Ophiuroid phylotranscriptomics enables discovery of novel echinoderm representatives of bilaterian neuropeptide families and reconstruction of neuropeptide precursor evolution over ~270 million years.

Meet Zandawala1,2,*, Ismail Moghul1,#, Luis Alfonso Yañez Guerra1,#, Jérôme Delroisse1,#, Nikara Abylkassimova1, Andrew F. Hugall3, Timothy D. O’Hara3 and Maurice R. Elphick1,*

Meet Zandawala: meet.zandawala@zoologi.su.se
Ismail Moghul: muhammad.moghul.16@ucl.ac.uk
Luis Alfonso Yañez Guerra: l.a.yanezguerra@qmul.ac.uk
Jérôme Delroisse: jerome.delroisse@umons.ac.be
Nikara Abylkassimova: abylkassimova@gmail.com
Andrew F. Hugall: ahugall@museum.vic.gov.au
Timothy D. O’Hara: tohara@museum.vic.gov.au
Maurice R. Elphick: m.r.elprick@qmul.ac.uk

1. Queen Mary University of London, School of Biological & Chemical Sciences, Mile End Road, London, E1 4NS, UK
2. Present address: Stockholm University, Department of Zoology, Stockholm, Sweden
3. Museums Victoria, GPO Box 666, Melbourne, VIC 3001, Australia

# Contributed equally
* Correspondence to: Prof. M.R. Elphick and Dr. M. Zandawala
Tel: 44 207 882 6664
Fax: 44 208 983 0973
E-mail: m.r.elprick@qmul.ac.uk and meet.zandawala@zoologi.su.se
Abstract

Background:

Neuropeptides are a diverse class of intercellular signaling molecules that mediate neuronal regulation of many physiological and behavioural processes, including feeding, reproduction and locomotion. Recent advances in genome/transcriptome sequencing are enabling identification of neuropeptide precursor proteins in species from a growing variety of animal taxa, providing new insights into the evolution of neuropeptide signaling. Here we report a phylo-transcriptomic analysis of neuropeptide precursors in over fifty species of brittle stars (Class Ophiuroidea; Phylum Echinodermata).

Results:

Detailed analysis of transcriptome sequence data from three brittle star species, *Ophionotus victoriae, Amphiura filiformis* and *Ophiopsila aranea*, enabled the first comprehensive identification of neuropeptide precursors in ophiuroids. Representatives of over thirty bilaterian neuropeptide precursor families were identified, some of which occur as paralogs (e.g. thyrotropin-releasing hormone, corticotropin-releasing hormone, cholecystokinin, somatostatin and pedal peptide). Furthermore, homologs of endothelin/CCHamide, eclosion hormone, neuropeptide-F/Y and nucleobinin/nesfatin were discovered here in a deuterostome/echinoderm for the first time. The majority of ophiuroid neuropeptide precursors contain a single copy of a neuropeptide, but several precursors comprise multiple copies of identical or non-identical, but structurally-related, neuropeptides. Here we performed an unprecedented investigation of the evolution of neuropeptide copy-number over a period of ~270 million years by analysing sequence data from over fifty ophiuroid species, with reference to a robust phylogeny. Interestingly, the number of neuropeptide copies in the majority of precursors was constant across all the species examined, but examples of clade-specific losses/gains of neuropeptides were also observed.

Conclusions:

We report here the most comprehensive analysis to date of neuropeptide precursors in the phylum Echinodermata, with novel representatives of several bilaterian neuropeptide families discovered for the first time in echinoderms. Furthermore, analysis of precursor proteins comprising multiple copies of identical or related neuropeptides across ~270 million years of ophiuroid evolution indicates that the composition of neuropeptide “cocktails” is functionally important, but with plasticity over long evolutionary time scales.
Keywords (3 to 10):
Neuropeptide; echinoderm; Ophiuroidea; eclosion hormone; CCHamide; neuropeptide-Y;
evolution
**Introduction**

The nervous systems of animals utilize a wide variety of chemicals for neuronal communication. These include amino acids (e.g. glutamate), biogenic amines (e.g. serotonin), and neuropeptides (e.g. vasopressin) amongst others. Neuropeptides are by far the most diverse and they control many physiological/behavioural processes, including feeding, reproduction and locomotion [1-3]. Recent advances in genome/transcriptome sequencing are enabling identification of neuropeptide precursor proteins in species from a growing variety of animal taxa, providing new insights into the evolution of neuropeptide signaling [4-8]. The echinoderms are notable in this regard because as deuterostomian invertebrates they occupy an “intermediate” phylogenetic position with respect to the vertebrates and intensely studied protostomian invertebrates such as insects (e.g. *Drosophila melanogaster*) and nematodes (e.g. *Caenorhabditis elegans*). Accordingly, characterisation of neuropeptide signaling systems in echinoderms has recently provided key “missing links” for determination of neuropeptide relationships and reconstruction of neuropeptide evolution [8-10].

The phylum Echinodermata comprises five extant classes: Echinoidea (sea urchins and sand dollars), Holothuroidea (sea cucumbers), Asteroidea (starfish), Ophiuroidea (brittle stars and basket stars) and Crinoidea (sea lilies and feather stars). Recent molecular phylogenetic studies support the hypothesis that Echinoidea and Holothuroidea are sister groups (Echinozoa) and Asteroidea and Ophiuroidea are sister groups (Asterozoa), with the Crinoidea basal to the Echinozoa + Asterozoa clade (Eleutherozoa) [11, 12]. Echinoderms are marine organisms that have several unique features including pentaradial symmetry as adults, a remarkable ability to autotomise and regenerate body parts, and neurally-controlled mutable collagenous tissue [13, 14]. Previous transcriptomic analyses have identified neuropeptide precursor complements in *Strongylocentrotus purpuratus* (purple sea urchin), *Apostichopus japonicus* (Japanese sea cucumber) and *Asterias rubens* (common European starfish) [8, 15, 16]. Furthermore, the identification of neuropeptides in these species has facilitated investigation of the evolution and physiological roles of various neuropeptide signaling systems [8-10, 17-21].

The recent progress in transcriptomic/genomic characterization of echinoderm neuropeptide systems has hitherto not been extended to ophiuroids or crinoids. The Ophiuroidea constitutes the largest class among extant echinoderms [22] with a long evolutionary history that extends back to the early Ordovician (around 480 million years ago) [23], whilst extant families date from the mid-Permian (~ 270 million years ago) [12]. Available molecular data for ophiuroids has increased significantly in recent years with the emergence of numerous transcriptomic studies [20, 24-29]. Here, we utilize transcriptome
sequence data from three brittle star species, *Ophionotus victoriae, Amphiura filiformis* and *Ophiopsila aranea* to perform the first comprehensive identification of neuropeptide precursors in ophiuroids. We identify representatives of over thirty neuropeptide families including homologs of endothelin/CCHamide, eclosion hormone (EH), neuropeptide-F/Y (NPF/NPY) and nucleobinin (NUCB)/nesfatin, which are the first to be discovered in a deuterostome/echinoderm.

