General and Comparative Endocrinology 282 (2019) 113217

Contents lists available at ScienceDirect

General and Comparative Endocrinology

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

What can transcriptomics reveal about the phylogenetic/structural conservation, tissue localization, and possible functions of CNMamide peptides in decapod crustaceans?

Andrew E. Christie[a,*], J. Joe Hull[b]

[a] Békésy Laboratory of Neurobiology, Pacific Biosciences Research Center, School of Ocean and Earth Science and Technology, University of Hawaii at Manoa, 1993 East-West Road, Honolulu, HI 96822, USA

[b] Pest Management and Biocontrol Research Unit, US Arid Land Agricultural Research Center, USDA Agricultural Research Services, Maricopa, AZ 85138, USA

ARTICLE INFO

Keywords:
In silico transcriptome mining
Nervous system
Ovary
Testis
Peptide precursors
Peptide receptors

ABSTRACT

Over the past several years, *in silico* analyses of arthropod genomes/transcriptomes have led to the identification of several previously unknown peptide families. The CNMamides are one such peptide group, having been discovered via computational analyses of the fruit fly, *Drosophila melanogaster*, genome; both a CNMamide precursor and receptor were identified. Recently, a CNMamide family member, VMCHFKICNLamide (disulfide bridging between the cysteine residues), was predicted via *in silico* mining of a crayfish, *Procambarus clarkii*, transcriptome, suggesting the presence of this peptide group in members of the Decapoda. Here, using publically accessible transcriptomic data, the phylogenetic/structural conservation, tissue localization, and possible functions of the CNMamide family in decapods were explored. Evidence for CNMamide precursors was found for members of each decapod infraorder for which significant sequence data are available, suggesting a ubiquitous conservation of the CNMamide family in the Decapoda. For the Penaeoidea, Caridea, Astacidea and Achelata, the isoform of CNMamide originally identified from *P. clarkii* appears to be ubiquitously conserved; in members of the Brachyura, VMCHFKICNMamide (disulfide bridging between the cysteine residues) is the native isoform. Interestingly, the decapod CNMamide gene appears to also have a splice variant in which the carboxy-terminal portion of the preprohormone containing the CNMamide peptide is replaced by one containing a different disulfide bridged peptide that is structurally unrelated to it; this second peptide shows considerable conservation within, but variation among, decapod infraorders. A highly conserved putative CNMamide receptor was identified from members of the Penaeoidea, Astacidea and Brachyura. Phylogenetic analyses support the annotation of the decapod receptor as a true member of the CNMamide receptor family. The presence of precursor and receptor transcripts in both nervous system- and reproductive tissue-specific transcriptomes suggests CNMamides serve as modulators of decapod neural and reproductive control systems.

1. Introduction

Improvements in technology for high-throughput nucleotide sequencing, as well as for assembling the resulting data, has led to the generation of genomes and transcriptomes for a growing number of species. Not surprisingly, over the past several years, much work has focused on generating genomes/transcriptomes for members of the Arthropoda, as species from this phylum are important biomedical/ecological/ecotoxicological models, vectors of human disease, agricultural pests, and targets for agriculture/aquaculture (*e.g.*, Adams et al., 2000; Charrier et al., 2018; Northcutt et al., 2016; Poynton et al., 2018; Tassone et al., 2016; Zhang et al., 2019). These genomic/transcriptomic datasets have been used for a variety of purposes, one of which is the identification of peptide paracrines/hormones and their cognate receptors (*e.g.*, Bao et al., 2015, 2018; Buckley et al., 2016; Christie, 2014; Dircksen et al., 2011; Hummon et al., 2006; Veenstra et al., 2012; Ventura et al., 2014). In addition to expanding the taxa for which previously described peptidergic signaling systems likely exist, *in silico* transcriptome mining has also allowed for the identification of previously unknown peptide groups, and in some cases, their receptors (*e.g.*, Christie, 2014; Dircksen et al., 2011; Jung et al., 2014; Veenstra et al., 2012).

One peptide family recently identified via *in silico* genome mining is the CNMamides, named for the structure of their carboxyl (C)-terminus

* Corresponding author.
E-mail address: crabman@pbrc.hawaii.edu (A.E. Christie).

https://doi.org/10.1016/j.ygcen.2019.113217
Received 28 April 2019; Received in revised form 27 June 2019; Accepted 4 July 2019
Available online 05 July 2019

(Jung et al., 2014). Originally described from the fruit fly, *Drosophila melanogaster*, *in silico* analyses conducted on other arthropod datasets suggested the CNMamides to be broadly, though not ubiquitously, conserved in the Arthropoda (Jung et al., 2014). While a single CNMamide receptor was identified in *D. melanogaster*, some species appear to possess two CNMamide receptors (Jung et al., 2014). Interestingly, in some species that lack a gene for the CNMamide precursor, a putative CNMamide receptor is present (Jung et al., 2014). Immunohistochemistry revealed the expression of CNMamide peptides in neurons in the brain and ventral nerve cord of *D. melanogaster* (Jung et al., 2014), which suggests that one role played by CNMamide is as a locally-released and/or circulating neuromodulator, though as of yet, there is no direct demonstration of this proposed function. In fact, the only demonstration of CNMamide function currently extant from any species is that knockdown of the CNMamide precursor gene in *D. melanogaster* increases the delay to sperm ejection by females following copulation (Lee et al., 2015).

While one crustacean, the cladoceran *Daphnia pulex*, was among the species from which CNMamide signaling systems were initially identified (Jung et al., 2014), very little is currently known about CNMamide in the Crustacea. In fact, the only reports currently extant are the prediction of a putative partial CNMamide precursor from a crayfish, *Procambarus clarkii*, eyestalk transcriptome (Veenstra, 2015) and the prediction of a putative CNMamide precursor and receptor from the crab, *Carcinus maenas* (Oliphant et al., 2018). To help further our understanding of CNMamide signaling systems in the Crustacea, and specifically members of the Decapoda, we have used transcriptomic analyses of publicly accessible decapod datasets to investigate four questions: (1) To what extent is the presence of the CNMamide family conserved in the order?; (2) To what extent is there structural conservation of the native decapod CNMamide isoform(s) and receptor(s)?; (3) Can any assessment be made about possible tissue localization of CNMamide peptides/receptors in members of the taxa; and (4) Can any physiological/behavioral functions be proposed for decapod CNMamides based on their tissue localization(s)?.

2. Materials and methods

2.1. In silico transcriptome mining and peptide/receptor structural prediction

2.1.1. In silico transcriptome mining

Searches of publicly accessible decapod transcriptomic datasets for putative CNMamide precursor- and receptor-encoding sequences were conducted on or before April 12, 2019 using a well-vetted protocol (*e.g.*, Christie et al., 2015, Christie et al., 2018; Christie and Yu, 2019). Briefly, the database of the online program tblastn (National Center for Biotechnology Information, Bethesda, MD; http://blast.ncbi.nlm.nih.gov/Blast.cgi) was set to "Transcriptome Shotgun Assembly (TSA)" and restricted to data from specific decapod taxa for which sequences have been publicly deposited, *i.e.*, "Penaeoidea (taxid:111520)", "Caridea (taxid:6694)", "Astacidea (taxid:6712)", "Achelata (taxid:6730)", and "Brachyura (taxid:6752)". For decapod CNMamide precursor searches, a *P. clarkii* partial CNMamide preprohormone deduced from an eyestalk-specific transcript (Accession No. GARH01035775; Manfrin et al., 2015) was used as the query sequence; the deduced protein was identified previously as a putative CNMamide precursor (Veenstra, 2015). For decapod CNMamide receptor searches the sequences of the *D. melanogaster* (Accession No. AAF50229; Adams et al., 2000) and *Daphnia magna* (Accession No. KZS12306; unpublished direct GenBank submission) CNMamide receptors were used as the initial query proteins.

