at position 62 with a glycine, which permits amidation of the C-terminal with the production of a mature peptide of 32aa. This sequence was followed by a PRP, which exists in the propeptide of other decapods too. The other three transcripts corresponded to potentially spliced isoforms since they possessed identical sequences to Cha-NPF1 described above but extended to the level of the neuropeptide F itself, thus creating a long form (NFP1L), or at the level of the PRP (NPF1') or at the level of the two sequences simultaneously (NPF1L') (Figure 6C). Such situation has previously been reported, at least for the NPF itself, in the krill *E. crystallorophias* (Toullec et al., 2013) as in the lobster *Homarus americanus* (Christie et al., 2017). However, it is the first time that such splicing was reported at the level of the PRP. The last transcript encoded a clearly different precursor sequence (NPF2) (Figure 6D). The NPF2 sequence is long with 61aa and follows a signal peptide of

ACCEPTED MANUSCRIPT

27aa and precedes a short PRP with 18 residues. Similar sequences have been found in other crustaceans or insects.

3.2.8 Orcokinin/Orcomyotropin

Two partial transcripts were extracted from the assembly (Figure 6E). The first one encoded a 106aa sequence that contains a 21aa signal peptide followed by a 25aa PRP, then a 11aa orcomyotropin sequence and three identical orcokinin sequences (NFDEIDRSGFGFN) separated by dibasic cleavage sites. The second transcript represented rather the C-terminal part of a precursor. However three different potential orcokinins were identified in this sequence. The variations were observed at the level of the eighth and last residues.

3.2.9 Pigment dispersing hormone (PDH)

Nine transcripts were extracted from the assembly potentially encoding for 6 different precursors counting from 79 to 81 residues for the five full-length sequences (Figure 6F). The 6 mature PDHs are designated α and β and share a conserved structure with a mature peptide of 18aa (Rao, 2001). Five α isoforms and one β were characterized. Whether the number of precursors is similar to that found in *Macrobrachium rosenbergii* (Suwansa-Ard et al., 2015), each precursor however codes for a different PDH sequence. The diversity of isoforms is the largest found in decapods to date. It is also interesting to note that the number of isoforms highlighted in the lobster eyestalks was only two (Christie et al., 2017). It is then very likely that not all these isoforms are originating from this tissue and some isoforms are expressed in peripheral neuroendocrine tissues. The expression levels are clearly different among these forms. PDH3α and especially PDHβ are, according to the FPKM values, the most highly expressed in the extracts, all peptides taken together. This observation confirms the one previously carried out on krill (Toullec et al., 2013) and highlights the functional importance of these peptides.

3.2.10 Short neuropeptide F/Y

In the *C. antarcticus* transcriptome, two transcripts coding for the same pre-pro-peptide were identified. Only one was full-length and a 167aa precursor sequence was deduced (Figure 7C). The signal peptide was 25aa. The characteristic C-terminal sequence X_n-P-X_2-R-L-R-Fa was found in three peptides separated by cleavage sites. A forth peptide could belong to this family, except it was ending with a Y rather the expected F. So, like for NPFs that can exist with the variant NPY, especially in mammals, sNPY could be a variant in shrimp as well. It is the first time this type of sNP is reported in arthropods.

ACCEPTED MANUSCRIPT

3.2.11 Tachykinin-related peptide (TKRP)

There were two transcripts coding for two precursors where identical sequences of TKRP were present (APSGFLGMRa). These precursors were identical with the exception of first part including signal peptides (Figure 7F). This part is different in sequence but in number of residues too. The two precursors are likely spliced variants of the same gene as previously mentioned for the lobster (Christie et al., 2017). The 39aa missing within one precursor included a TKRP. Thus, the longer one coded for seven sequences distinguished by classical dibasic cleavage sites while the shorter carried only six copies.

4. Focus on CHH peptide family

In recent years, the multiplication of transcriptome explorations in an increasing number of crustaceans has allowed us to deepen our knowledge of peptidomes and particularly of the families of key peptides such as CHH. As mentioned above, the number of members of this family is steadily expanding with the result that our vision is seriously complicated and even undermines our understanding of the structural and functional diversity of these peptides. The aim of this section is to try to make a review of the sequences of CHHs, MIHs and other ITPs available in the euarthropods by integrating recently available sequences, and more especially the ITP-like, in order to support the existence of this new type of peptide in Malacostraca. Overall, the state of the art relies on the existence of two types of peptides built on a similar architecture based on the presence of six cysteines that are particularly well preserved and are at the origin of three disulfide bridges. The type 1 groups the CHHs and the ITPs due in particular to the presence of a CPRP and a dibasic cleavage site upstream of mature peptide sequences and the type 2 groups together the MIH/VIH/MOIH without CPRP and with an additional glycine in position 5 after the first cysteine (Webster et al., 2012). Moreover, CHH has been found only in malacostracans and ITP has been first characterized in insects before in non-malacostracan crustaceans (like *Daphnia* or copepods).

4.1 Malacostracan ITP-like characteristics

Exploring the transcriptome of *C. antarcticus*, five CHH isoforms were found as well as 2 MIH/VIH and a new member of the CHH family, which blasted with ITP peptides from insect and *P. clarkii* (Manfrin et al., 2015). This ITP-like has an intermediate structure, which distinguishes it from both of the two types described above. Indeed, there is no dibasic cleavage site generating a CPRP. The sequence appears to begin just after the signal peptide with a longer N-terminal structure. This characteristic would make this isoform resembling an

ACCEPTED MANUSCRIPT

MIH/VIH. However, the characteristic glycine is also absent from the sequence, invalidating the hypothesis. This particular structure is clearly very similar to that found in *P. clarkii* and thus supports the reality of the assembly. The next step was therefore to investigate potentially orthologous sequences in new transcriptomes, in a first time from shrimps, available online or kindly provided by other researchers.

So, eight new sequences close to ChaITP-like peptide were obtained from different species such as shrimps, an amphipod (*Hyalella azteca*) or a crab (*Metograpsus thukuhar*) with two different isoforms. The different sequences of the four different members of the family were aligned and the residues significantly present in the sequences were collected in a synthetic figure in order to highlight the shared and specific structures of each potential paralogous (Figure 4). It is necessary to relativize the image given by this theoretical alignment since the numbers of sequences used are not equivalent. Thus the theoretical sequence of ITP-like peptide is based on nine sequences against several decades for each of the other three isoforms. Nevertheless, this alignment makes possible not only to highlight the signatures belonging to the family but also to identify the specific differences of these ITP-like peptides. So the global structure seems conserved. The DiANNA analyses have confirmed the positions of disulfide bridges. The N-terminal portion is particularly variable in length, but is always longer than that observed on the other three members with a well-preserved pattern (x_nPxT/SxEF). Thus it appears that these ITP-like would constitute a fourth form of peptide belonging to the CHH family. Considering, as for type 1, that the absence or the presence of a CPRP is decisive, this new group of peptides would integrate with the set of type 2 peptides.

4.2 CHH phylogeny

In order to validate this proposition, it seemed pertinent to confirm the structural observations by a phylogenetic study of the different members of the family. Manfrin et al. (2015) had already demonstrated the originality of the crayfish sequence by such an approach, but at a limited level. The new analysis presented here was carried out on a set of 202 sequences integrating new sequences from carideans, isopods and a crab (Figure 8). A study of this magnitude had not been carried out since the advent of new high-throughput sequencing techniques. In addition to confirming or denying the reality of this additional type of peptide, it will also enable us to update our vision of the diversity of CHH family members. If the length of the sequences used does not permit to obtain values of nodes that are always well supported, the fact is that the four sets of paralogous are clearly distinguishable with the confirmation that the new ITP-like is positioned distinctly from the other three isoforms.

ACCEPTED MANUSCRIPT

The ITP groups together, besides chelicerate sequences, insect and phyllopod sequences confirming, if needed, the hypothesis according to which the cladocerans and more generally the phyllopods would position at the base of insects. It is interesting to note that copepods whose position is often fluctuating associated with them. They also have the most important branch length attesting of an important evolution rate (Figure 9).

The multiplication of available sequences both supports established phylogeny and highlights the existence of isoforms, which makes it more complex and allows us to see an ever-increasing diversity. This is clearly the case for the CHHs for which at least a second set of peptides seems to take shape in this phylogeny, and potentially more with respect of the sequences that do not fit into any of the two defined sets (Figure 9). It is not surprising to note that the marginalized sequences in the tree are all obtained from transcriptomes, confirming the power of the technique and anticipating the importance of its future contributions in our understanding of the evolution of this family of peptides. If the existence of chimeras cannot be completely excluded, the grouping of sequences from several species on a branch according a parallel phylogeny seems to validate their reality (Figure 9).

5 Conclusions

Today, the study of the biology or physiology of a new species needs the primordial steps of sequencing and assembly of its transcriptome, thus constituting a true identity card. Besides the interest of identifying the potential actors of the main biological functions studied, the exploration of the transcriptome allows us to deepen our knowledge of the diversity and evolution of these same actors. The peptidome of *Chorismus antarcticus* does not escape this rule, especially because relatively few caridean peptidomes have been studied to date. This study described new mature peptide sequences (101) including in most of the cases the encoded pre-pro-peptides (55). Apart from the notion of the absence or presence of potentially orthologous sequences of crustaceans or insects, the functionality of these peptides remains purely speculative or is purely and simply unknown. This is particularly true since the more an analysis gains in depth, the more the number of paralogues highlighted tends to increase without precise information on a conservation or a modification of the function. The functional labeling of these new isoforms, which attests, by their number, the importance of the original function, cannot be done without experimental verification and constitutes a new challenge.

ACCEPTED MANUSCRIPT

Acknowledgments

JYT would like to thank all the IPEV staff at the base Dumont d'Urville and especially A. Pottier for his help for fishing during different Antarctic campaigns in Terre Adélie. JYT benefited from funding provided by Institut Paul Emile Victor (KREVET program) and also by la Région Bretagne (SAD1 - DRAKAR program). CYL is sponsored by Ministry of Science and Technology, Taiwan (Grant number MOST 104-2311-B-018-002).