Transcriptomes have also been employed to investigate the phylogenetic relationships of the ophiuroids, utilising data from fifty-two species [12]. In this the most comprehensive molecular analysis of ophiuroid phylogeny to date, previous morphology-based classification schemes [30] were rejected in favour of a new phylogeny comprising three primary ophiuroid clades [12, 31, 32]. This landmark study and the associated large dataset has provided a unique opportunity to investigate the conservation and diversification of neuropeptide precursor structure over a period of ~270 million years of ophiuroid evolution. Our analysis reveals that the majority of ophiuroid neuropeptide precursors contain a single copy of a neuropeptide, but several precursors comprise multiple copies of identical or non-identical, but structurally-related, neuropeptides. Interestingly, the number of neuropeptide copies in the majority of precursors is constant across all the ophiuroid species examined, but examples of clade-specific losses/gains of neuropeptides are also observed. This remarkable conservation in neuropeptide copy number across ~270 million years of ophiuroid evolution indicates that the composition of neuropeptide “cocktails” is functionally important, but with plasticity over long evolutionary time scales.
Results and discussion

Here we have identified ophiuroid homologs of neuropeptide precursors that have been identified previously in other echinoderms and these include, alphabetically: AN peptides, bursicon (α and β), calcitonin, cholecystokinin (CCK), corazonin, corticotropin-releasing hormone (CRH), glycoprotein hormones (α2 and β5), gonadotropin-releasing hormone (GnRH), insulin-like peptide (ILP), kisspeptin (KP), luqin, melanin-concentrating hormone (MCH), NG peptides (neuropeptide-S), orexin, pedal peptides, pigment-dispersing factor (PDF), relaxin-like peptide, SALMFamides (L-type and F-type), somatostatin, tachykinin, thyrotropin-releasing hormone (TRH) and vasopressin/oxytocin. Identification of ophiuroid representatives of these neuropeptide precursor types has in some cases provided new insights into neuropeptide precursor structure and evolution, as discussed in more detail below. First, however, we will highlight representatives of bilaterian neuropeptide precursor families that have been identified here for the first time in an echinoderm species.

Discovery of the first echinoderm representatives of bilaterian neuropeptide families

Comprehensive analysis of transcriptome sequence data from three ophiuroid species, O. victoriae, A. filiformis and O. aranea, has enabled the discovery of the first echinoderm representatives of four bilaterian neuropeptide families. Specifically, we have discovered the first deuterostomian homologs of eclosion hormone (Figure 2), the first ambulacrarian homolog of CCHamide/endothelin-type peptides (Figure 3A), and the first echinoderm homologs of neuropeptide-Y/neuropeptide-F (Figure 3B) and NUCB/nesfatin (Figure S1), as discussed in detail below.

Eclosion hormone

Eclosion hormone (EH) was first isolated and sequenced in the insects Manduca sexta (tobacco hornworm) and Bombyx mori (silk moth) and shown to alter the timing of adult emergence [33, 34]. EH is one of the main peptide/protein hormones involved in control of ecdysis (i.e. shedding of the cuticle) behavior in insects [35]. It binds to and activates a receptor guanylyl cyclase that is expressed in epitracheal Inka cells and causes the secondary release of ecdysis-triggering hormone (ETH) that is also expressed in Inka cells [36, 37]. In Drosophila, EH is important but not essential for ecdysis as some flies lacking EH are able to undergo ecdysis [38]. Insect EHs have six conserved cysteine residues that form three disulfide bridges [36]. EHs have not been discovered previously outside of arthropods. Interestingly, four EH-like precursors were identified in A. filiformis and O. aranea and two in O. victoriae (Figure S2-S4). The ophiuroid EH-like precursors are orthologous to
neuropeptide precursors previously identified in the sea-urchin *S. purpuratus* (Spnp11 and Spnp15, which we now rename as Spur EH1 and Spur EH2, respectively) [16] and the starfish *A. rubens* (Arnp11, Arnp15 and Arnp15b renamed as Arub EH1, Arub EH2a and Arub EH2b, respectively) [8]. The positions of cysteine residues are conserved across all echinoderm and insect EHs, but aside from this there is little sequence conservation (Figure 2A). The echinoderm EH-like precursor sequences were also analysed using a sequence-similarity-based clustering approach based on BLASTp e-values using CLANS software [39]. The analysis shows that echinoderm EH-like precursors (i) cluster in two compact subgroups (echinoderm EH-like precursor 1 and EH-like precursor 2 and (ii) have strong positive BLAST results with arthropod EHs and, to a lesser extent, with arthropod ion transport peptide (ITP) and vertebrate atrial natriuretic peptide (ANP) (Figure 2B). ITP precursors also possess six cysteine residues; however, the position of these residues is not conserved with cysteine residues found in echinoderm EH-like precursors (not shown).

To obtain further evidence for the presence of an EH-like signaling system in echinoderms, we performed a phylogenetic analysis of EH-type receptors. Insect EHs mediate their effects by binding to membrane guanylyl cyclase receptors [37]. EH receptors are closely related to vertebrate ANP receptors and various orphan receptors [40]. Specific BLAST searches enabled identification of transcripts in *O. victoriae*, *A. filiformis* and *O. aranea* that encode proteins similar to arthropod EH receptors. Maximum likelihood and Bayesian phylogenetic analyses confirmed that these sequences group with the receptor cluster containing EH receptors (Figure 2C). The discovery of the first deuterostomian EHs suggests an ancient bilaterian origin of EHs and indicates that these hormones may have other functions in invertebrates aside from their role in ecdysis.

CCHamide

CCHamides are neuropeptides that were discovered relatively recently in the silkworm *Bombyx mori* [41]. Later, it was found that insects have two CCHamide genes, CCHamide-1 and CCHamide-2, each encoding a single copy of the mature peptide [42]. These peptides are referred to as CCHamides because they contain two cysteine residues and a characteristic histidine-amide C-terminal motif. There are two CCHamide receptors in insects: CCHamide-1 specifically activates one receptor and CCHamide-2 specifically activates the second receptor [42, 43]. CCHamide-1 has a physiological role in starvation-induced olfactory modifications [44] whereas as CCHamide-2 regulates feeding, growth and developmental timing in flies [43, 45]. Recent studies examining the evolution of neuropeptides in the Bilateria have shown that protostomian CCHamides are related to
elevenin (another protostomian neuropeptide originally discovered from the mollusc *Aplysia californica* L11 neuron), lophotrochozoan GGNG peptides, endothelins and gastrin-releasing peptides (GRPs) [6, 7, 46, 47]. The latter two are neuropeptide types that have not been found outside chordates. Furthermore, the degree of sequence/structural conservation varies across these different peptide families. Hence, CCHamides are amidated and have a disulphide bridge, elevenins and endothelins have a disulphide bridge but are non-amidated and GRPs are amidated but lack the disulphide bridge. Furthermore, CCHamide-1 is located immediately after the signal peptide whereas there is a dibasic cleavage site separating the signal peptide and CCHamide-2 [42].

Here we have identified two neuropeptide precursors in brittle stars whose sequence and precursor structure resembles those of lophotrochozoan GGNG peptides and insect CCHamide-1 (*Figure 3A*). The CCHamide-like precursor 1 identified in *O. victoriae* is orthologous to an uncharacterized neuropeptide precursor (Arnp25) identified previously in the starfish *A. rubens* [8], whereas the CCHamide-like precursor 2 was only found in brittle stars. Both CCHamide-like precursors in *O. victoriae* comprise a single copy of a putative cyclic amidated peptide that is flanked by a signal peptide at the N-terminus and a dibasic cleavage site at the C-terminus. Interestingly, both of these peptides lack a penultimate histidine residue, just like the lophotrochozoan GGNG peptides (*Figure 3A*) [46, 47].