2.1.2. Peptide structural prediction

The putative mature structures of decapod CNMamide isoforms were predicted using a well-established workflow (*e.g.* Christie and Pascual, 2016; Christie et al., 2015, Christie et al., 2018; Christie and Yu, 2019). Specifically, the hits returned by a given BLAST search were translated using the ExPASy Translate tool (http://web.expasy.org/translate/) and assessed for completeness. Precursor proteins listed as "full-length" exhibit a functional signal sequence (including a "start" methionine) and are flanked on their C-terminus by a stop codon. Proteins listed as "partial" lack a start methionine (referred to as C-terminal partial proteins), a stop codon (referred to as amino [N]-terminal partial proteins), or both of these features (referred to as internal fragment proteins). Next, each full-length or N-terminal partial precursor protein was assessed for the presence of a signal peptide using the online program SignalP 4.1 (http://www.cbs.dtu.dk/services/SignalP/; Petersen et al., 2011); the D-cutoff values of SignalP were set to "Sensitive" to better match the sensitivity of version 3.0 of this freeware program. Prohormone cleavage sites were identified based on information presented in Veenstra (2000). When present, the sulfation state of tyrosine residues was predicted using the online program Sulfinator (http://www.expasy.org/tools/sulfinator/; Monigatti et al., 2002). Disulfide bonding between cysteine residues was predicted by homology to known peptide isoforms or by using the online program DiANNA (http://clavius.bc.edu/~clotelab/DiANNA/; Ferrè and Clote, 2005). Other post-translational modifications, *i.e.*, cyclization of N-terminal glutamine/glutamic acid residues and C-terminal amidation at glycine residues, were predicted by homology to known arthropod peptides. All protein/peptide alignments were done using the online program MAFFT version 7 (http://mafft.cbrc.jp/alignment/software/; Katoh and Standley, 2013). To determine amino acid conservation between selected precursor proteins or peptides, the sequences in question were aligned using MAFFT, and amino acid identity/similarity subsequently determined using the alignment output. Specifically, percent identity was calculated as the number of identical amino acids divided by the total number of residues in the longest sequence ($\times 100$), while amino acid similarity was calculated as the number of identical and similar amino acids divided by the total number of residues in the longest sequence ($\times 100$).

2.1.3. Peptide receptor structural prediction and vetting

A workflow developed to help provide provisional annotation to a variety of protein types, including peptide receptors, was used to vet the annotation of the deduced putative decapod CNMamide receptors (*e.g.*, Christie et al., 2015, Christie et al., 2018; Christie and Yu, 2019; Dickinson et al., 2019). First, nucleotide sequences were translated using the Translate tool of ExPASy and assessed for completeness (see Section 2.1.2). Next, to confirm that the *D. melanogaster* CNMamide receptor is the most similar protein to each of the putative decapod CNMamide receptors, each decapod sequence was employed as the input query in a BLAST search of the annotated *Drosophila* protein dataset present in FlyBase version FB2019_01 (http://flybase.org/blast/index.html; Thurmond et al., 2019). Finally, protein structural motifs were predicted for each of the putative decapod CNMamide receptors using the online program Pfam 32.0 (http://pfam.xfam.org/; El-Gebali et al., 2019). The same procedure used to determine amino acid conservation between peptide precursor/peptide isoforms was used to calculate identity/similarity scores for selected receptor proteins (see Section 2.1.2).

2.2. Assessment of phylogenetic relationships among receptor proteins

The phylogenetic relationships of the putative decapod CNMamide receptors with a diverse set of *D. melanogaster* receptors were inferred from a multiple sequence alignment constructed using default MUSCLE (Edgar, 2004) settings in Geneious v10.1.3. Evolutionary analyses were conducted in MEGA X (Kumar et al., 2018) using the maximum likelihood method based on the Le and Gascuel model (Le and Gascuel, 2008). Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix

A. Alignment of Type-1 CNMamide precursors from four different decapod infraorders

```
Penmo-prepro-CNMa-1 MFARGRHGCG----------SLPRLLLALLSLLMAARVQPSPFAEYEG-LP----F--A
Macto-prepro-CNMa-1 MVFRGRQGSG----------SLTWFLLVVVSLVVVVGVHPSPYPAYEDHLP----Y--Q
Homam-prepro-CNMa-1 MVSRTQQQQQQQGGGGGNNGWSLRWFVVVVVSLVLVARVQ--PYPGYGERLRSVRRY---
Carma-prepro-CNMa-1 MVSRGQQSSV----------SLTWLLVGAVSLVLVARVQ--PYPGYVDRLPSLRRYPQT
                    *. * ::              **  :::  :**::.. *:  *:. *   *     :
Penmo-prepro-CNMa-1 SSFESHFVPKAPSPAAPPSAASSSSSSSSSSSLPSSVARARSRSFPTRHDADARMNVVYE
Macto-prepro-CNMa-1 ISQED-----APGDVSPPMTYN-----------------------------------
Homam-prepro-CNMa-1 QAQQR-----ALHDDT--HTLD-----------------------------------
Carma-prepro-CNMa-1 QQQQQ-----ALGEGEPFAGAN-----------------------------------
                       :      *          .
Penmo-prepro-CNMa-1 ASDYDEQAAAAAAAVEAAVEVLQEYDFPKAKSLEAAGSSNYSPSQKASLDSTKNLQQILR
Macto-prepro-CNMa-1 --------------PDPHLSFPQELEFGSSNPERFPQIF-YDRRS----DSNRNLQQILR
Homam-prepro-CNMa-1 -NDYLQ-------GDEGELDYSQFYDYDSVLPRRDPDGYHFDQPE----DSNKNLQQILR
Carma-prepro-CNMa-1 -DDYVE--------DDDDLGYSKKYAYNTLMAGREPDIFPYDQAE----ASNKNLQQILR
                                 :  :   :   : .  . . .    :.  .     *.:*******
Penmo-prepro-CNMa-1 SFPLVVGGHVPPPSSSSPYDV--WRSFQEGVEGDHRPLSGGSGGGAGAGGV-PQARLPEL
Macto-prepro-CNMa-1 SFPLVVGGPSP----SDRVLPLSWSSLEEATLAQQ-----GRGTGVSSNLS-KGATLTQL
Homam-prepro-CNMa-1 SFRPVVGGTSP----SDRVMP--WGELDQEAAIDD-PTQGG-GREADAGGLPSRTQLTQL
Carma-prepro-CNMa-1 SFRPVVGGTYP----SDRVLP--WDDLHELGLKD--------GRSAEEQLSAPDTSLTQL
                    **  **** *    *.      * .:.:    :        *  .        : *.:*
Penmo-prepro-CNMa-1 PFGEPRPKRVMCHFKICNLGRRRRARQSLPLQGWLS
Macto-prepro-CNMa-1 TLEEPRPKRVMCHFKICNLGRRRRARMSSPLQGWLS
Homam-prepro-CNMa-1 TLSEPKQKRVMCHFKICNLGRRRRARQSSPLQGWLS
Carma-prepro-CNMa-1 SFAKHRQKRVMCHFKICNMGRRRRARHSNPLQGWLS
                    .: : : ***********:******* * *******
```