Legends

Figure 1:
Complete and partial sequences from *Chorismus antarcticus* of the pre-pro-peptides containing: A) Adipokinetic-corazonin peptides (ACP); B) Allatostatins A; C) Allatostatins B; D) Allatostatins C; E) Bursicon alpha. The mature peptides are highlighted in blue, the potential dibasic cleavage sites in red, the signal peptides in green and the precursor related peptides (PRP) in yellow.

Figure 2:
Complete and partial sequences from *Chorismus antarcticus* of the pre-pro-peptides containing: A) Calcitonin-Like Diuretic Hormone (CLDH 31); B) Corticotropin Related Factor LIKE Diuretic Hormone (CRFLD45); C) Crustacean Cardioactive Peptide (CCAP); D) CCHamide; E) Crustacean Hyperglycemic Hormone (CHH); F) Molt/Vitellogenesis Inhibiting Hormone (MIH/VIH). The mature peptides are highlighted in blue, the potential dibasic cleavage sites in red, the signal peptides in green and the precursor related peptides (PRP) in yellow.

Figure 3:
Alignment of the protein sequences of the ITP-like pro-peptides from various malacostracans. The mature peptides are highlighted in blue, the potential dibasic cleavage sites in red and the signal peptides in green. The sequences are grouped by similarity.

Figure 4:
Alignment of the protein sequences of the CHH family members. Beside conserved cysteines, bold letters highlight a totally or mainly conserved amino acid. Full line boxes show strictly conserved amino acid among CHH family members. Black full lines represent disulfide bridges. The dibasic cleavage sites are in red

ACCEPTED MANUSCRIPT

Figure 5:
Complete and partial sequences from *Chorismus antarcticus* of the pre-pro-peptides containing: A) Crustacean Sex Female Hormone (CSFH); B) Corazonin (CRZ), C) Eclosion Hormone; D) FLRFamide peptides; E) GSEFLamide peptides; F) Intocin; G) Leucokinin. The mature peptides are highlighted in blue, the potential dibasic cleavage sites in red, the signal peptides in green and the precursor related peptides (PRP) in yellow.

Figure 6:
Complete and partial sequences from *Chorismus antarcticus* of the pre-pro-peptides containing: A) Myosuppressin; B) Neuroparsins; C) Neuropeptides F1; D) Neuropeptide F2; E) Orkomyotropins and Orkokinins; F) Pigment dispersing Hormones (PDH). The mature peptides are highlighted in blue, the potential dibasic cleavage sites in red, the signal peptides in green and the precursor related peptides (PRP) in yellow.

Figure 7:
Complete sequences from *Chorismus antarcticus* of the pre-pro-peptides containing: A) Pyrokinins; B) Red pigment Dispersing Hormone (RPCH); C) small Neuropeptides F; D) SIFamide; E) Sulfakinins; F) Tachykinin Related Peptide (TKRP). The mature peptides are highlighted in blue, the potential dibasic cleavage sites in red, the signal peptides in green and the precursor related peptides (PRP) in yellow.

Figure 8:
Circular phylogenetic tree built using Maximum Likehood and Bayesian Inference from an alignment of 202 sequences of CHH family peptides with 71 sites. Chelicerate sequences were assigned as outgroup. Numbers above branches are bootstrap values (based on 1000 replicates) and posterior probabilities (italic) obtained from the analysis of the amino acid dataset.

Figure 9:
Synthetic representation of the tree represented figure 8, where sequence clusters are collapsed.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

References

Abascal, F., Zardoya, R., Posada, D., 2005. ProtTest: selection of best-fit models of protein evolution. Bioinformatics 21, 2104-2105.

Arntz, W.E., Gorny, M., 1991. Shrimp (Decapoda, Natantia) occurrence and distribution in the eastern Weddell Sea, Antarctica. Polar Biology 11, 169-177.

Basher, Z., Bowden, D.A., Costello, M.J., 2014. Diversity and distribution of deep-sea shrimps in the Ross Sea region of Antarctica. Plos One 9, e103195.

Bendena, W., Donly, B., Tobe, S., 1999. Allatostatins: a growing family of neuropeptides with structural and functional diversity. Ann. N. Y. Acad. Sci 897, 311-329.

Cascella, K., Jollivet, D., Papot, C., Leger, N., Corre, E., Ravaux, J., Clark, M.S., Toullec, J.Y., 2015. Diversification, evolution and sub-functionalization of 70kDa heat-shock proteins in two sister species of antarctic krill: differences in thermal habitats, responses and implications under climate change. Plos One 10, 1-23.

Christie, A.E., 2016a. Expansion of the neuropeptidome of the globally invasive marine crab *Carcinus maenas*. Gen Comp Endocrinol 235, 150-169.

Christie, A.E., 2016b. Prediction of *Scylla olivacea* (Crustacea; Brachyura) peptide hormones using publicly accessible transcriptome shotgun assembly (TSA) sequences. Gen Comp Endocrinol 230-231, 1-16.

Christie, A.E., Chi, M., Lameyer, T.J., Pascual, M.G., Shea, D.N., Stanhope, M.E., Schulz, D.J., Dickinson, P.S., 2015. Neuropeptidergic Signaling in the American Lobster *Homarus americanus*: New Insights from High-Throughput Nucleotide Sequencing. Plos One 10, e0145964.

Christie, A.E., McCoole, M.D., Harmon, S.M., Baer, K.N., Lenz, P.H., 2011. Genomic analyses of the *Daphnia pulex* peptidome. Gen Comp Endocr 171, 131-150.

Christie, A.E., Pascual, M.G., 2016. Peptidergic signaling in the crab *Cancer borealis*: Tapping the power of transcriptomics for neuropeptidome expansion. Gen Comp Endocrinol 237, 53-67.

Christie, A.E., Roncalli, V., Cieslak, M.C., Pascual, M.G., Yu, A., Lameyer, T.J., Stanhope, M.E., Dickinson, P.S., 2017. Prediction of a neuropeptidome for the eyestalk ganglia of the lobster *Homarus americanus* using a tissue-specific de novo assembled transcriptome. Gen Comp Endocrinol 243, 96-119.

Christie, A.E., Sousa, G.L., Rus, S., Smith, C.M., Towle, D.W., Hartline, D.K., Dickinson, P.S., 2008. Identification of A-type allatostatins possessing -YXFGI/Vamide carboxy-termini from the nervous system of the copepod crustacean *Calanus finmarchicus*. Gen Comp Endocrinol 155, 526-533.

Dircksen, H., Neupert, S., Predel, R., Verleyen, P., Huybrechts, J., Strauss, J., Hauser, F., Stafflinger, E., Schneider, M., Pauwels, K., Schoofs, L., Grimmelikhuijzen, C.J.P., 2011. Genomics, Transcriptomics, and Peptidomics of *Daphnia pulex* Neuropeptides and Protein Hormones. J Proteome Res 10, 4478-4504.

Finn, R.D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Heger, A., Hetherington, K., Holm, L., Mistry, J., Sonnhammer, E.L., Tate, J., Punta, M., 2014. Pfam: the protein families database. Nucleic Acids Res 42, D222-230.

Genomic Resources Development Consortium, C., Havird, J.C., Santos, S.R., 2014. Genomic Resources Notes accepted 1 June 2014 - 31 July 2014. Mol Ecol Resour 14, 1322.

Girardie, J., Boureme, D., Couillaud, F., Tamarelle, M., Girardie, A., 1987. Anti-Juvenile Effect of Neuroparsin-a, a Neuroprotein Isolated from Locust *Corpora Cardiaca*. Insect Biochem 17, 977-983.

Gorny, M., 1999. On the biogeography and ecology of the Southern Ocean decapod fauna. Scientia Marina 63, 367-382.

ACCEPTED MANUSCRIPT

Gouy, M., Guindon, S., Gascuel, O., 2010. SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Mol Biol Evol 27, 221-224.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B.W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., Regev, A., 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol 29, 644-652.

Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O., 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol 59, 307-321.

Guindon, S., Gascuel, O., 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol 52, 696-704.

Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., et al., 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nature Protocols 8, 1494–1512.

Jeon, J.M., Kim, B.K., Lee, J.H., Kim, H.J., Kang, C.K., Mykles, D.L., Kim, H.W., 2012. Two type I crustacean hyperglycemic hormone (CHH) genes in Morotoge shrimp (*Pandalopsis japonica*): cloning and expression of eyestalk and pericardial organ isoforms produced by alternative splicing and a novel type I CHH with predicted structure shared with type II CHH peptides. Comp Biochem Physiol B Biochem Mol Biol 162, 88-99.

Krogh, A., Larsson, B., von Heijne, G., Sonnhammer, E.L., 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol 305, 567-580.

Langmead, B., Trapnell, C., Pop, M., Salzberg, S.L., 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biology 10.

Li, B., Dewey, C.N., 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. Bmc Bioinformatics 12.

Ma, M., Szabo, T.M., Jia, C., Marder, E., Li, L., 2009a. Mass spectrometric characterization and physiological actions of novel crustacean C-type allatostatins. Peptides 30, 1660-1668.

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

Manfrin, C., Tom, M., De Moro, G., Gerdol, M., Giulianini, P.G., Pallavicini, A., 2015. The eyestalk transcriptome of red swamp crayfish *Procambarus clarkii*. Gene 557, 28-34.

Martin, M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal 17, 10-12.

Maule, A.G., Shaw, C., Halton, D.W., Thim, L., Johnston, C.F., Fairweather, I., Buchanan, K.D., 1991. Neuropeptide-F - a Novel Parasitic Flatworm Regulatory Peptide from Moniezia-Expansa (Cestoda, Cyclophyllidea). Parasitology 102, 309-316.

Moreau, R., Gourdoux, L., Girardie, J., 1988. Neuroparsin - a New Energetic Neurohormone in the African Locust. Arch Insect Biochem 8, 135-145.

Nassel, 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, 1335-1355.

Peck, L.S., 2004. Physiological flexibility: the key to success and survival for Antarctic fairy shrimps in highly fluctuating extreme environments. Freshwater Biol 49, 1195-1205.