**Neuropeptide-Y/Neuropeptide-F**

Neuropeptide-Y (NPY) was first isolated and sequenced from the porcine hypothalamus in 1982 [48, 49]. Although the NPY/NPF family of peptides are pleiotropic in nature [50], they are mainly known for their roles in regulation of feeding and stress [3, 51, 52]. The discovery of Neuropeptide-F (NPF) in the tapeworm *Moniezia expansa* in 1991 demonstrated for the first time the occurrence of NPY homologs in invertebrates [53]. Here, we have identified the first echinoderm representatives of the NPY/NPF family in brittle stars and starfish (*Figure 3B*). The brittle star precursors contain a peptide with a C-terminal RYamide, in common with NPY in vertebrates and an ortholog in the starfish *Patiria miniata*. In contrast, an ortholog in the starfish *A. rubens* has a C-terminal RFamide, a feature that it shares with NPY/NPF-type peptides in the hemichordate *S. kowalevskii* and in protostomes. Thus, our findings have revealed that NPY/NPF-type peptides with a C-terminal Yamide motif are not restricted to vertebrates. Echinoderm NPY/NPF-type peptides are located immediately after the signal peptide in the precursor proteins, as is the case in other bilaterian species. Surprisingly, we did not find NPY/NPF-type precursors in the sea urchin *S. purpuratus* or the sea cucumber *A. japonicus*. However, we suspect that this may reflect
sequence divergence rather than gene loss because a gene encoding a NPY/NPF-type receptor can be found in the *S. purpuratus* genome [54].

**NUCB**

Nucleobindins (NUCB1 and NUCB2) are multidomain Ca$^{2+}$ and DNA binding proteins. NUCB1 was first discovered in 1992 and thought to play a role in apoptosis and autoimmunity [55]. Interestingly, the NUCB1 precursor has both a signal peptide and a leucine zipper structure suggesting that it can bind DNA and act as an endocrine factor [56]. NUCB2 is a homolog of NUCB1 and was named based on high sequence similarity between the two precursors [57]. In 2006, an 82 amino acid peptide located in the N-terminal region of NUCB2 was reported. This peptide, Nesfatin-1 (Nucleobindin-2-Encoded Satiety and FAT-Influencing protein-1), was discovered as a satiety inducing factor in the rat hypothalamus [58]. Its role in inhibiting food intake in vertebrates is now well-established [57, 59]. Moreover, this pleiotropic peptide also modulates other processes including glucose and lipid metabolism, and cardiovascular and reproductive functions. Recently, nesfatin-1-like peptide derived from NUCB1 was shown to be anorexigenic in goldfish [60]. Surprisingly, the presence of NUCBs in invertebrates had not been reported, in spite of the potential therapeutic applications of these molecules in obesity related disorders. Here, we show that NUCB-type precursors are present in echinoderms (*Figure S1A*). Phylogenetic analysis of NUCB precursors reveals that a single copy of the NUCB precursor is found in invertebrate species and gene duplication in the vertebrate lineage gave rise to NUCB1 and NUCB2 (*Figure S1B*). In chordates, the NUCB precursors are predicted to generate three peptides (Nesfatin-1, 2 and 3); however, no biological role has been attributed specifically to nesfatin-2 and nesfatin-3. Interestingly, the prohormone convertase cleavage sites expected to generate Nesfatin-1, 2 and 3 are conserved between echinoderm and chordate NUCBs. Moreover, the *O. victoriae* precursor has an additional predicted cleavage site within the Nesfatin-1 containing region, which is not present in other species (except for *Drosophila melanogaster*). However, it remains to be determined whether or not this cleavage site in the *O. victoriae* precursor is functional.

**First comprehensive identification of neuropeptide precursors in ophiuroids**

We have identified neuropeptide precursors belonging to 32 families, which represents the first comprehensive analysis of neuropeptide precursors in ophiuroids (*Figure 4; Figure S2-S4*). Several of these neuropeptide families have been identified previously in echinoderms and include homologs of AN peptides, bursicon (α and β), calcitonin, CCK
corazonin, CRH, glycoprotein hormones (α2 and β5), GnRH, ILP, KP, luqin, MCH, NG peptides (neuropeptide-S), orexin, pedal peptides, PDF, relaxin-like peptide, SALMFamides (L-type and F-type), somatostatin, tachykinin, TRH and vasopressin/oxytocin (Figures 5-7 and S5-S9). With the exception of MCH (which may be unique to deuterostomes), the origins of all of the neuropeptide precursors identified here in ophiuroids predate the divergence of protostomes and deuterostomes. Of the three species examined here, the neuropeptide precursor complement of O. victoriae was the most complete (Figure 4) and therefore this species is used as a representative ophiuroid for sequence alignments, except in a few cases where a neuropeptide precursor was not found in O. victoriae. Below we highlight several interesting and/or unusual features of ophiuroid neuropeptides and neuropeptide precursors.

Neuropeptide precursors that occur in multiple forms in O. victoriae

Thyrotropin-releasing hormone (TRH)-type precursors

TRH (also known as thyrotropin-releasing factor or thyroliberin) was first isolated and sequenced in the 1960s. In mammals, TRH is produced in the hypothalamus and stimulates the release of thyroid-stimulating hormone (TSH) and prolactin from the anterior pituitary. The recent discovery of a TRH receptor in the annelid Platynereis dumerilii indicates that the evolutionary origin of this neuropeptide signaling system predates the divergence of protostomes and deuterostomes.

The human TRH precursor contains six copies of the tripeptide pQHPamide. Precursor proteins comprising multiple copies of TRH-like peptides have been identified previously in the sea urchin S. purpuratus, the sea cucumber A. japonicus and the starfish A. rubens, with a single TRH-type precursor found in each of these species. Interestingly, here we identified two TRH-type precursors (OvTRHP1 and OvTRHP2) in O. victoriae (Figure S2 and 6A). OvTRHP1 comprises 21 copies of putative TRH-like tetrapeptides with the motif pQXXXamide (where X is variable). OvTRHP2, on the other hand, comprises two copies of the putative tetrapeptide pQGPRamide and two longer peptides that also have a C-terminal GPRamide motif but lack the N-terminal pyroglutamate.