B. Alignment of *Portunus sanguinolentus* Type-1 and Type-2 CNMamide precursors

```
Porsa-prepro-CNMa-1 MVSRGQQGSGSLTWLLVGAVLLVLVGRVQPYPGYVERLPSLRRYPQTQQQQGPAESEPFA
Porsa-prepro-CNMa-2 MVSRGQQGSGSLTWLLVGAVLLVLVGRVQPYPGYVERLPSLRRYPQTQQQQGPAESEPFA
                    ************************************************************
Porsa-prepro-CNMa-1 GTNDDYVEDDDDELGYSKKYAYDTLMAGREPDIFPYDQAEASNKNLQQILRSFRPVVGGT
Porsa-prepro-CNMa-2 GTNDDYVEDDDDELGYSKKYAYDTLMAGREPDIFPYDQAEASNKNLQQILRSFRPVVGGT
                    ************************************************************
Porsa-prepro-CNMa-1 YPSDRVLPWDDLHELGLKGGRSAEDQLSAPDTSLTQLSFAKHRQKRVMCHFKICNMGRRR
Porsa-prepro-CNMa-2 YPSDRVLPWDDLHELGLKGGRSAEDQLSAPDTSLTQLSFAKHRQKR------------E
                    ******************************************               .
Porsa-prepro-CNMa-1 RARHSN---PLQGWLS
Porsa-prepro-CNMa-2 RKWHCGLWMPICPFSG
                    *  *..   *:  : .
```

Fig. 1. MAFFT alignment of selected full-length putative decapod CNMamide (CNMa) precursor proteins. (A) Alignment of Type-1 precursors from members of four different decapod infraorders. (B) Alignment of Type-1 and Type-2 *Portunus sanguinolentus* CNMamide gene splice variants. In each protein, the signal peptide is shown in gray, the CNMamide isoform in the Type-1 precursor is shown in red, the variant disulfide bridged peptide in the Type-2 precursor is shown in pink, and linker/precursor-related peptides in both precursor types are shown in blue. In the line immediately below each sequence grouping, the symbol "*" indicates amino acids that are identical in all four proteins, while "." and ":" denote amino acids that are similar in structure among all four sequences. Species abbreviations: Penmo, *Penaeus monodon*; Macto, *Macrobrachium tolmerum*; Homam, *Homarus americanus*; Carma, *Carcinus maenas*; Porsa, *Portunus sanguinolentus*. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

of pairwise distances estimated using a Jones-Taylor-Thornton model (Jones et al., 1992) and then selecting the topology with superior log likelihood value. A discrete gamma distribution was used to model evolutionary rate differences among sites (two categories [+G, parameter = 1.3245]). The analysis involved 38 amino acid sequences and all positions with less than 95% site coverage were eliminated such that fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position. The final dataset consisted of total of 280 positions. The tree shown reflects the topology with the highest log likelihood (−21213.82). Phylogenetic inferences made using neighbor joining (Saitou and Nei, 1987) and minimum evolution (Rzhetsky and Nei, 1992) approaches in MEGAX generated trees with similar topologies.

A more refined examination of the phylogenetic relationships among Type-1 and Type-2 CNMamide receptor (Jung et al., 2014) sequences from diverse arthropods was also conducted. As before, a multiple sequence alignment was constructed using default MUSCLE settings and phylogeny estimated using the maximum likelihood method in MEGAX. A discrete gamma distribution was used to model evolutionary rate differences among sites (two categories [+G, parameter = 1.2178]). The rate variation model allowed for some sites to be evolutionarily invariable ([+I], 12.72% sites). The analysis involved 22 amino acid sequences. As before, all positions with less than 95% site coverage were eliminated and the final dataset consisted of 217 positions. The tree shown reflects the topology with the highest log likelihood (−5911.06). Similar tree topologies were generated using neighbor joining and minimum evolution approaches.

3. Results and discussion

3.1. Transcriptome mining suggests the CNMamides are broadly conserved in decapod species

Prior to this study, the presence of members of the CNMamide family in decapods was limited to the crayfish, *P. clarkii* and the crab, *C. maenas*, where transcripts encoding a putative CNMamide precursors were identified via transcriptome mining (Oliphant et al., 2018; Veenstra, 2015). Using the C-terminal partial protein deduced from the *P. clarkii* transcript, the extant decapod TSA datasets in NCBI were searched for transcripts encoding putative homologs. Via these searches, evidence for the presence of CNMamide precursors was found for at least one member of each decapod infraorder for which significant sequence data has been publicly deposited (Supplemental Table 1). Specifically, putative precursor-encoding transcripts were identified from *Penaeus monodon* and *Litopenaeus vannamei*, two members of the Penaeoidea (penaeid shrimp), and from *Macrobrachium tolmerum*,