Peck, L.S., Morley, S.A., Clark, M.S., 2010. Poor acclimation capacities in Antarctic marine ectotherms. Marine Biology 157, 2051-2059.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

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

Portner, H.O., Peck, L., Somero, G., 2007. Thermal limits and adaptation in marine Antarctic ectotherms: an integrative view. Philosophical transactions of the Royal Society of London. Series B, Biological sciences 362, 2233-2258.

Rao, K.R., 2001. Crustacean pigmentary-effector hormones: Chemistry and functions of RPCH, PDH, and related peptides. Am Zool 41, 364-379.

Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572-1574.

Stay, B., Tobe, S.S., 2007. The role of allatostatins in juvenile hormone synthesis in insects and crustaceans. Annu Rev Entomol 52, 277-299.

Stemmler, E.A., Bruns, E.A., Cashman, C.R., Dickinson, P.S., Christie, A.E., 2010. Molecular and mass spectral identification of the broadly conserved decapod crustacean neuropeptide pQIRYHQCYFNPISCF: The first PISCF-allatostatin (*Manduca sexta*- or C-type allatostatin) from a non-insect. Gen Comp Endocr 165, 1-10.

Suwansa-Ard, S., Thongbuakaew, T., Wang, T., Zhao, M., Elizur, A., Hanna, P.J., Sretarugsa, P., Cummins, S.F., Sobhon, P., 2015. In silico Neuropeptidome of Female *Macrobrachium rosenbergii* Based on Transcriptome and Peptide Mining of Eyestalk, Central Nervous System and Ovary. Plos One 10, e0123848.

Suwansa-Ard, S., Zhao, M., Thongbuakaew, T., Chansela, P., Ventura, T., Cummins, S.F., Sobhon, P., 2016. Gonadotropin-releasing hormone and adipokinetic hormone/corazonin-related peptide in the female prawn. Gen Comp Endocrinol 236, 70-82.

Szabo, T.M., Chen, R., Goeritz, M.L., Maloney, R.T., Tang, L.S., Li, L., Marder, E., 2011. Distribution and physiological effects of B-type allatostatins (myoinhibitory peptides, MIPs) in the stomatogastric nervous system of the crab *Cancer borealis*. The Journal of Comparative Neurology, 519, 2658-2676.

Thatje, S., Arntz, W.E., 2004. Antarctic reptant decapods: more than a myth? Polar Biology 27, 195-201.

Toullec, J.Y., Corre, E., Bernay, B., Thorne, M.A., Cascella, K., Ollivaux, C., Henry, J., Clark, M.S., 2013. Transcriptome and Peptidome Characterisation of the Main Neuropeptides and Peptidic Hormones of a Euphausiid: The Ice Krill, *Euphausia crystallorophias*. Plos One 8, e71609.

Veenstra, J.A., 2015. The power of next-generation sequencing as illustrated by the neuropeptidome of the crayfish *Procambarus clarkii*. Gen Comp Endocrinol 224, 84-95.

Webster, S.G., Keller, R., Dircksen, H., 2012. The CHH-superfamily of multifunctional peptide hormones controlling crustacean metabolism, osmoregulation, moulting, and reproduction. Gen Comp Endocrinol 175, 217-233.

Whelan, S., Goldman, N., 2001. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. Mol Biol Evol 18, 691-699.

Yasuda-Kamatani, Y., Yasuda, A., 2006. Characteristic expression patterns of allatostatin-like peptide, FMRFamide-related peptide, orcokinin, tachykinin-related peptide, and SIFamide in the olfactory system of crayfish *Procambarus clarkii*. Journal of Comparative Neurology 496, 135-147.

Yin, G.L., Yang, J.S., Cao, J.X., Yang, W.J., 2006. Molecular cloning and characterization of FGLamide allatostatin gene from the prawn, *Macrobrachium rosenbergii*. Peptides 27, 1241-1250.

Zmora, N., Chung, J.S., 2014. A novel hormone is required for the development of reproductive phenotypes in adult female crabs. Endocrinology 155, 230-239.

ACCEPTED MANUSCRIPT

Table 1

Alphatical list of peptide precursors, contig expression values (FPKM) and associated BLAST matches

Peptide precursor designation	Size (aa)	Contig ID	Size (pb)	FPKM	BLAST matches
Adipokinetic-corazonin peptide 1	97	176907_c1_seq1	1612	9.1	AKH/corazonin-related peptide (*Nilaparvata lugens*) BAO00933 - 0.55
Adipokinetic-corazonin peptide 2	100	84549_c0_seq1	878	27.2	AKH/corazonin-related peptide (*Nasonia vitripennis*) NP_001161199 - 1.00e-3
Allatostatin A	616	173416_c7_seq2	4356	55.9	Type A pre-pro-allatostatin (*Machrobrachium rosenbergii*) AAY82901 - 0.00
		173416_c7_seq3	323	16.6	
		173416_c2_seq1	1312	14.7	
Allatostatin B	345	163527_c0_seq1	1652	75.4	Type B pre-pro-allatostatin (*Scylla paramamosain*) ALQ28584 - 8.96e-86
Allatostatin C1	106	145520_c1_seq1	1140	120.3	Type C pre-pro-allatostatin (*Nilaparvata lugens*) BAO00971 - 2.62e-24
Allatostatin C2	96	174424_c0_seq1	668	44.5	Type C pre-pro allatostatin, (*Nilaparvata lugens*) BAO00935.1 - 9.07e-7
Allatostatin C3	148	171290_c0_seq1	2798	16.5	Type C pre-pro allatostatin, (*Neocaridina denticulata*) AIY69122.1 - 1.76e-10
Bursicon α□	147	103777_c0_seq1	794	0.3	Bursicon hormone alpha subunit (*Penaeus monodon*) AKJ74864 - 7.02 e-79
□ursicon β	87 partial	144574_c2_seq1	572	0.24	Bursicon hormone beta subunit (*Homarus gammarus*) ADI86243 - 2.23 e-48
C□□□	137	145611_c0_seq1	1115	105.1	Crustacean cardioactive peptide (*Procambarus clarkii*) BAF34910 - 2.86e-52
CCH1	132	167871_c0_seq1	1386	7.6	CCHamide 1 (*homarus americanus*) GFDA01105168.1- 2e-20
CCH2	221	171770_c0_seq6	1345	0.8	CCHamide (*homarus americanus*) GFDA01145210.1 - 3e-14
CHH1	147	176012_c10_seq2	2144	0.1	CPRP/CHH precursor (*Pandalopsis japonica*) AFG16933.1 - 6.84e-56
		176012_c10_seq4	2070	0.5	
		176012_c10_seq8	2086	34.4	
		176012_c10_seq9	2128	80.8	
CHH1L	146	176012_c10_seq1	2191	0.13	Hyperglycemic hormone (*Pandalopsis japonica*) AFG16932.1 - 3e-61
		176012_c10_seq3	2249	16.6	
		176012_c10_seq5	2035	2.6	
		176012_c10_seq6	2207	6.6	
		176012_c10_seq7	2265	0.1	
CHH2	130	176651_c0_seq4	816	3.8	CHH isoform 2 (*Rimicaris kairei*) ACS35347 - 1.49e-35
CHH3	147	162039_c2_seq1	865	1.5	CHH gill form (*Macrobrachium rosenbergii*) AAL40916 - 2.54e-18
CHH4	73 partial	1025550_c0_seq1	320	0.3	Hyperglycemic hormone (*Pandalopsis japonica*) AFG16934.1 - 6e-12
CFSH-like1	224 partial	157251_c0_seq3	677	12.2	Crustacean female sex hormone precursor (*C. sapidus*) ADO00266 - 6e-29
CFSH-like2	239	148772_c1_seq1	940	1.5	Crustacean female sex hormone precursor (*C. sapidus*) ADO00266 - 2e-31
CFSH-like3	319	148623_c0_seq3	1361	6.1	Crustacean female sex hormone precursor (*C. sapidus*) ADO00266 - 1e-6
CLDH31	141	145195_c0_seq1	1025	53.6	Prepro-calcitonin-like diuretic hormone (*H. americanus*) ACX46386 - 2.32e-59
CRFLDH45	140	177023_c0_seq1	723	0.2	Corticotropin releasing factor-like protein (*P. americana*) ALG35940 - 1.98e-5
Corazonin	111	83574_c0_seq1	1092	0.3	Corazonin (*Macrobrachium rosenbergii*) ALA65535 - 4.66e-7
Eclosion hormone	82	175814_c1_seq1	332	207.2	Eclosion hormone (*Scylla paramamosain*) ALQ28581 - 9.16e-31
FLRFamide	380	173443_c7_seq1	1348	0.7	FLRFamide (*Scylla paramamosain*) ALQ28593 - 2.09e-90
GSELFamide	270	170259_c1_seq2	2847	14	GSEFLamide, [*Scylla paramamosain*] ALQ28590.1 - 2e-44
Intocin	147	172469_c1_seq1	671	8,9	vasotocin-neurophysin, partial (*Scylla paramamosain*) ALQ28600.1 - 8e-29
ITP-like	118	173384_c1_seq2	1465	60.9	Ion transport protein (*Procambarus clarkii*) AIZ05253.1 - 6e-29
		173384_c1_seq3	1447	126.7	
		173384_c1_seq7	1029	4.1	
Leucokinin	202 partial	176505_c0_seq1	2927	8.6	kinin, partial (*Scylla paramamosain*) ALQ28594.1 - 7e-35