Cholecystokinin (CCK)-type precursors
A CCK-type peptide (formerly pancreozymin) was first sequenced in the 1960s [72]. CCK-type peptides play numerous roles in feeding and digestion related physiology. CCK mediates satiety, stimulates the release of digestive enzymes and gall bladder contractions [73-75]. CCK-type peptides are involved in mechanisms of learning and memory, and analgesia [76]. A neuropeptide precursor comprising two CCK-like peptides was recently identified in the starfish A. rubens [8]. Here we have identified two CCK-type precursors in O. victoriae (OvCCKP1 and OvCCKP2) and orthologs of both of these precursors were also identified in the sea urchin S. purpuratus (Figure S2) [16]. The CCK-type precursor 1 comprises three CCK-like peptides in both O. victoriae and S. purpuratus and this precursor is similar to the A. rubens CCK-type precursor, which comprises two CCK-like peptides. In contrast, the CCK-type precursor 2 comprises a single CCK-like peptide in both O. victoriae and S. purpuratus. Interestingly, the sequence of the S. purpuratus CCK-type precursor 2 was reported previously as part of a genome-wide search for neuropeptides [77], but the authors of this study did not identify it as a CCK-type precursor. However, based on the presence of a conserved tyrosine residue and a C-terminal F-amide motif in the predicted neuropeptide derived from this protein, it is evident that it belongs to the family of CCK-type precursors (Figure 6B). A search of a preliminary genome assembly of the starfish Patiria miniata (http://www.echinobase.org) [78] did not reveal a gene encoding a CCK-type precursor 2. Therefore, it appears that this neuropeptide precursor type may have been lost in the Asteroidea; nevertheless, further analysis of a wider range of starfish species will be required to draw definitive conclusions. With a broader evolutionary perspective, CCK-type peptides in deuterostomes are orthologs of sulfakinin (SK)-type neuropeptides found in insects [6, 7]. Interestingly, insects have a single SK precursor, which comprises two neuropeptides, SK-1 and SK-2 [79], and this may reflect the ancestral condition in the common ancestor of protostomes and deuterostomes. Thus, the occurrence of two CCK-type peptides on a single precursor in A. rubens and insects may be an ancestral characteristic and the occurrence of two CCK-type precursors that comprise one and three CCK-type peptides appears to be a derived characteristic.

Somatostatin-type precursors

Somatostatin was first isolated and sequenced from sheep hypothalamus in 1973 [80]. This peptide inhibits the release of pituitary hormones such as growth hormone, prolactin and thyroid-stimulating hormone [81]. Moreover, it also inhibits the release of gastrointestinal (cholecystokinin and gastrin amongst others) and pancreatic (insulin and glucagon) hormones [82-84]. Aside from its effects on release of hormones, somatostatin also has central actions.
that influence motor activity [82]. Here, we have identified two somatostatin-type precursors (OvSSP-1 and OvSSP-2) in O. victoriae. (Figure S2 and 6C). Homologs of both of these precursors are present in the sea urchin S. purpuratus (Figure S2 and 6C), one of which was previously referred to as Spnp16 [16]. By comparison, only a single somatostatin-type precursor has been found in the starfish A. rubens, which is an ortholog of OvSSP-1 [8]. All somatostatin-type precursors comprise a single copy of the bioactive neuropeptide, which is located in the C-terminal region of the precursor [85, 86]. Interestingly, the type-1 somatostatins in echinoderms have a phenylalanine residue located in the middle part of the peptide and this conserved feature is found in human somatostatin. Conversely, type-2 somatostatins in echinoderms lack the phenylalanine residue but have a neighbouring tryptophan-lysine (WK) motif that is also conserved in human and B. floridae somatostatins (Figure 6C). The deuterostomian somatostatins are orthologous to the allatostatin-C neuropeptide family in arthropods [7]. This family of peptides comprises three precursor-types: allatostatin-C, allatostatin-CC and the recently discovered allatostatin-CCC [86, 87]. Both allatostatin-C and allatostatin-CC are non-amidated, like somatostatins; however, allatostatin-CCC has a C-terminal amide. Hence, non-amidated peptides may be representative of the ancestral condition in the common ancestor of protostomes and deuterostomes, with the amidated allatostatin-CCC probably having evolved only within the arthropod lineage [87]. It remains to be determined whether or not the duplication of somatostatin-type precursors in echinoderms and the duplication of allatostatin C (to give rise to allatostatin-CC) represent independent duplications. Further insights into this issue may be obtained if the receptors for somatostatin-type peptides in echinoderms are deorphanised.

Corticotropin-releasing hormone (CRH)-type precursors

CRH-type peptides are a family of related neuropeptides that include CRH, urocortins and urotensin-I in chordates, egg-laying hormone (ELH) in lophotrochozoans and diuretic hormone 44 (DH44) in arthropods [6, 7]. Arthropods usually have a single DH44 precursor, which comprises a single copy of the mature peptide. In some insects, such as Tribolium castaneum and Bombyx mori, alternative splicing of DH44 transcripts results in multiple mature peptide isoforms of varying lengths [41, 88]. The situation in lophotrochozoans is more complex, with several species having multiple precursors and some of these precursors comprising multiple ELH mature peptides [4, 89]. A single CRH-type precursor was found previously in the starfish A. rubens, whereas here we have identified four CRH-type precursors in O. victoriae (Figure S2 and 6D). Thus, expanded families of CRH-type
peptides and receptors appear to have evolved independently in multiple animal lineages, including chordates and ophiuroid echinoderms [90, 91].

Diversity in neuropeptide precursor structure: new insights from ophiuroids

Tachykinins

The mammalian neuropeptide substance P was the first tachykinin-type peptide to be isolated and sequenced [92-94]. Subsequently, tachykinin-type peptides were discovered in other animals including tunicates [95], insects [96, 97], annelids [98] and molluscs [99]. Tachykinin-type peptides regulate various physiological processes including muscle contractility [100], nociception [101] and stress responses [102] amongst others [103].

Analysis of genomic/transcriptomic sequence data from the sea urchin S. purpuratus and the sea cucumber A. japonicus did not identify candidate tachykinin-type precursors [6, 7, 15, 16]. However, recently a putative tachykinin-type precursor was discovered in the starfish A. rubens (ArTKP), indicating that this signaling system does occur in some echinoderms [8]. Here we have identified orthologs of ArTKP in O. victoriae and other ophiuroids (Figure 4 and 7A). Collectively, these findings indicate that this signaling system has been retained in the Asterozoa but lost in the Echinozoa. Comparison of the structure of the asterozoan tachykinin-type precursors reveals that the A. rubens precursor (ArTKP) comprises two putative mature peptides, whereas the O. victoriae precursor comprises four mature peptides (Figure 7B). It remains to be determined, however, which of these two conditions represents the ancestral state in the common ancestor of the Asterozoa. Further insights into this issue may be obtained if sequence data from a variety of starfish species are analysed.

Kisspeptins (KP)

Kisspeptin (formerly known as metastin) is encoded by the KiSS1 gene in humans. KiSS1 was originally discovered as a gene that may suppress the metastatic potential of malignant melanoma cells [104]. Subsequently, it was found to play a vital role in regulating the onset of puberty. Thus, in vertebrates kisspeptin binds to its receptor GPR54 to stimulate pituitary release of gonadotropin-releasing hormone (GnRH) [105]. The first KP-type precursors to be identified in non-chordates were discovered recently in ambulacrarians - the echinoderms A. rubens and S. purpuratus and the hemichordate S. kowalevskii [8]. Accordingly, here we have identified KP-type precursors in O. victoriae and other ophiuroids. All of the ambulacrarian precursor proteins comprise two KP-type peptides and the first putative neuropeptide in the echinoderm precursors has two cysteine residues at the
N-terminus, which could form an N-terminal disulphide bridge similar that of calcitonin-type peptides (see below). In contrast, the second putative neuropeptide does not contain any cysteine residues and is typically shorter than the first peptide (Figure 7C and D). Interestingly, comparison of the sequences of the first (long) and second (short) KP-type peptides in echinoderms reveals that the long and short peptides share less sequence similarity with each other within a species than they do with respective peptides in other species (Figure 7C). This indicates that the duplication event that gave rise to the occurrence of the long and short peptides occurred before the divergence of the Asterozoa and Echinozoa. Interestingly, previous studies have revealed that there has been an expansion of KP-type receptors in ambulacraria (S. purpuratus and S. kowalevskii) and in the cephalochordate, Branchiostoma floridae, with 16 KP receptors present in the latter [6, 54]. Further studies are now needed to identify the proteins that act as receptors for the KP-type peptides identified here in ophiuroids and previously in other echinoderms [8].