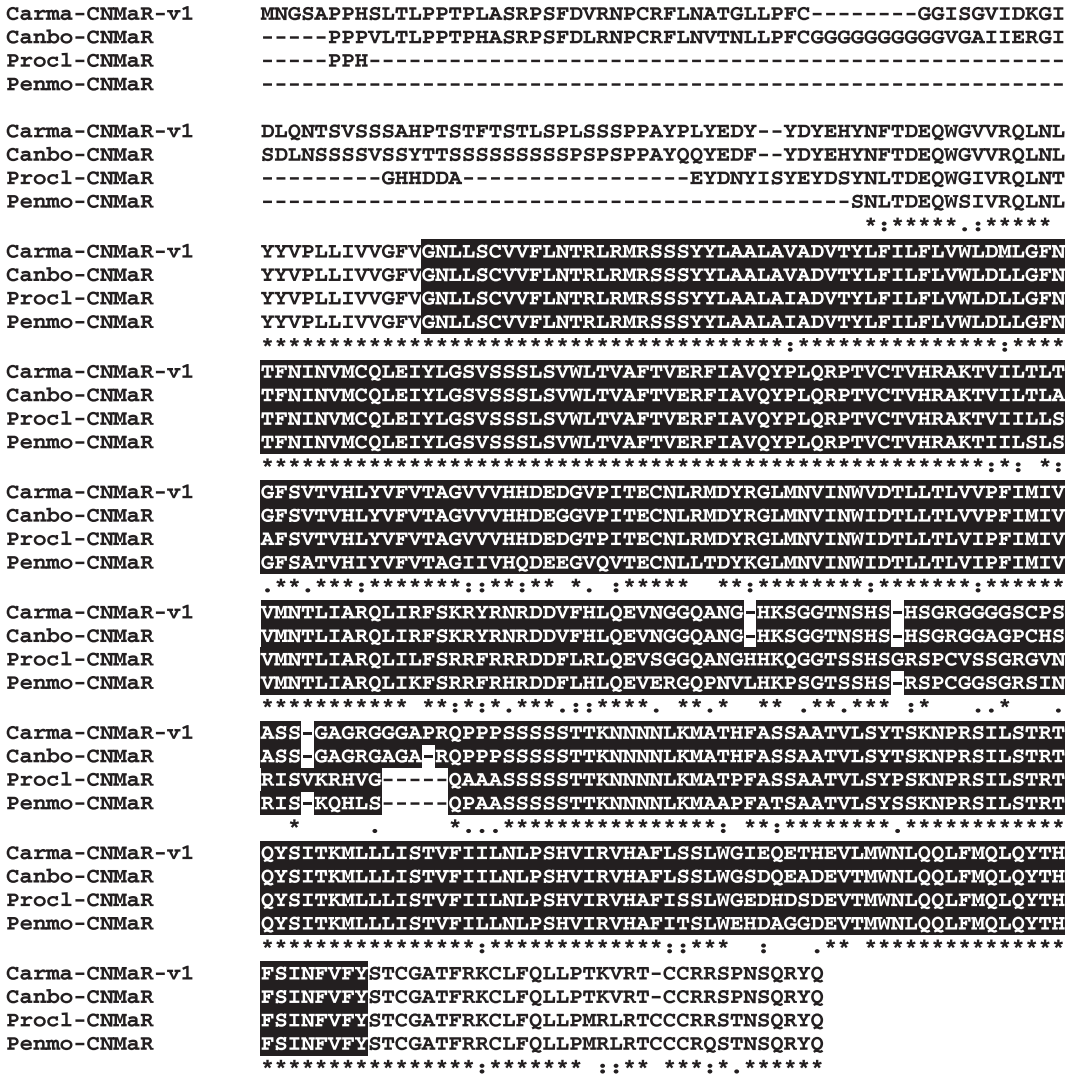

Fig. 2. MAFFT alignment of full-length/near full-length putative decapod CNMamide receptor (CNMaR) proteins. In the line immediately below each sequence grouping, the symbol "*" indicates amino acids that are identical in all four proteins, while "." and ":" denote amino acids that are similar in structure among all four sequences. Rhodopsin family seven-transmembrane receptor domains identified by Pfam in the putative decapod CNMamide receptors are highlighted in black. Species abbreviations: Carma, *Carcinus maenas*; Canbo, *Cancer borealis*; Procl, *Procambarus clarkii*; Penmo, *Penaeus monodon*.

Macrobrachium novaehollandiae, and *Macrobrachium australiense*, three members of the Caridea (caridean shrimp). In addition to *P. clarkii*, putative CNMamide-encoding transcripts were identified from *Homarus americanus* and *Cherax quadricarinatus*, two other members of the Astacidea (clawed lobsters and freshwater crayfish), and from *Jasus edwardsii*, a member of the Achelata (spiny and slipper lobsters). Finally, putative CNMamide-encoding transcripts were identified from *C. maenas*, *Cancer borealis*, *Portunus sanguinolentus*, *Eriocheir sinensis*, and *Scylla olivacea*, five members of the Brachyura (true crabs). Taken collectively, these data support the CNMamide family being broadly, if not ubiquitously, conserved in the Decapoda.

3.2. At least two isoforms of CNMamide appear to exist in the Decapoda

The putative mature structures of native decapod CNMamides were predicted from the precursor proteins deduced from the identified transcriptomic sequences (Fig. 1A and Supplemental Fig. 1). The peptide VMCHFKICNLamide (a disulfide bridge between the two cysteine residues) was the sole CNMamide isoform identified for members of the Penaeoidea, Caridea, Astacidea and Achelata, while VMCHFKI-CNMamide (a disulfide bridge between the two cysteine residues) was the sole CNMamide isoform identified for members of the Brachyura. The two peptides differ via a conserved substitution at position 10 of their sequences, a leucine in the former and methionine in the latter. Thus, there appear to be at least two CNMamide in the Decapoda, one broadly conserved, and the other, possibly infraorder-specific.

3.3. At least two splice variants of the CNMamide gene appear be present in decapod species

An unexpected finding from our *in silico* investigation was that the decapod CNMamide gene appears to give rise to two distinct precursors, one containing the CNMamide peptide (Type-1 precursors), and one in which the C-terminal portion of the preprohormone containing the CNMamide is replaced by one having a different disulfide bridged peptide that is structurally unrelated to CNMamide (Type-2 precursors). With the exception of this relatively short region of putative alternative splicing, the two variants are otherwise identical in all species for which both precursor types were identified (Fig. 1B and Supplemental Fig. 1). Evidence for the second putative splice variant was found for members of the Penaeoidea (*P. monodon*), Caridea (*M. tolmerum*), Astacidea (*C. quadricarinatus* and *Astacus astacus*), and the Brachyura (*C. maenas*, *P.*

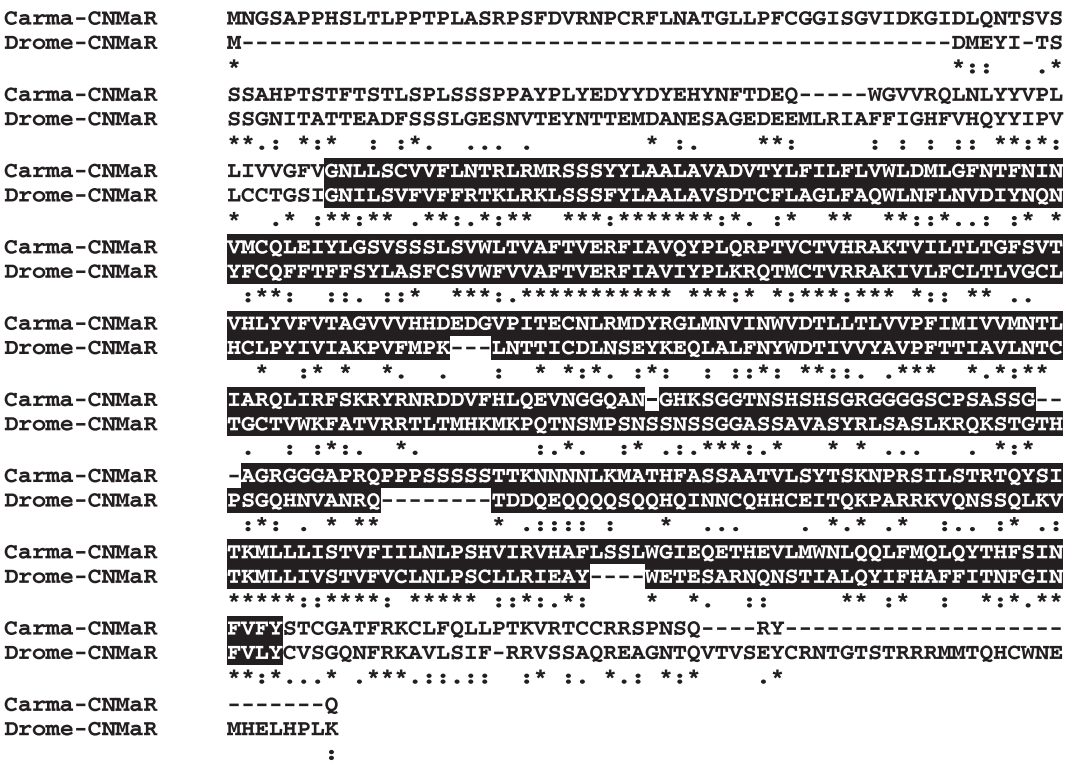

Fig. 3. MAFFT alignment of *Carcinus maenas* (Carma) and *Drosophila melanogaster* (Drome) CNMamide receptor (CNMaR) proteins. In the line immediately below each sequence grouping, the symbol "*" indicates amino acids that are identical in all four proteins, while "." and ":" denote amino acids that are similar in structure among all four sequences. Rhodopsin family seven-transmembrane receptor domains identified by Pfam in the CNMamide receptors are highlighted in black.