ACCEPTED MANUSCRIPT

MIH/VIH1	111	171447_c7_seq1 171447_c7_seq2	1046 960	13.4 0.5	SGP A precursor (*Macrobrachium rosenbergii*) AAL37948 - 6.9e-57
MIH/VIH2	110	145525_c0_seq1	465	79.3	SGP B precursor (*Macrobrachium rosenbergii*) AAL37949 - 1.11e-52
Myosuppressin	107	175121_c0_seq1	1215	39.2	myosuppressin-like precursor (*Procambarus clarkii*) BAG68789.1 - 8e-39
Neuroparsin 1	97	160864_c1_seq1	650	26.6	Neuroparsin 1 (*Scylla paramamosain*) ALQ28570 - 5.00e-15
Neuroparsin 2	99	174206_c0_seq1	614	47.0	Neuroparsin (*Metapenaeus ensis*) AHX39208 - 1.57e-31
Neuroparsin 3	100	166524_c0_seq1	3527	52.3	Neuroparsin (*Jasus lalandii*) AHG98659 - 6.49e-10
Neuropeptide F1	100	160229_c2_seq4	591	6.4	Preproneuropeptide F1 (*Litopenaeus vannamei*) AEC12204 - 6.52e-28
Neuropeptide F1L	137	160229_c2_seq3	702	6.1	Preproneuropeptide F2 (*Litopenaeus vannamei*) AEC12205 - 4.99e-52
Neuropeptide F1'	90	160229_c2_seq1	561	25.6	Preproneuropeptide F1 (*Litopenaeus vannamei*) AEC12204 - 3.06e-31
Neuropeptide F1L'	127	160229_c2_seq2	672	6.1	Preproneuropeptide F1 (*Litopenaeus vannamei*) AEC12205 - 2.69e-55
Neuropeptide F2	109	180522_c0_seq1	579	54.9	Neuropeptide F1 (*Scylla paramamosain*) ALQ28586.1 - 2e-23
Orcokinin 1	106	175130_c1_seq1	441	160.1	Orcokinin precursor (*Procambarus clarkii*) Q9NL83 - 6.49e-44
Orcokinin 2	69 partial	175130_c1_seq2	1044	70.7	Prepro-orcokinin 2 (*Homarus americanus*) ACD13197 - 1.42e-25
Orcomyotropin	106	175130_c1_seq1	441	160.1	Orcokinin precursor (*Procambarus clarkii*) Q9NL83 - 6.49e-44
PDH1α	79	166116_c0_seq1 166116_c0_seq2	547 651	0.9 28.7	Pigment dispersing hormone (*Marsupenaeus japonicus*) BAE78495 - 1.80e-14
PDH□α□	70	171809_c1_seq3	598	7.2	Pigment dispersing hormone 2 (*Litopenaeus vannamei*) P91964.2 - 5.27e-20
PDH□α□	80	176495_c0_seq1 176495_c0_seq2	902 648	103.8 46.3	Pigment dispersing hormone 2 (*Litopenaeus vannamei*) P91964.2 - 5.88e-14
PDH□α□	49 partial	171809_c1_seq2 171809_c1_seq4	455 1006	4.1 9.7	Pigment dispersing hormone I (*Marsupenaeus japonicus*) BAB91010.1 - 6e-16
PDH□α	80	82278_c0_seq1	649	3.7	Pigment dispersing hormone 2 (*Litopenaeus vannamei*) P91964.2 - 3e-10
PDHβ	74	155387_c0_seq1	473	352.1	Pigment dispersing hormone precursor (*L. vannamei*) CAA72409 1.54e-15
Pyrokinin	357	171276_c0_seq1	1733	3.52	Pyrokinin precursor (*Scylla paramamosain*) ALQ28575.1 - 7e-37
RPCH/AKH	97	165820_c2_seq1	863	96.9	Red pigment concentrating hormone (*M. rosenbergii*) ABV46765 - 9.20e-34
SIFamide	76	172635_c15_seq2	523	182.9	SIFamide (*Scylla paramamosain*) ALQ28576 - 7.91e-25
sNPF	167 122 partial	163533_c0_seq2 163533_c0_seq1	940 618	33.8 43.3	Short neuropeptide F precursor (*Scylla paramamosain*) ALQ28574 - 3.97e-30
Sulfakinin	122	89102_c0_seq1	819	11.6	Preprosulfakinin (*Homarus americanus*) ABQ95346 - 1.61e-34
Tachykinin RP	210 182	163516_c0_seq2 163516_c0_seq1	2356 2121	239.2 6.4	Preprotachykinin (*Panulirus interruptus*) BAD06363 - 3.23e-86

Contigs corresponding to the selected peptide sequences are in column three. Size (aa) refers to the coding portion derived from the assembly and size (bp) refers to total size of corresponding contigs. FPKM = Fragments Per Kilobase of exon per Million fragments mapped.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Table 2 - List of mature peptides of *Chorismus antarcticus*

Peptide name	Peptide sequence	Previous identification in Arthropods	Pfam/Interpro accession N°
Adipokinetic-corazonin		Daphnia, lobster	PF00473/IPR000187
Cha-ACP	pQITFSRSWVPQa		
Allatostatins A family		Cirriped, copepod, daphnia, decapods, krill and insects	PF05953/IPR010276
Cha-AST A1	HNDYVFGLa		
Cha-AST A2	SPGYAFGLa		
Cha-AST A3	DRMYSFGLa		
Cha-AST A4	EGLYAFGLa		
Cha-AST A5	SGTYNFGLa		
Cha-AST A6	SKAFNFGLa		
Cha-AST A7	DRSYSFGLa		
Cha-AST A8	PQHYAFGLa		
Cha-AST A9	ALQYAFGLa		
Cha-AST A10	PNNYAFGLa		
Cha-AST A11	PQQYAFGLa		
Cha-AST A12	EQNYAFGLa		
Cha-AST A13	YSDDNANRMYAFGLa		
Cha-AST A14	ASSYGFGLa		
Cha-AST A15	AGKYTFGLa		
Cha-AST A16	GGSYAFGLa		
Cha-AST A17	AGYAFGLa		
Cha-AST A18	PDAYSFGLa		
Cha-AST A19	SGPYQFGLa		
Cha-AST A20	PSGSYAFGLa		
Cha-AST A21	AGQYSFGLa		
Cha-AST A22	SNPYAFGLa		
Cha-AST A23	SSPYAFGLa		
Cha-AST A24	SGSYSFGLa		
Cha-AST A25	VPGSYAFGLa		
Allatostatin B family		Shrimp, krill	
Cha-AST B1	ADWSSMRGTWa		
Cha-AST B2	SGWNKFQGSWa		
Cha-AST B3	ANWNKFQGSWa		
Cha-AST B4	DGWQNFQGSWa		
Cha-AST B5	DGWQNFQGSWa		
Cha-AST B6	NNWSSLQGTWa		
Cha-AST B7	AWQNLHGAWa		
Cha-AST B8	PQYPTRVSPRSANWSSLRGTWa		
Cha-AST B9	NADWSSLRGAWa		
Cha-AST B10	NSDWSQFKGSWa		
Allatostatin C family			
Cha-AST C1	SYWKQCAFNAVSCFa	Cirriped, daphnia, decapods and insects	
Cha-AST C2	GNNEGGRLYWRCYFNAVSCF	Insects, daphnia	
Cha-AST C3	pQIRYHQCYFNPISCF	Decapods, daphnia, insects	
Bursicon family			
Cha-Bursicon α	DECSLTPVIHILSYPGCNSKPIPSFACQGRCTSYVQVSGSKIWQT ERSCMCCQESGEREATVVLNCPKARVGDPKRRKVLTRAPVDC MCRPCTDVEEGTVLAQEIANFIADDPMAHMPFLK	Cirriped, daphnia, decapods and insects	
Cha-Bursicon β	----------RTCEEDLAVNKCEGACLSKVQPSVNTPS GFLKDCRCCRETHLRSREVILTHCYDVDGNRLVGGKGQLSLK MSEPADCQCSKCGDSTR	Cirriped, daphnia, decapods and insects	
Calcitonin-Like Diuretic H.		Ixod, Cirriped, copepod, daphnia, lobster and insects	
Cha-CLDH31	GLDLGLGRGFSGSQAAKHLMGLAAANFAGGPa		-
Corticotropin Related F.		Daphnia	PF00473/IPR000187
Cha-CRFLDH45	TSGLSLSIDASMKVLREALYLEMARKKQRQQMLRARHNQALLTTIa		
Crust. Cardioactive Pept.		Ixod, daphnia, decapods and insects	PF11105/IPR024276
Cha-CCAP	PFCNAFTGCa		
CCHamide		Insects and lobster	
Cha-CCHa1	SCSQYGHSCFGAHa		
Cha-CCHa2	RRIPKGGCLSYGHSCLGAHa		
CHH family		Daphnia, isopod, decapods and insects	PF01147/IPR001166
Cha-CHH1	AVLDQSCKGIYDRELFKKLDRVCEDCYNLYRKPYVGIDCRNNCYG NLVFRQCLDDLLLVENLDEYVNAVQMVa		
Cha-CHH1L	AVLDQSCKGIYDRELFKKLDRVCEDCYNLYRKPYVGIDCRKHCFST KTFNQCVGDLLLDEKLYTAMRDHIAYF		
Cha-CHH2	VILDQSCKGIFDRNLFRKLDRVCEDCYNLYRKPHVGIDCRSNCYGN MIFRQCLDDLMMMDVVDEYIKKVQVVa		
Cha-CHH3	SVQGSSCRGIDSRVLWNKLDRVCGDCYNLYRKAIVAIGCRKGCFTS DYFTMCVGDLLLPTKEYDIYVSALSGVW		
Cha-CHH4	----GSCKGPAYTRGLFNTLDKICDDCYNLYRKVDVDINCRKNCFGE FQFFVCLKKLKYNKTEIDELLQIGYAIAKF		
Cha-ITPlike	SFIRIRPNTYKEFQYINCQGRFDKEQYASLTNICEDCHNVYRNPDVLL GCKADCFRNSLFPKCVSMLLLDQREPELSKMVYTVS		
Cha-MIH/VIH1	RYLDDECPGVMGNRDLYEKVVRVCDDCSNIFRMNDVGSRCKKDCF YNEDFLWCVYATERHGEVDQLNRWMSILKA	Isopod, decapods	
Cha-MIH/VIH2	RFLDDECRGVMGNRDLYEKVARVCDDCVNIFRNSNVGPKCRTNCF YNEDFLWCVIATQRKNDLDQMNRSMSILRA	Decapods (peneids)	
CFSH family		Malacostraca	
Cha-CFSH1	QQYLNTDELQYFSKEQVDEASKVEFKVVPDPVIYTSQIIHKGVNCSSI RTDLHENHIRPELQLHPGWIHSSQLIGSCPTHYVTRELPPMYSPSVVV EAVCTCNGSKCSREGHQCLPVSRHIPVWVRQGPNLHVLDVEELTVA CACIRRPSESGNFIYASAVHS		
Cha-CFSH2	NREDLGGDLLQYFSEEQVKDATRAEYKVVPYPIVYTSQILHEGVNC SSIRMNLHKNHVKPELQLRPNWIHKSELIGDCPTHYVARELPPMYSP AIILEAVCTCGGSQCSRSGHQCVPVSHHVPVWVRRGPNFHVLDVEE VTVACACVRRPSGIGNFLYAAAVEN		
Cha-CFSH3	SRACVNQSQGRCRRGQVSMIPAEQVQQDWEDDYSSVPDVLIQFSQQ QAEEAACNDLSVQLFQVDLREHYLEPVWVREIVHLGMCPSKLQMR NFGKDVWPSSVVETKCLCHNQPCSNLGGDFRCQAVRRPIPTWVRH VDNFMPVQEMVTVGCVCVQRTSPEGKYAKPSVES		
Corazonin		Ixod, daphnia, decapods and insects	-/IPR020190
Cha-Arg^{7}-CRZ1	pQTFQYSRGWTNa		
Eclosion hormone		Cirriped, daphnia, decapods and insects	-/IPR006825
Cha-EH	ATITSMCIRNCGQCKEMYGDYFHGQACAESCIMTQGNSIPDCNNPA TFNRFL		