Calcitonin

Calcitonin was first discovered in 1962 by Copp and Cheney [106]. The sequencing of the porcine calcitonin in 1968 revealed that this polypeptide is composed of 32 amino acids [107]. In vertebrates, calcitonin is produced by the thyroid gland [108] and regulates calcium (Ca²⁺) levels in the blood, antagonizing the effects of parathyroid hormone [109, 110]. The evolutionary antiquity of calcitonin-related peptides was first revealed with the discovery that a diuretic hormone in insects (DH₃₁) is a calcitonin-like peptide [111]. However, DH₃₁ shares modest sequence similarity with vertebrate calcitonins and lacks the N-terminal disulphide bridge that is characteristic of calcitonin-type peptides in vertebrates. More recently, it has been discovered that both DH₃₁-type and vertebrate calcitonin-type neuropeptides occur in some protostomian invertebrates, including the annelid Platynereis dumerilii and the insect Locusta migratoria [4, 112]. Hence, it is proposed that an ancestral-type calcitonin precursor gene duplicated in the common ancestor of protostomes to give rise to DH₃₁-type and calcitonin-type peptides, but with subsequent loss of calcitonin-type peptides in some protostomes. Consistent with this hypothesis, calcitonin-type precursors but not DH₃₁-type precursors have been identified in deuterostomian invertebrates, including echinoderms [8, 15, 16, 113].

An interesting feature of calcitonin/DH₃₁ precursors is the occurrence of multiple splice variants. In vertebrates, alternative splicing of the calcitonin gene results in two transcripts: one transcript encodes calcitonin and the other transcript encodes calcitonin gene-related peptide [114]. Furthermore, a complex interplay of receptors and accessory proteins...
determines the pharmacological profile of these peptides [115, 116]. Alternative splicing of DH31 and calcitonin precursors in insects has also been previously reported [112, 117, 118]. Interestingly, alternative splicing of insect calcitonin genes also generates variants that give rise to different mature peptides [112]. However, unlike the calcitonin gene, DH31 splice variants all produce an identical mature peptide [117, 118].

Our analysis of the ophiuroid transcriptomes also identified two transcript variants for calcitonin (Figure 7E and F). Based on our analysis of transcript sequences, ophiuroid calcitonin genes comprise at least three putative coding regions or ‘exons’. It is unclear if these three coding regions represent three or more exons due to the lack of genomic data, but for the sake of simplicity, we refer to them here as ‘exons’. Transcript variant 1 comprises ‘exons’ 1 and 3 but lacks ‘exon’ 2 whereas transcript variant 2 contains all 3 ‘exons’. Interestingly, ‘exons’ 2 and 3 both encode a calcitonin-type peptide. Hence, transcript variant 1 encodes a precursor that produces one calcitonin-type peptide and transcript variant 2 encodes two non-identical calcitonin-type peptides. These alternatively spliced transcripts were found in several brittle star species (Figure 8) and thus this may represent an ancient and conserved feature, although transcript variant 1 was not found in O. victoriae.

Previous studies have identified precursors comprising a single calcitonin-type peptide in the starfish A. rubens and the sea urchin S. purpuratus [8, 16], and a precursor comprising two calcitonin-type peptides in the sea cucumber A. japonicus [15]. Informed by the identification here of two transcript types in ophiuroids (transcript variant 1 and 2), we have now discovered that two transcript types also occur in A. japonicus transcriptome. Hence, alternative splicing of calcitonin-type precursor genes can be traced back in the echinoderm lineage to the common ancestor of the Asterozoa and Echinozoa, but with subsequent loss of this characteristic in some lineages.

GPA2 and GPB5

The vertebrate glycoprotein hormone family comprises luteinizing hormone (LH) follicle-stimulating hormone (FSH), chorionic gonadotropin (CG), thyroid-stimulating hormone (TSH) and the recently discovered thyrostimulin (TS) [119, 120]. Thyrostimulin is a heterodimer composed of two subunits, glycoprotein alpha 2 (GPA2) and glycoprotein beta 5 (GPB5). Orthologs of GPA2 and GPB5 have been identified and characterized in the insect Drosophila melanogaster [121] and in other invertebrates, including echinoderms [122]. Insect GPA2 and GPB5 both contain 10 conserved cysteine residues that are important in forming a heterodimeric cysteine-knot structure. Surprisingly, A. japonicus GPA2 contains only 7 cysteine residues (having lost residues 7, 8 and 9) while O. victoriae GPB5.1, A.
rubens GPB5.1 and S. purpuratus GPB5 all contain 8 cysteine residues (having lost the final two cysteine residues) (Figure S5). It is difficult to predict the structural differences that may arise in the heterodimer due to this variability in the number of cysteine residues. The possibility of GPA2 and/or GPB5 monomers or homodimers exerting their own biological functions has not been ruled out [123]. Additional investigations are needed to investigate if GPA2 and GPB5 are co-localized in echinoderms and if the monomers and dimers (both homo and hetero) exert different effects.

Uncharacterized neuropeptides

In addition to the neuropeptides discussed above, we have also identified three neuropeptide precursors that could not be classified into any known neuropeptide families. These include *O. victoriae* neuropeptide precursor (Ovnp) 18 (*O. victoriae* ortholog of Spnp18 in *S. purpuratus*) [16], Ovnp26 and Ovnp27, with the latter two identified for the first time in echinoderms. The choice of nomenclature for Ovnp26 and Ovnp27 is based on a previously used numerical nomenclature in *S. purpuratus* and/or *A. rubens*, which goes up to Arnp25 in *A. rubens*.

Ovnp18

Ovnp18 comprises four copies of a predicted mature peptide with the sequence LFWVD and the C-terminal region of the precursor (partial sequence) contains at least four cysteine residues (Figure 5F). Interestingly, this precursor type only comprises a single mature peptide in *A. rubens*, *S. purpuratus* and *A. japonicus* and the C-terminal region contains 9, 8 and 8 cysteine residues, respectively (data not shown) [8, 15, 16].