sanguinolentus, *S. olivacea*, and *Charybdis feriata*), suggesting it is a conserved phenomenon for the decapod CNMamide gene. Interestingly given the extreme level of conservation seen for the decapod CNMamide isoforms, the bridged peptide in the Type-2 splice variant, while showing considerable conservation within members of the same infraorder, is quite variable among infraorders, *i.e.*, PCILYIRICPFRSL in *P. monodon*, FCINWLKFCPF in *M. tolmerum*, pEASWPCILWVKFCPLamide and pEAIWPCVLWVKFCPLamide in *C. quadricarinatus* and *A. astacus*, respectively, and GRKWHCGLWMPICPFSamide in *C. maenas*, pERKWHCGLWMPICPFSamide in both *P. sanguinolentus* and *S. olivacea*, and ARKWHCGLWMPICPFSamide in *C. feriata* (disulfide bridging between the two cysteines in each peptide).

3.4. Identification of a putative highly conserved decapod CNMamide receptor

Using the sequences of the *D. melanogaster* CNMamide receptor (Accession No. AAF50229; Adams et al., 2000) and a putative CNMamide receptor from the cladoceran crustacean *D. magna* (Accession No. KZS12306; unpublished direct GenBank submission), the extant decapod TSA datasets were also searched for transcripts encoding putative decapod CNMamide receptor proteins. Transcripts encoding a highly conserved putative CNMamide receptor were found in members of the Penaeoidea, Astacidea and Brachyura (Supplemental Table 1). A full-length protein was deduced from a *C. maenas* transcript (Brachyura), likely the same receptor reported by Oliphant et al. (2018), with presumably near full-length C-terminal partial proteins deduced from *C. borealis* (Brachyura), *P. monodon* (Penaeoidea), and *P. clarkii* (Astacidea) transcripts (Fig. 2). Alignment of these sequences shows that the putative decapod CNMamide receptors are highly conserved in terms of their amino acid sequences (Fig. 2), *e.g.*, the putative *C. maenas* and *P. monodon* CNMamide receptors are 81% identical/93% similar in amino acid sequence over their region of overlap. Fragments of the putative CNMamide receptor were also identified from *L. vannamei* (Penaeoidea) and *E. sinensis* and *S. olivacea* (both members of the Brachyura) transcripts (Supplemental Fig. 1).

To increase confidence that the putative decapod proteins annotated as CNMamide receptors truly represent members of the CNMamide receptor family, the full-length *C. maenas* and near full-length *C. borealis*, *P. monodon*, and *P. clarkii* proteins were used as the query sequences in searches of the annotated *D. melanogaster* proteins curated in FlyBase. As expected, the CNMamide receptor was returned as the top hit for each search; the full-length *C. maenas* putative CNMamide receptor is 30% identical/59% similar in amino acid sequence to its *D. melanogaster* counterpart (Fig. 3). Assessment of the phylogenetic relationships among the putative decapod CNMamide receptors and a large set of *D. melanogaster* receptors (Supplemental Fig. 2) revealed the decapod sequences formed a receptor specific clade with significant bootstrap support (98%) with the *D. melanogaster* CNMamide receptor (Fig. 4). Assessment of the phylogenetic relationships among the putative decapod CNMamide receptors and previously identified arthropod Type-1 and Type-2 CNMamide receptors (Jung et al., 2014) (Supplemental Fig. 3) showed the decapod proteins clustered with significant bootstrap support (100%) with the Type-2 sequences (Fig. 5). Pfam analyses of the full-length/near full-length decapod receptors revealed each to contain a single rhodopsin family seven-transmembrane receptor domain, a domain also predicted for the *D. melanogaster* CNMamide receptor (Fig. 3).

3.5. Possible functions for CNMamides in decapod crustaceans

While most of the transcripts encoding putative decapod CNMamide precursors/receptors come from assemblies derived from mixed tissues, several were identified from tissue-specific transcriptomes (Supplemental Table 1), providing insight into possible functions the CNMamide signaling system may play in members of the Decapoda.