ACCEPTED MANUSCRIPT

FLRFamide		Decapods	
Cha-FLRF1	GYVDRNFLRFa		
Cha-FLRF2	GVGKNFLRFa		
Cha-FLRF3	NRNFLRFa		
Cha-FLRF4	DPDRNFLRFa		
Cha-FLRF5	GSNNFLRFa		
Cha-FLRF6	NYNKNFLRFa		
Cha-FLRF7	DRNFLRFa		
GSEFLamide		Lobster	
Cha-GSEFLa1	IGSEFLa		
Cha-GSEFLa2	MGSEFLa		
Cha-GSEFLa3	AMGSEFLa		
Intocin		Arthropods	
Cha-intocin	CFITNCPPGa		
Leucokinin		Insects, decapods	
Cha-lkn1	pQAFSAWAa		
Cha-lkn2	pQPFSAWAa		
Cha-lkn3	pQAFNAWAa		
Cha-lkn4	pQPFSPWAa		
Cha-lkn5	pQSFSSWAa		
Myosuppressin		Decapods, insects	
Cha-Myosup	QDLDHVFLRFa		
Neuroparsin family		Copepod, daphnia, krill and insects	PF07327/IPR010850
Cha-NP1	APRCTQHDLPAARKCDYGTVLDWCRNAVCAQGPGYPCGGNRWEL GKCGEGTFCSCGTCTGCSSITRECYRSALVC		
Cha-NP2	APSCSTTRHTVDEAECKYGTFVDWCRNTVCAKGPGQTCGGDWWE NGKCGEGTYCTCGICSGCSVNLECWFGTFC	Copepod, daphnia and krill	
Cha-NP3	SPLCPSSQQTDEDLSKCMYGTAIGWCGNLECAKGPGERCGGNWLE HGSCGDGMYCGCGYCAGCYIVKCATRMFC		
Neuropeptide F family		Daphnia, krill, decapods, insects	PF00159/IPR001955
Cha-NPF1	KPDPTQLAAMADALKYLQELDKYYSQVSRPRFa		
Cha-NPF1-L	KPDPTQLAAMADALKYLQELDKYYSQVSRPSTRSAPGPASQIQALE KTLKFLQLQELGKFYSLRARPRFa	Decapods and krill	
Cha-NPF2	SSARTENTAEALQAMHEAALANMLGSAEVQYPSRPNVFKSPVELRQ YLEALNAYYAIAGRPRFa	Decapod and krill	
Orcokinin		Decapods and krill	
Cha-OCK1	NFDEIDRSGFGFN		
Cha-OCK2a	NFDEIDRAGFGFY		
Cha-OCK2b	NFDEIDRQGFGFA		
Cha-OCK2c	NFDEIDRSGFGFV		
Orcomyotropin		Decapods	
Cha-OCM	FDSFTTGFGHS		
PDH/PDF family		Arthropods	PF06324/IPR009396
Cha-PDHα□	NSGMINSLLGIPRVMTAAa		
Cha-PDHα□	NSGMINSILGIPKVMAEAa		
Cha-PDHα□	NSGMINSLLGIPQVMNNAa		
Cha-PDHα□	NSGMINSLLGIPKVMTEAa		
Cha-PDHα□	AAGLINSILGIPKILVLAa		
Cha-PDHβ	NSELINSLLGLPKVMNDAa		
Pyrokinin		Decapods	
Cha-Pkn1	SPFSPRLa		
Cha-Pkn2	DELHYGLMYDDDDDDTTDMDNLRDDESDDNLFEDATSQDYTDEA VSPQRLALRSALVPRLa		
Cha-Pkn3	AIAFSPRLa		
Cha-Pkn4	GTAFIPRLa		
Cha-Pkn5	GDFAFSPRLa		
Cha-Pkn6	ADFAFSPRLa		
Cha-Pkn7	SDFAFSPRLa		
Cha-Pkn8	GNAFIPRLa		
Cha-Pkn9	DAVASSSEDTWSDNSNDVTQLQQRSVAFSPRLa		
RPCH/AKH		Daphnia, decapods and insects	PF06377/IPR010475
Cha-RPCH/AKH	pQLNFSPGWa		
SIFamide		Ixod, daphnia, decapods and insects	
Cha-SIFamide	GYRKPPFNGSIFa		
Short Neuropeptide F		Ixod, daphnia, decapods and insects	
Cha-sNPF1	GGPPSMRLRFa		
Cha-sNPY	GNIRSWQQVSQRSEPSLRLRYa		
Cha-sNPF2	DRTPALRLRFa		
Cha-sNPF3	TSELEQEEPFGDTDFLRQDRGAPALRLRFa		
Sulfakinin family		Peneids, lobster, insects	
Cha-SK1	pQFDEYGHMRFa		
Cha-SK2	AGGDYDDYGHLRFa		
Tachykinin Related Pept.		Daphnia, decapods and insects,	-/IPR013206
Cha-TKRP	APSGFLGMRa		

a = amide ; amphipod = *Talitrus saltator* ; cirriped = *Amphibalanus amphitrite* ; daphnia = *Daphnia pulex* ; decapods = identified in more than two species of decapods ; insects = identified in more than two species of hexapods ; isopod = *Armadillidium vulgare* ; Ixod = *Ixodus scapularis* ; lobster = *Homarus americanus* ; krill = *Euphausia superba* and/or *E. crystallorophias*

ACCEPTED MANUSCRIPT

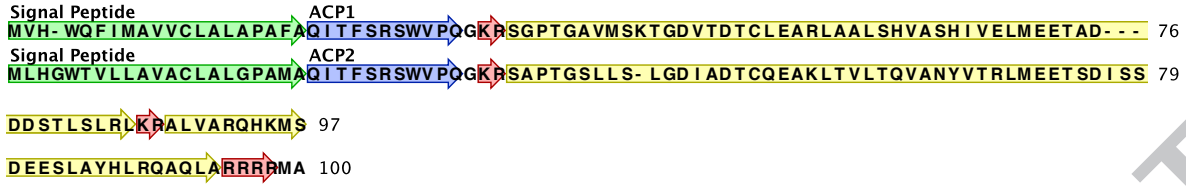
1- Illumina sequencing was used to produce a transcriptome of *Chorismus antarcticus*.
2- Analysis of the assembly produced 55 pre-pro-peptides coding for 111 neuropeptides
3- A new member of the CHH family blasting with ITP peptides was characterized
4- This new group of peptides would integrate with the set of type 2 CHH peptides