Ovnp26

Ovnp26 was identified following an analysis of *O. victoriae* transcriptome sequence using NpSearch [8]. Orthologs of Ovnp26 were identified in other brittle stars but not in other echinoderms (Figure S2-S4). Ovnp26 comprises seven copies of peptides with a conserved C-terminal GW motif, whereas orthologs in *O. aranea* and *A. filiformis* are predicted to generate eight copies of the mature peptide. Some of the mature peptides have a C-terminal SGW motif, which is similar to the C-terminus of predicted mature peptides derived from *O. victoriae* pedal peptide precursor 3 (Figure S7). However, the lack of sequence similarity in other parts of the peptide suggests that the C-terminal similarity may reflect convergence rather than homology.
Ovnp27 was identified following a HMM-based search for SIFamide-type peptides [124, 125], albeit with a high E-value. This neuropeptide precursor comprises two putative amidated mature peptides that are located immediately after the signal peptide (Figure S2-S4), as seen in SIFamide precursors [126]. The first peptide of the *O. victoriae* precursor has a C-terminal IFamide motif just like in insect SIFamides (Figure S9). However, there is no sequence similarity with SIFamides in the rest of the peptide. This coupled with the fact that SIFamide-type receptors have not been identified in echinoderms [6] suggests that the sequence similarity that peptides derived from Ovnp27-type precursors share with SIFamides may reflect convergence rather than homology.

**Neuropeptide precursors not found in brittle stars**

Our analysis of ophiuroid transcriptome sequence data did not reveal orthologs of the Spnp9 precursor from *S. purpuratus* or the Arnp21, Arnp22, Arnp23 and Arnp24 precursors from *A. rubens* [8, 16]. An Spnp9 ortholog is found in *A. japonicus* but not in *A. rubens* [15] and therefore this neuropeptide precursor type may be restricted to the Echinozoa. Orthologs of Arnp21-24 have not been found in *O. victoriae, S. purpuratus* or *A. japonicus*, which suggests that these may be Asteroidea-specific precursors.

Previous studies have shown that receptors for leucokinin, ecdysis-triggering hormone, QRFP, parathyroid, galanin/allatostatins-A and Neuromedin-U/CAPA are present in ambulacraria [6, 7, 15]. The presence of these receptors suggests that their cognate ligands should also be present in ambulacraria. However, our search approaches failed to identify any proteins in ophiuroids that resemble precursors of these neuropeptides.

**Evolutionary conservation and variation of neuropeptide copy number in the Ophiuroidea**

Many neuropeptide precursors comprise several structurally similar but non-identical bioactive peptides – i.e. the precursor protein gives rise to a neuropeptide “cocktail”. This feature of neuropeptide precursors occurs throughout metazoans. But how do these “cocktails” of neuropeptides evolve and what is their functional significance? Are the copies of mature peptides functionally redundant or do they have their own specific functions? These are important questions in neuroendocrinology for which answers remain elusive.

Evidence that neuropeptide copy number may be functionally important has been obtained from comparison of the sequences of neuropeptide precursors in twelve *Drosophila* species, the common ancestor of which dates back ~50 million years [127]. The number of peptide copies in each neuropeptide precursor was found to be identical (except for the
FMRFamide precursor) when compared between the twelve species, suggesting that stabilising selection has acted to conserve neuropeptide “cocktails” in the *Drosophila* lineage.

Here, a comparison of *O. victoriae*, *A. filiformis* and *O. aranea* neuropeptide precursors and their putative mature peptides revealed that fourteen neuropeptide precursors comprised multiple neuropeptide copies. In certain cases, the number of the mature peptides derived from a particular precursor varied across species, whereas in other cases the numbers remained constant (Figure 4). Interestingly, these three species belong to two of the three major clades of brittle stars that evolved ~270 million years ago [12]. While *O. victoriae* belongs to the Chilophiurina infraorder (clade A), *A. filiformis* and *O. aranea* belong to the Gnathophiurina infraorder (clade C). Hence, this prompted us to examine the evolution of neuropeptides and neuropeptide copy number variation at a higher level of phylogenetic resolution. To do this, we utilized a unique dataset comprising 52 ophiuroid transcriptomes. These transcriptomes were recently used as part of a phylotranscriptomic approach to reconstruct the phylogeny of ophiuroids, generating a robust phylogenetic tree that comprises three major clades [12]. Hence, this dataset allowed us to explore the evolution of neuropeptide precursors in the context of an established phylogenetic framework spanning over an unprecedented timescale of ~270 million years.

We selected for analysis neuropeptide precursors comprising more than a one putative mature neuropeptide, which include AN peptide, calcitonin, cholecystokinin 1, kisspeptin, np18, np26, np27, NG peptide, PDF, SALMFamide (L-type and F-type), tachykinin and TRH (1 and 2). Pedal peptide precursors (1, 2 and 3) were excluded from the analysis because orthology relationships between these precursors could not be established with confidence across all species (data not shown). We used *O. victoriae* representatives of these neuropeptide precursor families and the *A. filiformis* AN peptide precursor to mine 52 ophiuroid transcriptomes using BLAST. Multiple sequence alignments were generated based on the search hits (Figure S10) and the number of predicted mature peptides were compared (Figure 8). Interestingly, the number of peptides within the majority of precursors remained constant across all the species examined, which share a common ancestor estimated to date from ~270 million years ago [12].

Some studies that have investigated the physiological significance of neuropeptide “cocktails” indicate that neuropeptides derived from the same precursor protein are functionally redundant. For example, this was found for myomodulin neuropeptides in the mollusk *Aplysia californica* using the accessory radula closer muscle preparation as a bioassay [128] and for FMRFamide-related neuropeptides in *Drosophila melanogaster* when analysing effects on nerve-stimulated contraction of larval body-wall muscles [129].
However, the authors of the latter study cautiously highlighted the need to “search for additional functions or processes in which these peptides may act differentially”. Importantly, studies employing use of multiple bioassays have obtained data indicating that neuropeptides derived from a single precursor protein are not functionally redundant. For example, when the actions of fourteen structurally related neuropeptides derived from a precursor of *Mytilus* Inhibitory Peptide-related peptides in *Aplysia* were tested on three organ preparations (oesophagus, penis retractor, body wall) it was found that the rank order of potency for the peptides differed between preparations [130]. Similarly, when assaying the effects of allatostatin neuropeptides in cockroaches, tissue-specific differences in potency were observed [131]. The conservation of peptide copy number across a timescale of ~270 million years in the Ophiuroidea supports the idea that the occurrence of multiple copies of identical or structurally related neuropeptides is functionally important.

For those neuropeptide precursors that did exhibit variation in neuropeptide copy number, TRH-type precursors exhibited the highest variation, with numbers ranging from 16 to 20 copies (Figure 9). F-type SALMFamide precursors also showed variation in copy numbers (Figure 10) but loss of peptides was more frequent in F-type SALMFamide precursors than in TRH-type precursors. Furthermore, detailed analysis of sequence alignments for these precursors revealed that loss of neuropeptide copies is usually a consequence of non-synonymous mutations in codons for residues that form dibasic cleavage sites or for glycine residues that are substrates for the C-terminal amidation. This is not surprising since the C-terminal amide in smaller-sized peptides is usually important for receptor binding and activation. What is unclear at the moment is how the peptide copy number increases within a given precursor. Perhaps the increase in peptide copy number occurs as a result of unequal crossing-over during recombination [127].