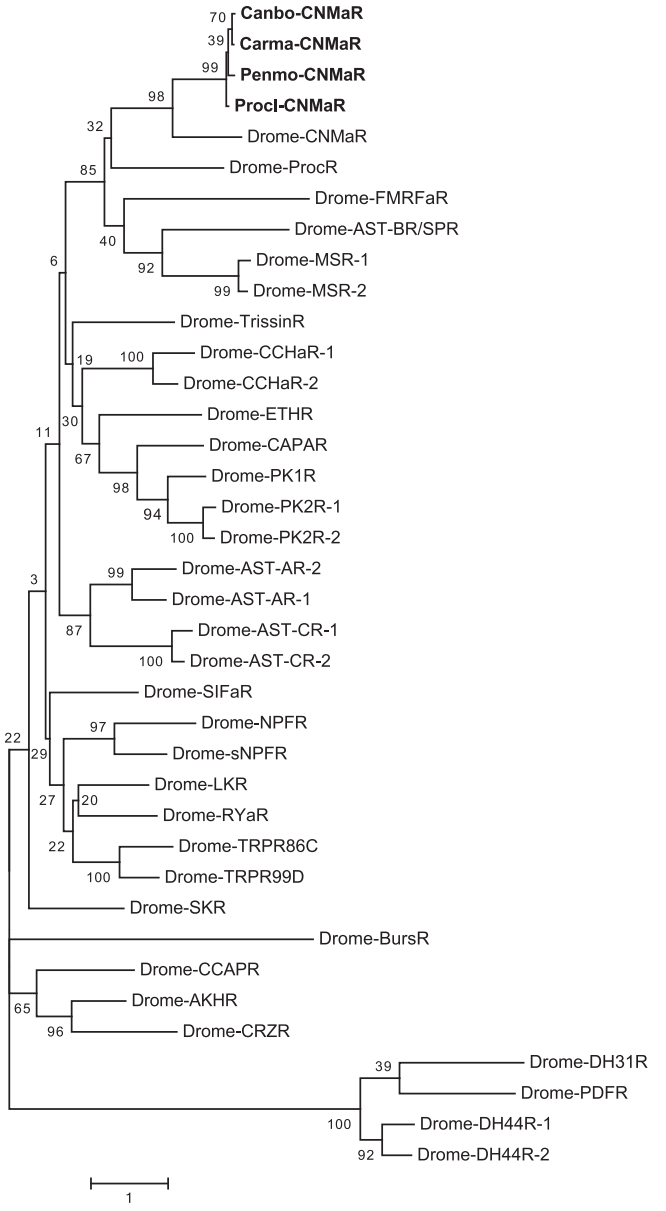

Fig. 4. Phylogenetic relationship of putative decapod CNMamide receptors with select *Drosophila melanogaster* receptors. Maximum likelihood tree depicting the inferred evolutionary history of the respective receptor sequences. The tree with the highest log likelihood is drawn to scale with branch lengths measured in the number of substitutions per site. The percentage of trees in which the associated taxa clustered together across 1000 replicates is shown next to the branches. Decapod sequences are shown in bold font. Receptor abbreviations: AKHR (adipokinetic hormone receptor); AST-AR (allatostatin A receptor); AST-BR/SPR (allatostatin B receptor/sex peptide receptor); AST-CR (allatostatin C receptor); BursR (bursicon receptor); CAPAR (Capa receptor); CCHaR (CCHamide receptor); CNMaR (CNMamide receptor); CRZR (corazonin receptor); CCAPR (crustacean cardioactive peptide receptor); DH31R (diuretic hormone 31 receptor); DH44R (diuretic hormone 44 receptor); ETHR (ecdysis-triggering hormone receptor); FMRFaR (FMRFamide receptor); LKR (leucokinin receptor); MSR (myosuppressin receptor); NPFR (neuropeptide F receptor); PDFR (pigment-dispersing factor receptor); ProcR (proctolin receptor); PKR (pyrokinin receptor); RYaR (RYamide receptor); SIFaR (SIFamide receptor); SKR (sulfakinin receptor); sNPFR (short NPF receptor); TRPR (tachykinin-related peptide receptor); TrissinR (trissin receptor). Species abbreviations: Canbo, *Cancer borealis*; Carma, *Carcinus maenas*; Drome, *Drosophila melanogaster*; Penmo, *Penaeus monodon*; Procl, *Procambarus clarkii*. Accession numbers of sequences used for the analyses are listed in Supplemental Fig. 2.

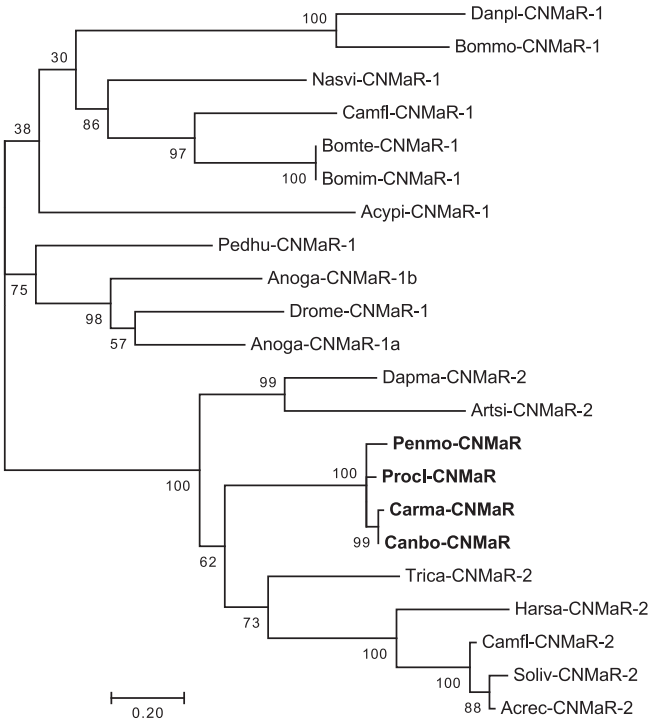

Fig. 5. Phylogenetic relationship of putative decapod CNMamide receptors with Type-1 and Type-2 CNMamide receptors (see Jung et al. (2014)) from diverse arthropods. Maximum likelihood tree depicting the inferred evolutionary history of the respective receptor sequences. The tree with the highest log likelihood is drawn to scale with branch lengths measured in the number of substitutions per site. The percentage of trees in which the associated taxa clustered together across 1000 replicates is shown next to the branches. Decapod sequences are shown in bold font. Species abbreviations: Acrec, *Acromyrmex echinatior*; Acypi, *Acyrthosiphon pisum*; Anoga, *Anopheles gambiae*; Artsi, *Artemia sinica*; Bomim, *Bombus impatiens*; Bommo, *Bombyx mori*; Bomte, *Bombus terrestris*; Camfl, *Camponotus floridanus*; Canbo, *Cancer borealis*; Carma, *Carcinus maenas*; Danpl, *Danaus plexippus*; Dapma, *Daphnia magna*; Drome, *Drosophila melanogaster*; Harsa, *Harpegnathos saltator*; Nasvi, *Nasonia vitripennis*; Pedhu, *Pediculus humanus corporis*; Penmo, *Penaeus monodon*; Procl, *Procambarus clarkii*; Soliv, *Solenopsis invicta*; Trica, *Tribolium castaneum*. Accession numbers of sequences used for the analyses are listed in Supplemental Fig. 3.

Specifically, transcripts encoding both a putative CNMamide precursor and putative CNMamide receptor were identified in a *C. borealis* nervous system-specific assembly (BioProject No. PRJNA310325; Northcutt et al., 2016), suggesting that the CNMamide signaling system serves to modulate nervous system functioning, as has been found for many other peptidergic signaling systems in members of the Decapoda (*e.g.*, Christie, 2011; Christie et al., 2010), a function also proposed for the CNMamides in *D. melanogaster* (Jung et al., 2014). At least one likely site of CNMamide production in the decapod nervous system appears to be the eyestalk ganglia, which is the locus of the X-organ-sinus gland system (XO-SG), a major neuroendocrine organ (*e.g.*, Christie, 2011), as transcripts encoding putative CNMamide precursors were identified from *P. clarkii* (BioProject No. PRJNA205889; Manfrin et al., 2015), *H. americanus* (BioProject No. PRJNA338672; Christie et al., 2017), and *E. sinensis* (BioProject No. PRJNA259036; Xu et al., 2015) eyestalk-specific transcriptomes. The presence of a transcript encoding a CNMamide receptor in the *E. sinensis* eyestalk assembly suggests the possibilities that CNMamide modulates photic input to the nervous system and its integration and/or modulation of peptide hormone production/release from the XO-SG system.

In addition to the nervous system, there is evidence for expression of both the CNMamide precursor and its cognate receptor in decapod reproductive tissues. Specifically, transcripts for both are present in a *S.*

olivacea testis-specific transcriptome (BioProject No. PRJNA289610; Waiho et al., 2017), whereas putative precursor-encoding sequences were identified from *P. sanguinolentus* (BioProject No. PRJNA415705; Zhang et al., 2018) and *C. feriata* (BioProject No. PRJNA415670; unpublished direct GenBank submission) mixed testis-ovary assemblies. Enrichment of CNMamide precursor-encoding transcripts in the ovary was reported previously for *P. clarkii* (Veenstra, 2015). Collectively these findings suggest a role for the CNMamides in decapod reproductive control and/or the signaling of reproductive state to other physiological/behavioral control systems, *e.g.*, the nervous system. Interestingly, the only assemblies that contain transcripts encoding Type-2 CNMamide precursor variants (the variant missing the CNMamide) are those that contain reproductive tissue as at least one source of RNA, raising the possibility that expression of this splice variant is reproductive tissue-specific.

4. Summary and conclusions

In this study, publicly accessible transcriptomic data were used to assess the phylogenetic/structural conservation, tissue localization, and possible functions of CNMamide signaling systems in members of the Decapoda. This peptidergic signaling system appears broadly, if not ubiquitously, conserved in this taxon. Interestingly, the decapod CNMamide gene appears to give rise to two splice variants, one containing CNMamide and one in which the C-terminal portion of the precursor containing the CNMamide peptide is replaced by a structurally distinct and unrelated bridged peptide. Furthermore, the respective cognate CNMamide receptor is present in members of the Decapoda as a single highly conserved protein that aligns with the Type-2 CNMamide receptor group. Given the presence of both precursor and receptor transcripts in nervous system- and reproductive tissue-specific transcriptomes, we propose that members of the CNMamide family are likely to serve as modulators of both the nervous and reproductive systems. These data provide a foundation for initiating molecular and physiological studies of the CNMamide signaling system in decapod species.