ACCEPTED MANUSCRIPT

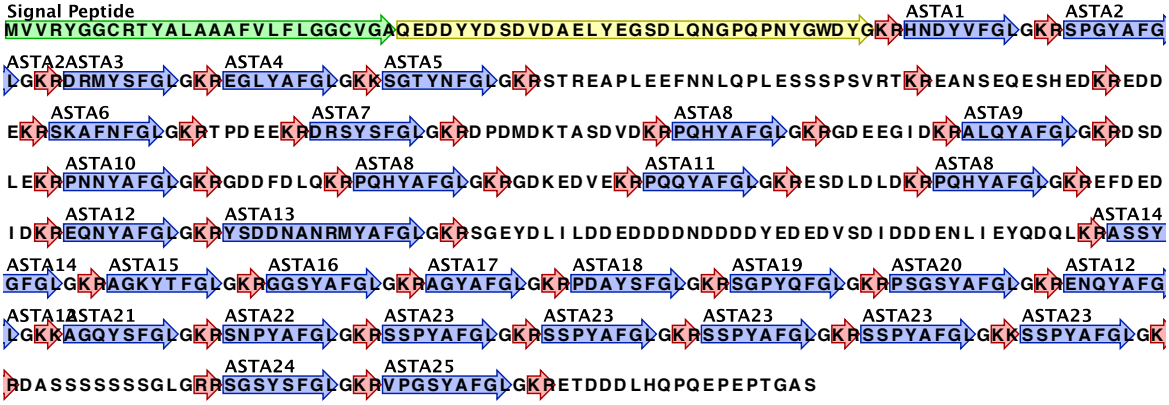
ACCEPTED MANUSCRIPT

Figure 1

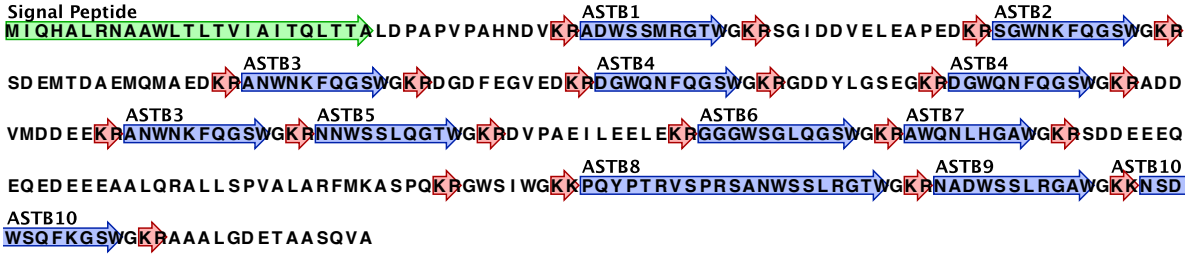
A



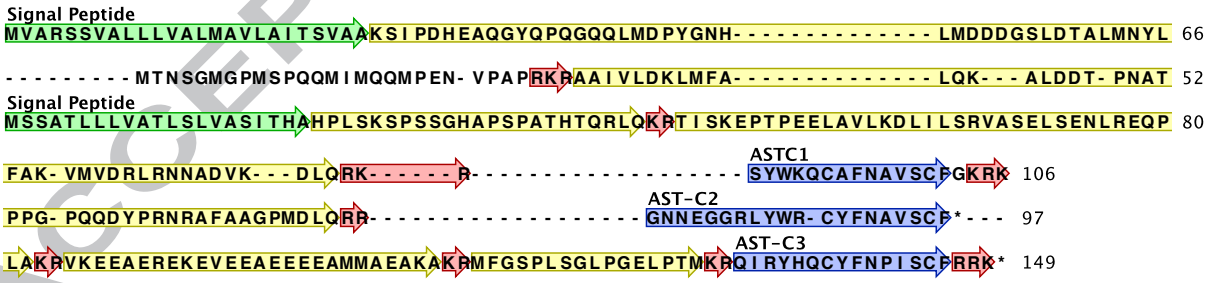
B



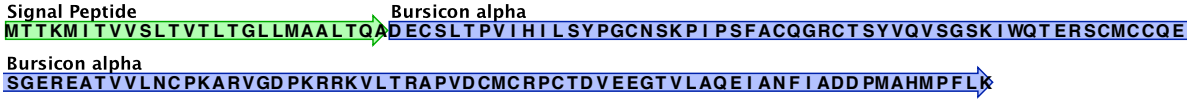
C



D

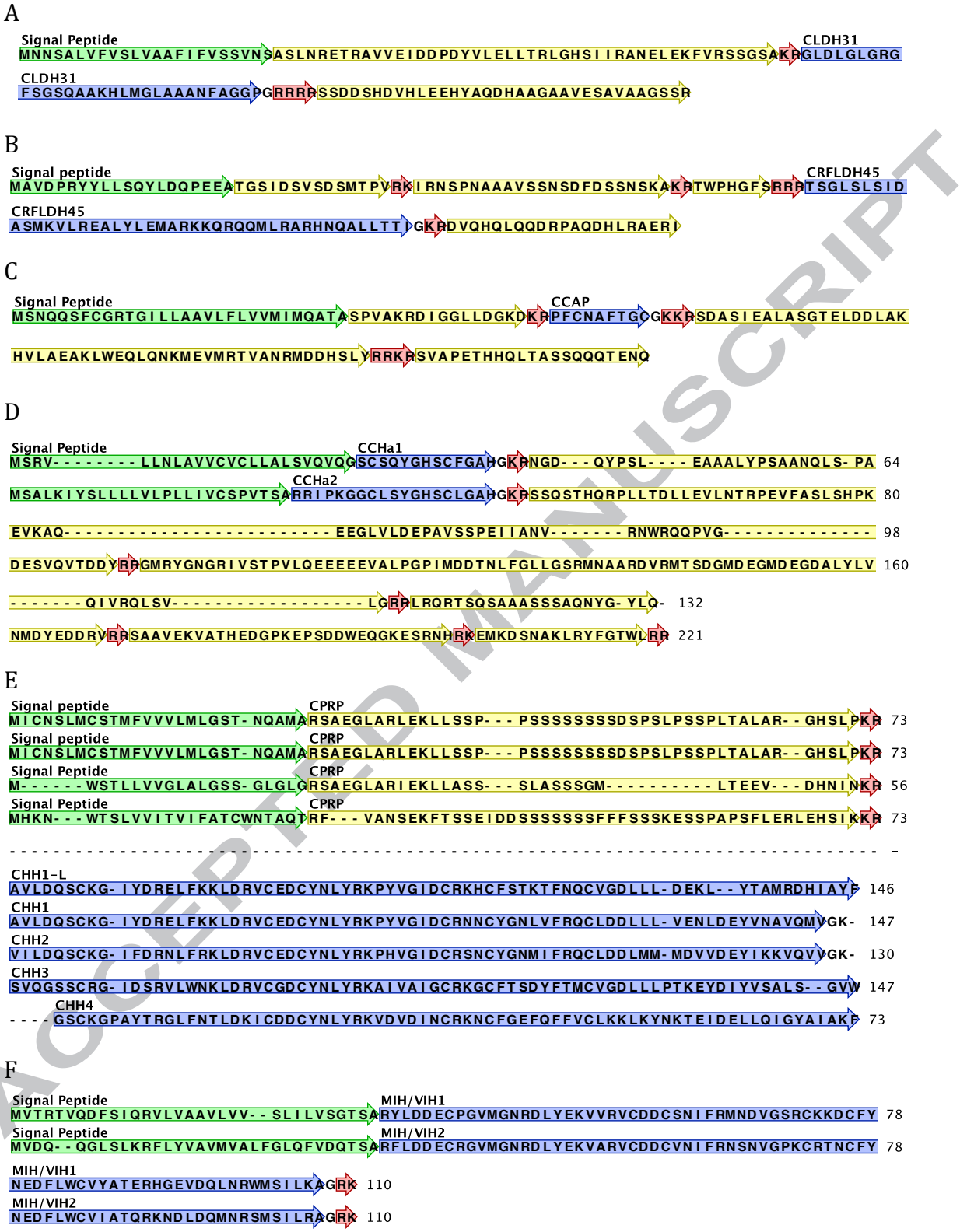


E



ACCEPTED MANUSCRIPT

Figure 2



ACCEPTED MANUSCRIPT

Figure 3

A

ACCEPTED MANUSCRIPT

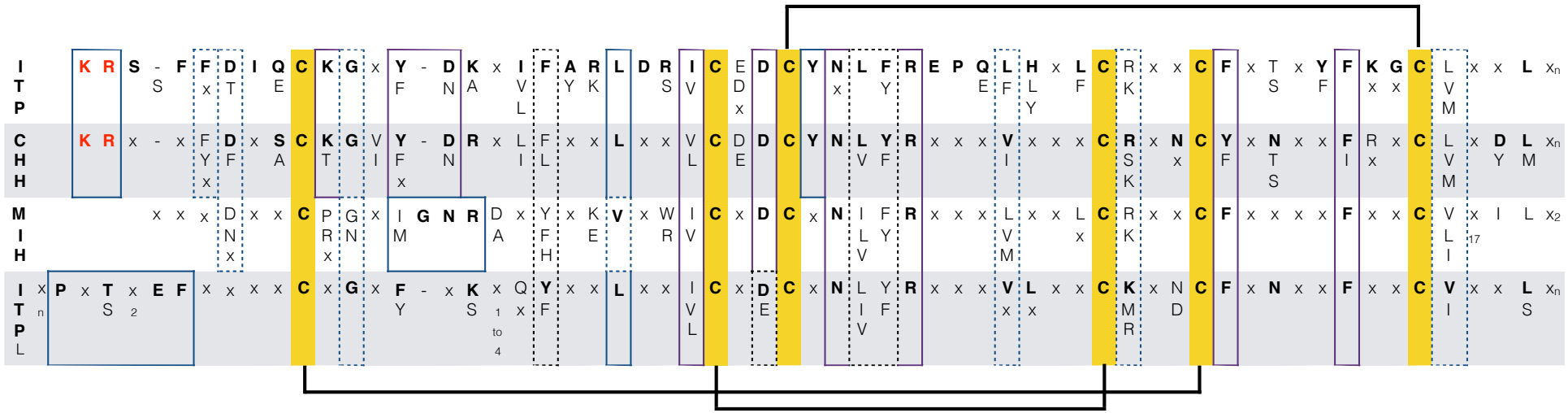
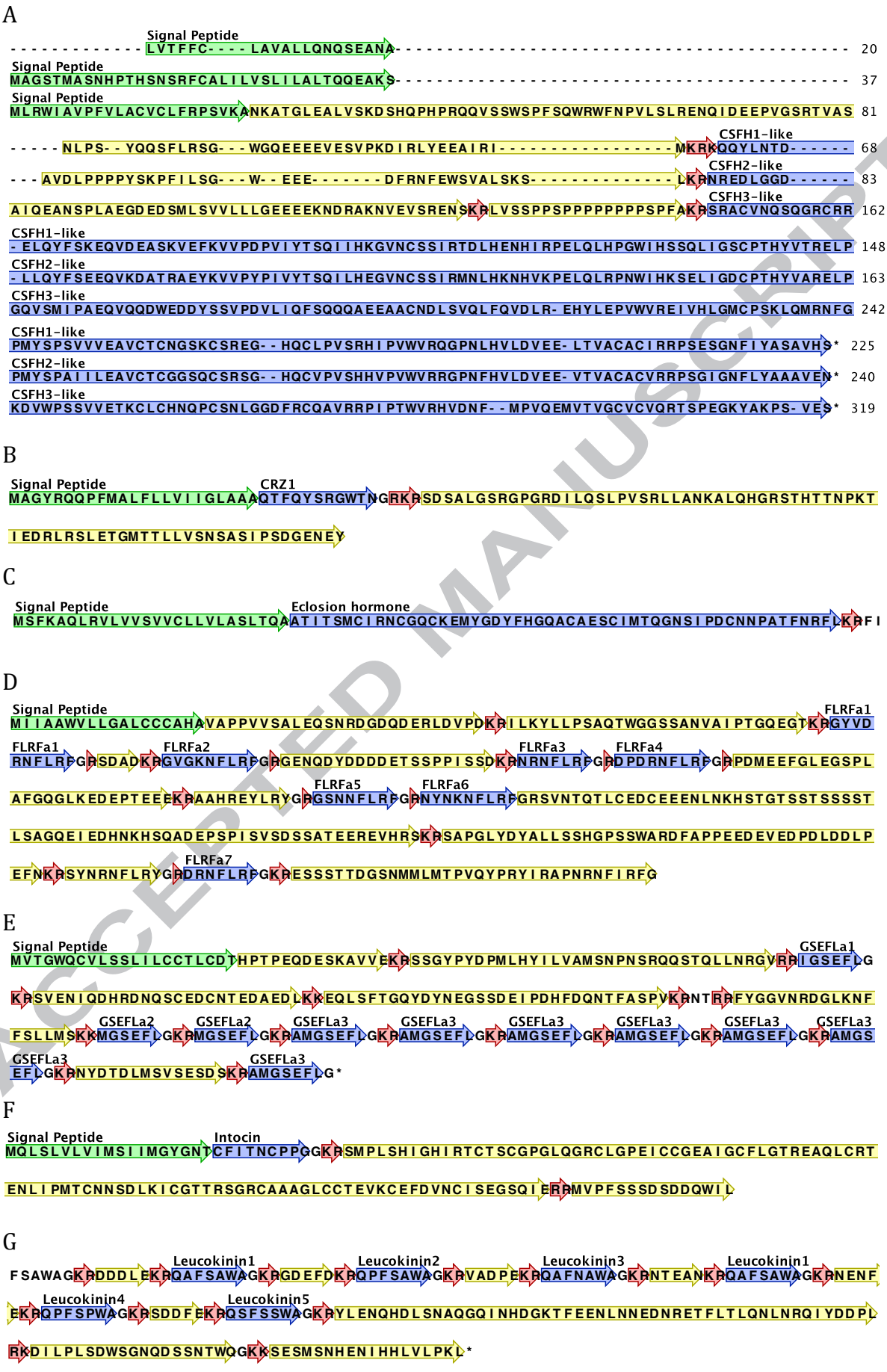