The number of peptides within the F-type SALMFamide precursors appear to be clade specific. Thus, the average/median number of F-type SALMFamides in precursors from clade A is 13, clade B is 12 and clade C is 11, with a few exceptions (Figure 8). Similarly, the number of peptides within NP26-type precursors also appears to be clade specific. Hence the number of peptides is highly stable at 7 peptides within clades A and B but a high variation in peptide copy number is observed in clade C. When examining peptide copy number within clades, there are a few cases where the number of peptides within a given precursor for certain species appears to be an exception/outlier. For instance, 16 copies of the mature peptide in *Ophioplax lamellosa* TRH-1 precursor is distinctly different to the 19 copies found in other species within that clade (clade C). Likewise, *Ophiactis savignyi* only has 3 copies of kisspeptin-type peptides compared to 4 copies found in other species of that clade (Figure 8).
It could be argued that misalignments during transcriptome assembly may have influenced the number of predicted peptides found in a given precursor. However, it is unlikely that misalignments have affected the predicted sequences of neuropeptide precursors comprising multiple copies of peptides that are similar but non-identical, which applies to the majority of the precursor proteins analysed here in ophiuroids. The only exception to this are the TRH-type precursors, where the encoded peptide sequences are short and often identical, even at the nucleotide level (data not shown). Another limitation of using transcriptome data is that the sequences of neuropeptide precursors may be partial or unknown for some species and where this applies a peptide copy number is not shown in Fig. 8. An extreme example of this is the AN peptide precursor, where complete precursors sequences were only obtained from the three reference species and three other species. However, for the majority of precursor types, sequence data was obtained from a variety of species from each of the three clades of ophiuroids. For example, complete F-type SALMFamide precursor sequences were found in most of the investigated species (39 species + 3 reference species).

Conclusion

Here we report the first detailed analysis of the neuropeptide precursor complement of ophiuroids and the most comprehensive identification of echinoderm neuropeptide precursors to date. We have identified novel representatives of several bilaterian neuropeptide families in echinoderms for the first time, which include orthologs of endothelin/CCHamide, eclosion hormone, neuropeptide-F/Y and nucleobinin/nesfatin. Furthermore, analysis of precursor proteins comprising multiple copies of identical or related neuropeptides across ~270 million years of ophiuroid evolution indicates that the precise composition of neuropeptide “cocktails” is functionally important as evident from the conservation of neuropeptide copy number for multiple precursors.

Methods

Sequencing and assembly of transcriptomes

Ophiuroid transcriptomes used in this study were sequenced and assembled as reported previously [12, 20, 24].

Identification of neuropeptide precursors in ophiuroids

In order to identify neuropeptide precursors in O. victoriae, A. filiformis and O. aranea, sequences of neuropeptide precursors identified previously in other echinoderms (including the starfish, A. rubens, the sea urchin S. purpuratus and the sea cucumber, A.
were used as queries for tBLASTn analysis of a transcriptome database, using an e-value of 1000. Sequences identified as potential neuropeptide precursors by BLAST were translated using the ExPASy Translate tool (http://web.expasy.org/translate/) and then analysed for features of neuropeptide precursors. Specifically, sequences were evaluated based on 1) the presence of an N-terminal signal peptide (using Signal P v 4.1 with the sensitive cut-off of 0.34) and 2) the presence of monobasic or dibasic cleavage sites flanking the putative bioactive peptide(s).

To identify novel neuropeptide precursors or highly-divergent precursors with low sequence similarity to known precursors, we utilized two additional approaches. In the first approach, we used NpSearch [8], software that identifies putative neuropeptide precursors based on various characteristics (presence of signal peptide and dibasic cleavage sites amongst others). In the second approach, NpHMMer (http://nphmmer.sbcs.qmul.ac.uk/), a Hidden Markov Models (HMM) based software was used to identify neuropeptides not found using the above approaches.

Neuropeptide precursors identified in O. victoriae (which represented a more comprehensive neuropeptide precursor repertoire compared to A. filiformis and O. aranea) were then submitted as queries for BLAST analysis of sequence data from 52 Ophiuroidea species, using an E-value of 1e-06. BLAST hits were then further analysed using an automated ruby script (available on Github). Each BLAST hit was translated using BioRuby and the open reading frame (ORF) containing the BLAST high-scoring segment pair was extracted. These ORFs were then examined for the presence of a signal peptide using Signal P 4.1 using a sensitive cut-off of 0.34. All sequences were then aligned using MAFFT, with the number of maximum iterations set to 1000 to ensure an optimal alignment. These alignments were then further optimized by manually adjusting the location of the bioactive peptide and cleavage sites. Finally, the alignments were annotated using different colours for the signal peptide (blue), the bioactive peptide(s) (red) and cleavage sites (green).

**Phylogenetic and clustering analyses of sequence data**

Phylogenetic analysis of membrane guanylyl cyclase receptors and nucleobindins was performed using maximum likelihood and Bayesian methods. Prior to these analyses, corresponding multiple alignments were trimmed using BMGE [132] with the following options: BLOSUM30, max –h = 1, –b = 1, as described previously [10, 91]. The maximum likelihood method was implemented in the PhyML program (v3.1/3.0 aLRT). The WAG substitution model was selected assuming an estimated proportion of invariant sites (of 0.112) and 4 gamma-distributed rate categories to account for rate heterogeneity across sites.
The gamma shape parameter was estimated directly from the data. Reliability for internal branch was assessed using the bootstrapping method (500 bootstrap replicates). The Bayesian inference method was implemented in the MrBayes program (v3.2.3). The number of substitution types was fixed to 6. The poisson model was used for substitution, while rates variation across sites was fixed to "invgamma". Four Markov Chain Monte Carlo (MCMC) chains were run for 100000 generations, sampling every 100 generations, with the first 500 sampled trees discarded as "burn-in". Finally, a 50% majority rule consensus tree was constructed.

CLANS analysis was performed on echinoderm EH-like, arthropod EH, arthropod ITP and vertebrates ANP precursors based on all-against-all sequence similarity (BLAST searches) using BLOSUM 45 matrix (https://toolkit.tuebingen.mpg.de/clans/) [39] and the significant high-scoring segment pairs (HSPs). Neuropeptide precursors were clustered in a three-dimensional graph represented here in two dimensions.

**Competing Interests**
The authors declare that no competing interests exist.

**Author contributions**

**Acknowledgements**
The authors would like to acknowledge Zuraiha Waffa and Giulia Oluwabunmi Olayemi for their assistance with sequence alignments. This work was supported by Leverhulme Trust grant (RPG-2013-351) and a BBSRC grant (BB/M001644/1) awarded to M.R.E. L.A.Y.G is supported by a PhD studentship awarded by CONACYT.
References


5. Veenstra JA: Neuropeptide Evolution: Chelicerate Neurohormone and Neuropeptide Genes may reflect one or more whole genome duplications. Gen Comp Endocrinol 2016.


73. Johnson LP, Magee DF: Inhibition of Gastric Motility by a Commercial Duodenal Mucosal Extract Containing Cholecystokinin and Pancreozymin. Nature 1965, 207(5004):1401-&.


112. Veenstra JA: The contribution of the genomes of a termite and a locust to our understanding of insect neuropeptides and neurohormones. *Front Physiol* 2014, 5:454.


**Figure captions**

**Figure 1**: Bilaterian animal phylogeny. The diagram shows i) the phylogenetic position of the phylum Echinodermata in the ambulacrarian clade of the deuterostomes and ii) relationships between the five extant classes of echinoderms, which include the focal class for this study – the Ophiuroidea (e.g. *Ophionotus victoriae*).