Acknowledgements

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

Funding

This work was supported by funds from the Cades Foundation of Honolulu, Hawaii (to AEC) and base CRIS funding (Project #2020-22620-022-00D to JJH).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ygcen.2019.113217.

References

Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Scherer, S.E., Li, P.W., Hoskins, R.A., Galle, R.F., George, R.A., Lewis, S.E., Richards, S., Ashburner, M., Henderson, S.N., Sutton, G.G., Wortman, J.R., Yandell, M.D., Zhang, Q., Chen, L.X., Brandon, R.C., Rogers, Y.H., Blazej, R.G., Champe, M., Pfeiffer, B.D., Wan, K.H., Doyle, C., Baxter, E.G., Helt, G., Nelson, C.R., Gabor, G.L., Abril, J.F., Agbayani, A., An, H.J., Andrews-Pfannkoch, C., Baldwin, D., Ballew, R.M., Basu, A., Baxendale, J., Bayraktaroglu, L., Beasley, E.M., Beeson, K.Y., Benos, P.V., Berman, B.P., Bhandari, D., Bolshakov, S., Borkova, D., Botchan, M.R., Bouck, J., Brokstein, P., Brottier, P., Burtis, K.C., Busam, D.A., Butler, H., Cadieu, E., Center, A., Chandra, I., Cherry, J.M., Cawley, S., Dahlke, C., Davenport, L.B., Davies, P., de Pablos, B., Delcher, A., Deng, Z., Mays, A.D., Dew, I., Dietz, S.M., Dodson, K., Doup, L.E., Downes, M., Dugan-Rocha, S., Dunkov, B.C., Dunn, P., Durbin, K.J., Evangelista, C.C., Ferraz, C., Ferriera, S., Fleischmann, W., Fosler, C., Gabrielian, A.E., Garg, N.S., Gelbart, W.M., Glasser, K., Glodek, A., Gong, F., Gorrell, J.H., Gu, Z., Guan, P., Harris, M., Harris, N.L., Harvey, D., Heiman, T.J., Hernandez, J.R., Houck, J., Hostin, D., Houston, K.A., Howland, T.J., Wei, M.H., Ibegwam, C., Jalali, M., Kalush, F., Karpen, G.H., Ke, Z., Kennison, J.A., Ketchum, K.A., Kimmel, B.E., Kodira, C.D., Kraft, C., Kravitz, S., Kulp, D., Lai, Z., Lasko, P., Lei, Y., Levitsky, A.A., Li, J., Li, Z., Liang, Y., Lin, X., Liu, X., Mattei, B., McIntosh, T.C., McLeod, M.P., McPherson, D., Merkulov, G., Milshina, N.V., Mobarry, C., Morris, J., Moshrefi, A., Mount, S.M., Moy, M., Murphy, B., Murphy, L., Muzny, D.M., Nelson, D.L., Nelson, D.R., Nelson, K.A., Nixon, K., Nusskern, D.R., Pacleb, J.M., Palazzolo, M., Pittman, G.S., Pan, S., Pollard, J., Puri, V., Reese, M.G., Reinert, K., Remington, K., Saunders, R.D., Scheeler, F., Shen, H., Shue, B.C., Sidén-Kiamos, I., Simpson, M., Skupski, M.P., Smith, T., Spier, E., Spradling, A.C., Stapleton, M., Strong, R., Sun, E., Svirskas, R., Tector, C., Turner, R., Venter, E., Wang, A.H., Wang, X., Wang, Z.Y., Wassarman, D.A., Weinstock, G.M., Weissenbach, J., Williams, S.M., Woodage, T., Worley, K.C., Wu, D., Yang, S., Yao, Q.A., Ye, J., Yeh, R.F., Zaveri, J.S., Zhan, M., Zhang, G., Zhao, Q., Zheng, L., Zheng, X.H., Zhong, F.N., Zhong, W., Zhou, X., Zhu, S., Zhu, X., Smith, H.O., Gibbs, R.A., Myers, E.W., Rubin, G.M., Venter, J.C., 2000. The genome sequence of *Drosophila melanogaster*. Science 287, 2185–2195.

Bao, C., Yang, Y., Huang, H., Ye, H., 2015. Neuropeptides in the cerebral ganglia of the mud crab, *Scylla paramamosain*: transcriptomic analysis and expression profiles during vitellogenesis. Sci. Rep. 5, 17055.

Bao, C., Yang, Y., Zeng, C., Huang, H., Ye, H., 2018. Identifying neuropeptide GPCRs in the mud crab, *Scylla paramamosain*, by combinatorial bioinformatics analysis. Gen. Comp. Endocrinol. 269, 122–130.

Buckley, S.J., Fitzgibbon, Q.P., Smith, G.G., Ventura, T., 2016. *In silico* prediction of the G-protein coupled receptors expressed during the metamorphic molt of *Sagmariasus verreauxi* (Crustacea: Decapoda) by mining transcriptomic data: RNA-seq to repertoire. Gen. Comp. Endocrinol. 228, 111–127.

Charrier, N.P., Couton, M., Voordouw, M.J., Rais, O., Durand-Hermouet, A., Hervet, C., Plantard, O., Rispe, C., 2018. Whole body transcriptomes and new insights into the biology of the tick *Ixodes ricinus*. Parasit. Vectors 11, 364.

Christie, A.E., 2011. Crustacean neuroendocrine systems and their signaling agents. Cell Tissue Res. 345, 41–67.

Christie, A.E., 2014. Prediction of the peptidomes of *Tigriopus californicus* and *Lepeophtheirus salmonis* (Copepoda, Crustacea). Gen. Comp. Endocrinol. 201, 87–106.

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., 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., Pascual, M.G., Yu, A., 2018. Peptidergic signaling in the tadpole shrimp *Triops newberryi*: a potential model for investigating the roles played by peptide paracrines/hormones in adaptation to environmental change. Mar. Genomics 39, 45–63.

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., Stemmler, E.A., Dickinson, P.S., 2010. Crustacean neuropeptides. Cell. Mol. Life Sci. 67, 4135–4169.

Christie, A.E., Yu, A., 2019. Identification of peptide hormones and their cognate receptors in *Jasus edwardsii* – a potential resource for the development of new aquaculture management strategies for rock/spiny lobsters. Aquaculture 503, 636–662.

Dickinson, P.S., Hull, J.J., Miller, A., Oleisky, E.R., Christie, A.E., 2019. To what extent may peptide receptor gene diversity/complement contribute to functional flexibility in a simple pattern-generating neural network? Comp. Biochem. Physiol. Part D Genomics Proteomics 30, 262–282.

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., 2011. Genomics, transcriptomics, and peptidomics of *Daphnia pulex* neuropeptides and protein hormones. J. Proteome Res. 10, 4478–4504.

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792–1797.