Figure 4

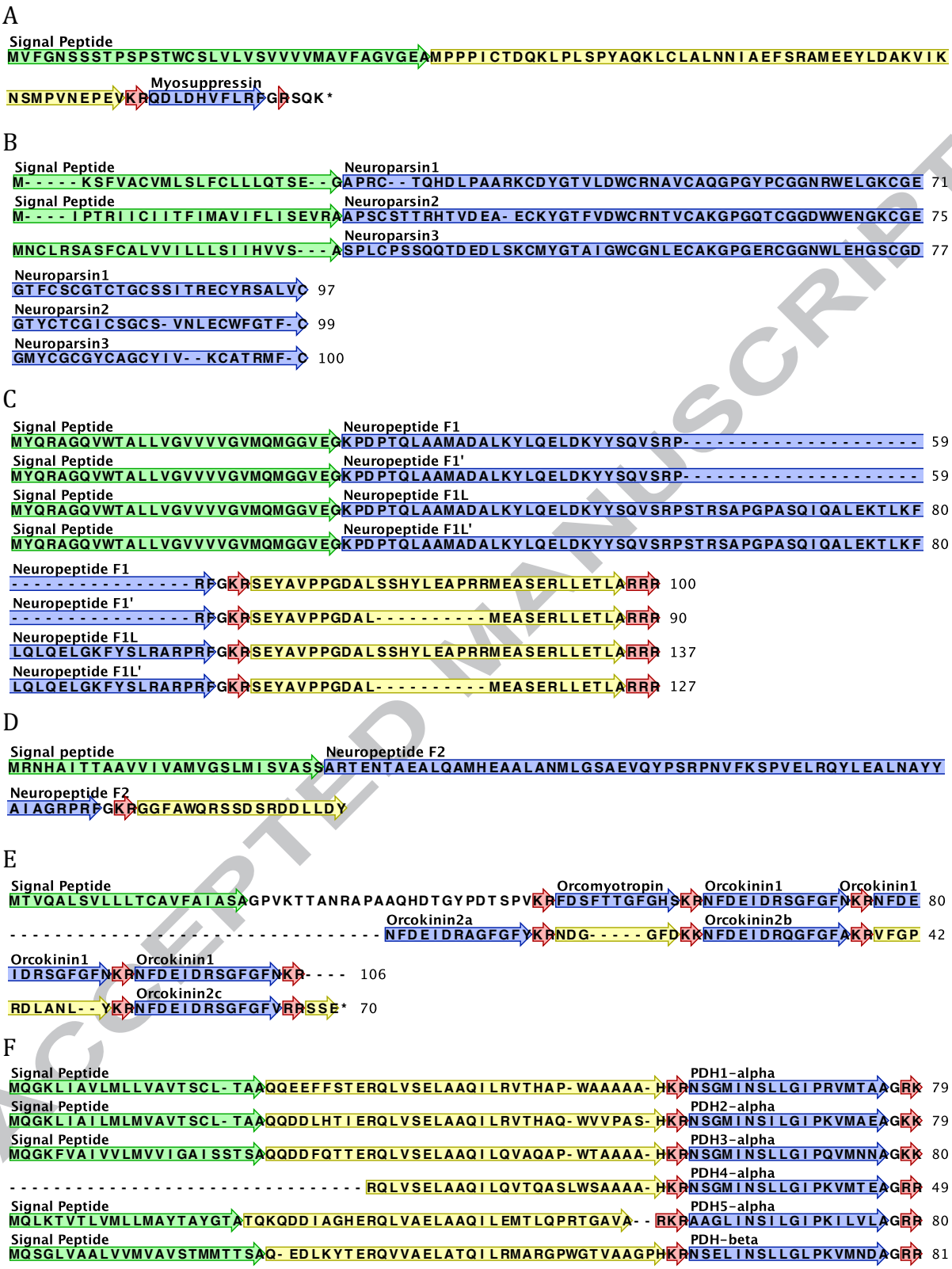
ACCEPTED MANUSCRIPT

Figure 5



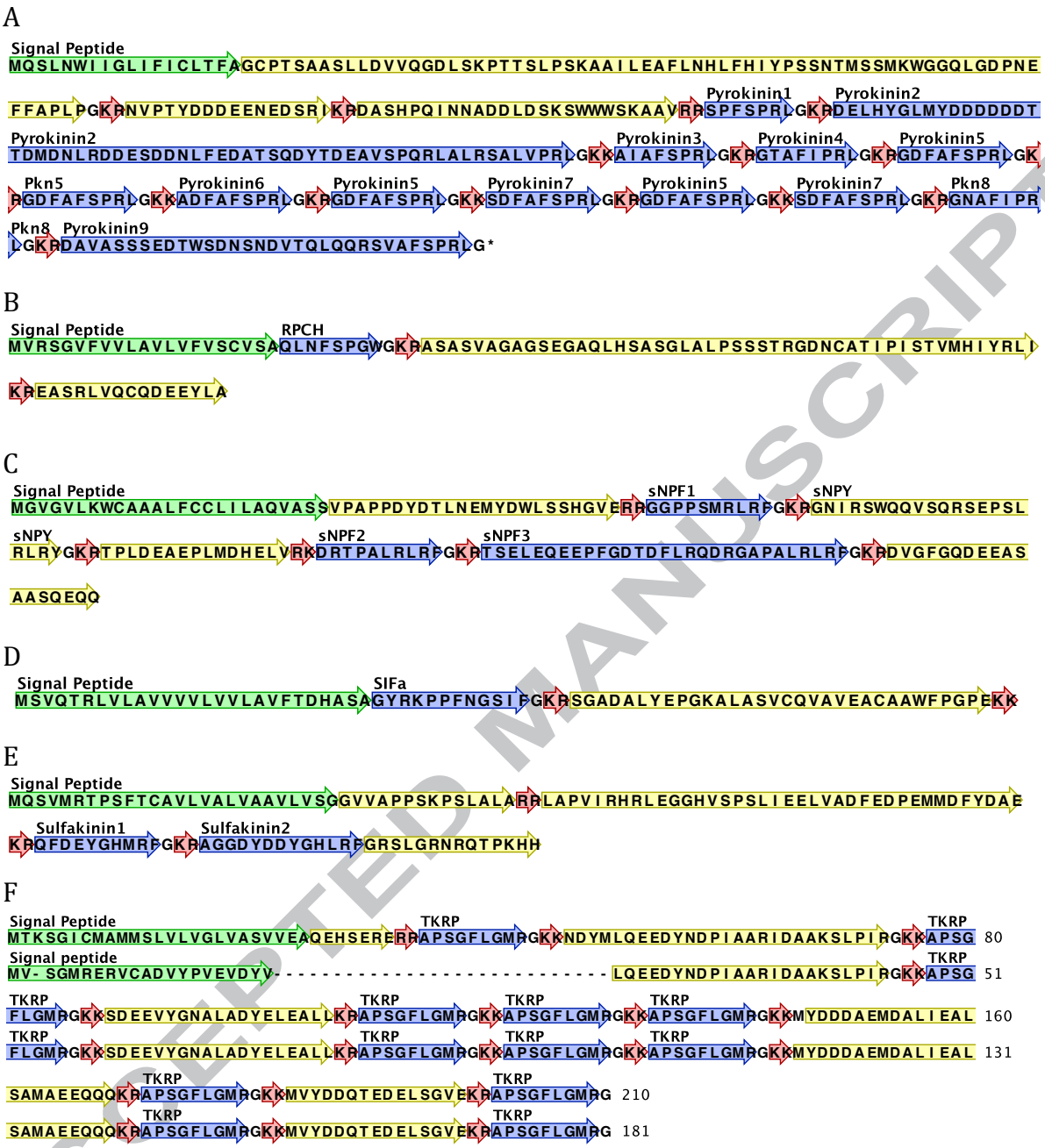
ACCEPTED MANUSCRIPT

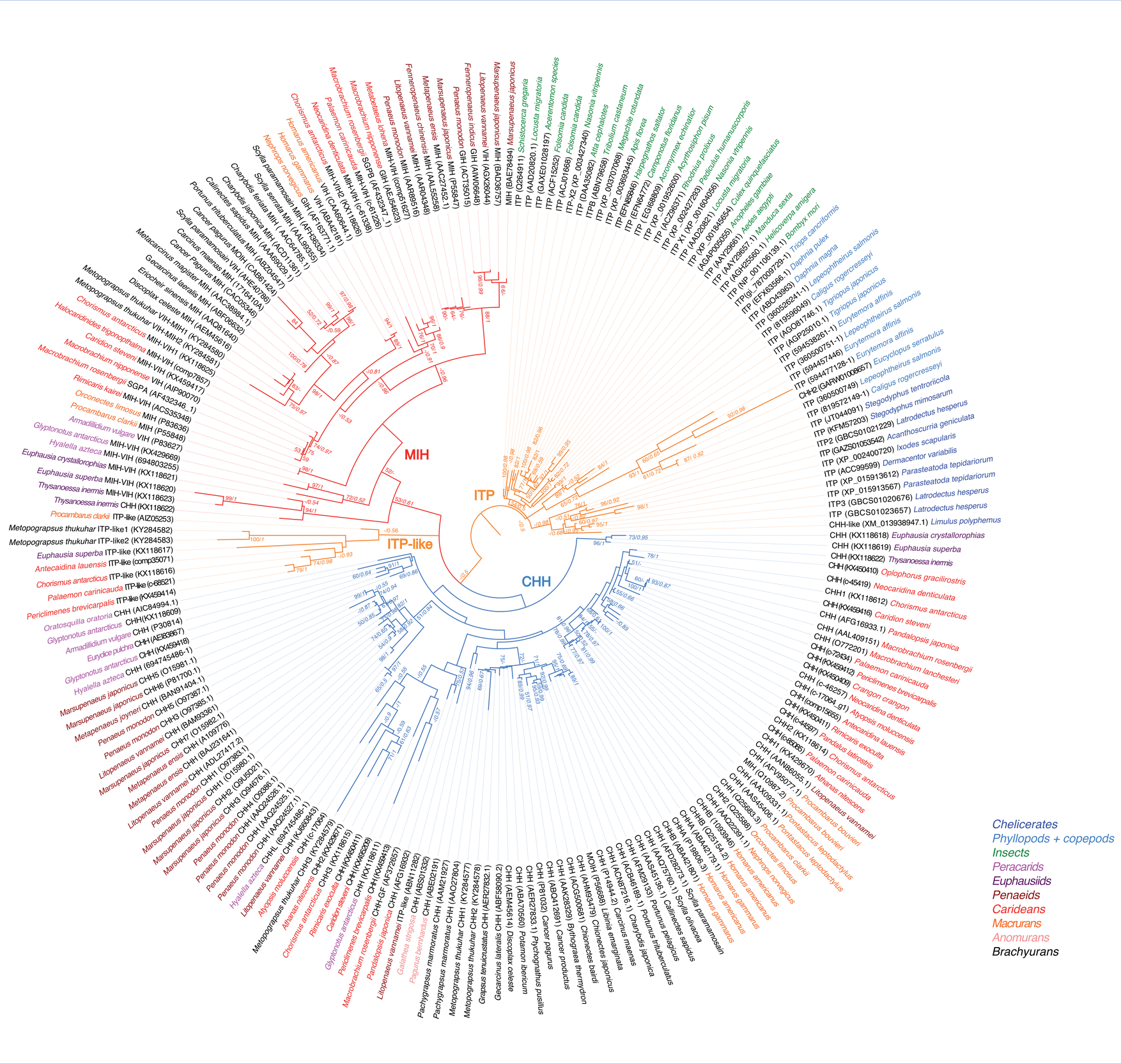
Figure 6

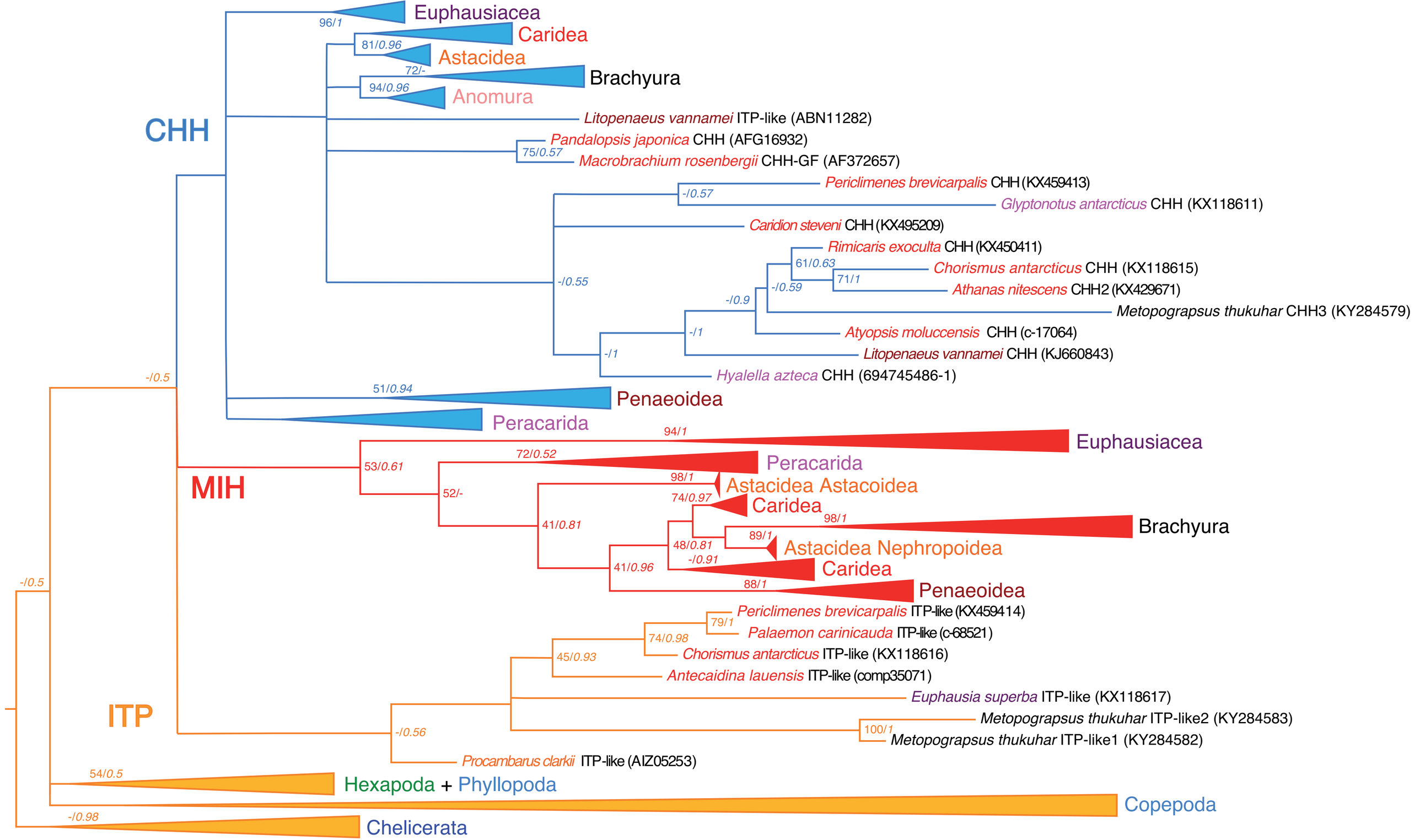


ACCEPTED MANUSCRIPT

Figure 7




MIH
ITP
ITP-like
CHH
Chelicerates
Phyllopods + copepods
Insects
Peracarids
Euphausiids
Penaeids
Carideans
Macrurans
Anomurans
Brachyurans


CHH
Euphausiacea
96/1
Caridea
81/0.96
Astacidea
72/-
Brachyura
94/0.96
Anomura
Litopenaeus vannamei ITP-like (ABN11282)
Pandalopsis japonica CHH (AFG16932)
75/0.57
Macrobrachium rosenbergii CHH-GF (AF372657)
Periclimenes brevicarpalis CHH (KX459413)
-/0.57
Glyptonotus antarcticus CHH (KX118611)
Caridion steveni CHH (KX495209)
Rimicaris exoculta CHH (KX450411)
61/0.63
Chorismus antarcticus CHH (KX118615)
71/1
-/0.59
Athanas nitescens CHH2 (KX429671)
-/0.55
-/0.9
Metopograpsus thukuhar CHH3 (KY284579)
-/1
Atyopsis moluccensis CHH (c-17064)
-/1
Litopenaeus vannamei CHH (KJ660843)
Hyalella azteca CHH (694745486-1)
-/0.5
51/0.94
Penaeoidea
Peracarida
94/1
Euphausiacea
72/0.52
Peracarida
53/0.61
MIH
98/1
Astacidea Astacoidea
52/-
74/0.97
Caridea
98/1
Brachyura
41/0.81
89/1
48/0.81
Astacidea Nephropoidea
-/0.91
Caridea
41/0.96
88/1
Penaeoidea
-/0.5
Periclimenes brevicarpalis ITP-like (KX459414)
79/1
Palaemon carinicauda ITP-like (c-68521)
74/0.98
Chorismus antarcticus ITP-like (KX118616)
45/0.93
Antecaidina lauensis ITP-like (comp35071)
Euphausia superba ITP-like (KX118617)
ITP
Metopograpsus thukuhar ITP-like2 (KY284583)
100/1
-/0.56
Metopograpsus thukuhar ITP-like1 (KY284582)
Procambarus clarkii ITP-like (AIZ05253)
54/0.5
Hexapoda + Phyllopoda
Copepoda
-/0.98
Chelicerata