El-Gebali, S., Mistry, J., Bateman, A., Eddy, S.R., Luciani, A., Potter, S.C., Qureshi, M., Richardson, L.J., Salazar, G.A., Smart, A., Sonnhammer, E.L.L., Hirsh, L., Paladin, L., Piovesan, D., Tosatto, S.C.E., Finn, R.D., 2019. The Pfam protein families database in 2019. Nucleic Acids Res. 47, D427–D432.

Ferrè, F., Clote, P., 2005. DiANNA: a web server for disulfide connectivity prediction. Nucleic Acids Res. 33, W230–W232.

Hummon, A.B., Richmond, T.A., Verleyen, P., Baggerman, G., Huybrechts, J., Ewing, M.A., Vierstraete, E., Rodriguez-Zas, S.L., Schoofs, L., Robinson, G.E., Sweedler, J.V., 2006. From the genome to the proteome: uncovering peptides in the *Apis* brain. Science 314, 647–649.

Jones, D.T., Taylor, W.R., Thornton, J.M., 1992. The rapid generation of mutation data matrices from protein sequences. Comput. Appl. Biosci. 8, 275–282.

Jung, S.H., Lee, J.H., Chae, H.S., Seong, J.Y., Park, Y., Park, Z.Y., Kim, Y.J., 2014. Identification of a novel insect neuropeptide, CNMa and its receptor. FEBS Lett. 588, 2037–2041.

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

improvements in performance and usability. Mol. Biol. Evol. 30, 772–780.
Kumar, S., Stecher, G., Li, M., Knyaz, C., Tamura, K., 2018. MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol. Biol. Evol. 35, 1547–1549.
Le, S.Q., Gascuel, O., 2008. An improved general amino acid replacement matrix. Mol. Biol. Evol. 25, 1307–1320.
Lee, K.M., Daubnerová, I., Isaac, R.E., Zhang, C., Choi, S., Chung, J., Kim, Y.J., 2015. A neuronal pathway that controls sperm ejection and storage in female *Drosophila*. Curr. Biol. 25, 790–797.
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.
Monigatti, F., Gasteiger, E., Bairoch, A., Jung, E., 2002. The Sulfinator: predicting tyrosine sulfation sites in protein sequences. Bioinformatics 18, 769–770.
Northcutt, A.J., Lett, K.M., Garcia, V.B., Diester, C.M., Lane, B.J., Marder, E., Schulz, D.J., 2016. Deep sequencing of transcriptomes from the nervous systems of two decapod crustaceans to characterize genes important for neural circuit function and modulation. BMC Genomics 17, 868.
Oliphant, A., Alexander, J.L., Swain, M.T., Webster, S.G., Wilcockson, D.C., 2018. Transcriptomic analysis of crustacean neuropeptide signaling during the moult cycle in the green shore crab, *Carcinus maenas*. BMC Genomics 19, 711.
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.
Poynton, H.C., Hasenbein, S., Benoit, J.B., Sepulveda, M.S., Poelchau, M.F., Hughes, D.S.T., Murali, S.C., Chen, S., Glastad, K.M., Goodisman, M.A.D., Werren, J.H., Vineis, J.H., Bowen, J.L., Friedrich, M., Jones, J., Robertson, H.M., Feyereisen, R., Mechler-Hickson, A., Mathers, N., Lee, C.E., Colbourne, J.K., Biales, A., Johnston, J.S., Wellborn, G.A., Rosendale, A.J., Cridge, A.G., Munoz-Torres, M.C., Bain, P.A., Manny, A.R., Major, K.M., Lambert, F.N., Vulpe, C.D., Tuck, P., Blalock, B.J., Lin, Y.Y., Smith, M.E., Ochoa-Acuña, H., Chen, M.M., Childers, C.P., Qu, J., Dugan, S., Lee, S.L., Chao, H., Dinh, H., Han, Y., Doddapaneni, H., Worley, K.C., Muzny, D.M., Gibbs, R.A., Richards, S., 2018. The toxicogenome of Hyalella azteca: a model for sediment ecotoxicology and evolutionary toxicology. Environ. Sci. Technol. 52, 6009–6022.
Rzhetsky, A., Nei, M., 1992. A simple method for estimating and testing minimum evolution trees. Mol. Biol. Evol. 9, 945–967.
Saitou, N., Nei, M., 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4, 406–425.
Tassone, E.E., Geib, S.M., Hall, B., Fabrick, J.A., Brent, C.S., Hull, J.J., 2016. *De novo* construction of an expanded transcriptome assembly for the western tarnished plant bug, *Lygus hesperus*. Gigascience 5, 6.
Thurmond, J., Goodman, J.L., Strelets, V.B., Attrill, H., Gramates, L.S., Marygold, S.J., Matthews, B.B., Millburn, M., Antonazzo, G., Trovisco, V., Kaufman, T.C., Calvi, B.R., the FlyBase Consortium, 2019. FlyBase 2.0: the next generation. Nucleic Acids Res. 47, D759–D765.
Veenstra, J.A., 2000. Mono- and dibasic proteolytic cleavage sites in insect neuroendocrine peptide precursors. Arch. Insect. Biochem. Physiol. 43, 49–63.
Veenstra, J.A., 2015. The power of next-generation sequencing as illustrated by the neuropeptidome of the crayfish *Procambarus clarkii*. Gen. Comp. Endocrinol. 224, 84–95.
Veenstra, J.A., Rombauts, S., Grbić, M., 2012. *In silico* cloning of genes encoding neuropeptides, neurohormones and their putative G-protein coupled receptors in a spider mite. Insect Biochem. Mol. Biol. 42, 277–295.
Ventura, T., Cummins, S.F., Fitzgibbon, Q., Battaglene, S., Elizur, A., 2014. Analysis of the central nervous system transcriptome of the eastern rock lobster *Sagmariasus verreauxi* reveals its putative neuropeptidome. PLoS ONE 9, e97323.
Waiho, K., Fazhan, H., Shahreza, M.S., Moh, J.H., Noorbaiduri, S., Wong, L.L., Sinnasamy, S., Ikhwanuddin, M., 2017. Transcriptome analysis and differential gene expression on the testis of orange mud crab, *Scylla olivacea*, during sexual maturation. PLoS ONE 12, e0171095.
Xu, Z., Zhao, M., Li, X., Lu, Q., Li, Y., Ge, J., Pan, J., 2015. Transcriptome profiling of the eyestalk of precocious juvenile Chinese mitten crab reveals putative neuropeptides and differentially expressed genes. Gene 569, 280–286.
Zhang, Y., Miao, G., Wu, Q., Lin, F., You, C., Wang, S., Aweya, J.J., Ma, H., 2018. Transcriptome sequencing and molecular markers discovery in the gonads of *Portunus sanguinolentus*. Sci. Data 5, 180131.
Zhang, X., Yuan, J., Sun, Y., Li, S., Gao, Y., Yu, Y., Liu, C., Wang, Q., Lv, X., Zhang, X., Ma, K.Y., Wang, X., Lin, W., Wang, L., Zhu, X., Zhang, C., Zhang, J., Jin, S., Yu, K., Kong, J., Xu, P., Chen, J., Zhang, H., Sorgeloos, P., Sagi, A., Alcivar-Warren, A., Liu, Z., Wang, L., Ruan, J., Chu, K.H., Liu, B., Li, F., Xiang, J., 2019. Penaeid shrimp genome provides insights into benthic adaptation and frequent molting. Nat. Commun. 10, 356.