**Figure 2**: Eclosion hormone (EH)-type peptides and receptors in echinoderms A) Partial multiple sequence alignment of eclosion hormone-type precursor sequences, excluding the N-terminal signal peptide; B) Cluster analysis of arthropod EH precursors, echinoderm EH-like precursors, arthropod ion transport peptides (ITPs) and vertebrate atrial natriuretic peptides shows that echinoderm EH-like precursors are more closely related to arthropod EH than ITP C) Phylogenetic analysis of membrane guanylate cyclase receptors shows that EH-like receptors are found in echinoderms but are absent in vertebrates as seen for the EH-like precursors. Species names: *Ophionotus victoriae* (Ovic), *Asterias rubens* (Arub), *Strongylocentrotus purpuratus* (Spur), *Drosophila melanogaster* (Dmel), *Bombyx mori* (Bmor) and *Pediculus humanus corporis* (Pcor).

**Figure 3**: Multiple sequence alignments of A) CCHamide-type and B) Neuropeptide-F/Y-type peptides. Species names: *Ophionotus victoriae* (Ovic), *Asterias rubens* (Arub), *Apostichopus japonicus* (Ajap), *Drosophila melanogaster* (Dmel), *Apis mellifera* (Amel), *Lottia gigantea* (Lgig), *Aplysia californica* (Acal), *Homo sapiens* (Hsap), *Ophiopsila aranea* (Oara), *Amphiura filiformis* (Afil), *Patiria miniata* (Pmin), *Saccoglossus kowalevskii* (Skow), *Branchiostoma floridae* (Bflo) and *Daphnia pulex* (Dpul).

**Figure 4**: Summary of neuropeptide precursors identified in *Ophionotus victoriae*, *Amphiura filiformis* and *Ophiopsila aranea*. Neuropeptide precursors are classified based on the type of G-protein coupled receptor (GPCR) their constituent peptides are predicted to activate (see Mirabeau and Joly, 2013). Some peptides bind to receptors other than GPCRs and these are grouped with peptides where the receptor is unknown. Ophiuroids have neuropeptide precursors from up to 32 families. The number of putative mature peptides derived from each precursor has been indicated along with the presence of amidation and pyroglutamation.

**Figure 5**: Multiple sequence alignments of mature peptides belonging to selected neuropeptide families. A) corazonin alignment; B) gonadotropin-releasing hormone (GnRH) alignment; C) orexin alignment; D) luqin alignment; E) vasopressin/oxytocin (VP/OT)
alignment; F) Ovn18 alignment; G) melanin-concentrating hormone (MCH) alignment; H) NP peptide alignment; I) pigment dispersing factor (PDF) alignment. Species names: *Ophionotus victoriae* (Ovic), *Asterias rubens* (Arub), *Strongylocentrotus purpuratus* (Spur), *Apostichopus japonicus* (Ajap), *Saccoglossus kowalevskii* (Skow), *Branchiostoma floridae* (Bflo), *Anopheles gambiae* (Agam), *Daphnia pulex* (Dpul), *Strigamia maritima* (Smar), *Lottia gigantea* (Lgig) and *Homo sapiens* (Hsap).

**Figure 6:** Alignments of neuropeptides derived from precursors that exist in multiple forms in ophiuroids. A) thyrotropin-releasing hormone (TRH) alignment; B) cholecystokinin alignment; C) somatostatin alignment; D) corticotropin-releasing hormone (CRH) alignment. Species names: *Ophionotus victoriae* (Ovic), *Asterias rubens* (Arub), *Strongylocentrotus purpuratus* (Spur), *Apostichopus japonicus* (Ajap), *Branchiostoma floridae* (Bflo), *Homo sapiens* (Hsap), *Drosophila melanogaster* (Dmel) and *Lottia gigantea* (Lgig).

**Figure 7:** Comparative analysis of ophiuroid tachykinin, kisspeptin and calcitonin-type precursors and neuropeptides. A) Alignment of tachykinin-type peptides in *O. victoriae* (Ophiuroidea) and *A. rubens* (Asteroidea); B) Schematic diagrams of the *O. victoriae* and *A. rubens* tachykinin precursors showing the location of the signal peptide (SP) and predicted neuropeptides (labelled 1 to 4); C) Alignments of the long and short forms of kisspeptin-type neuropeptides in *O. victoriae*, *A. rubens* and *S. purpuratus* (Echinoidea) D) Schematic diagrams of the *O. victoriae* and *A. rubens* kisspeptin precursors showing the locations of the SP, short and long orthocopies and cysteine (C) residues; E) Alignment of calcitonin-type peptides from *O. victoriae*, *A. rubens*, *S. purpuratus* and *A. japonicus* (Holothuroidea); F) Predicted alternative splicing of the calcitonin gene in ophiuroids, with the location of the SP and neuropeptides (CT1 and CT2) labelled. Species names: *Ophionotus victoriae* (Ovic), *Asterias rubens* (Arub), *Strongylocentrotus purpuratus* (Spur) and *Apostichopus japonicus* (Ajap).

**Figure 8:** Comparison of neuropeptide copy numbers across the Ophiuroidea for precursors comprising multiple copies of neuropeptides. Neuropeptide precursors were mined from 52 ophiuroid transcriptomes, with the phylogeny adapted from O’Hara et al. (2014) [12]. Am_laud: *Amphiophiura laudata*, Am_spat: *Amphiophiura spatulifera*, Am_cipu: *Amphioplus cipus*, Am_cten: *Amphioplus ctenacantha*, Am_squa: *Amphipholis squamata*, Am_cons1: *Amphiura constricta* 1, Am_cons2: *Amphiura constricta* 2, As_love: *Asteronyx loveni*, As_bidw: *Asteroschema bidwillae*, As_tubi: *Asteroschema tubiferum*, Ba_hero: *Bathymetra heroica*. 

Figure 9: A partial multiple sequence alignment of ophiuroid thyrotropin-releasing hormone (TRH) precursors showing clade-specific gain/loss of neuropeptide copies. Mono- and dibasic cleavage sites are highlighted in green, mature peptides in red with the glycine residue for amidation in pink. Species have been grouped and coloured (clade A in purple, clade B in blue and clade C in orange) based on the phylogeny determined by O’Hara et al. (2014) [12].

Figure 10: A partial multiple sequence alignment of ophiuroid F-type SALMFamide precursors showing clade-specific gain/loss of neuropeptide copies. Di-basic cleavage sites are highlighted in green, mature peptides in red with the glycine residue for amidation in pink. Species have been grouped and coloured (clade A in purple, clade B in blue and clade C in orange) based on the phylogeny determined by O’Hara et al. (2014) [12].

Supplementary files

Figure S1: Alignment and phylogenetic analysis of nucleobinds (NUCB). A) Partial sequence alignment (excludes the signal peptide) of NUCB precursors. The locations of Homo sapiens nesfatin-1, 2 and 3 are indicated. A dibasic cleavage site in O. victoriae nesfatin-1 is marked in red. B) Phylogenetic analysis of NUCB precursors. Species names: Ophionotus victoriae (Ovic), Amphiura filiformis (Afil), Ophiopsila aranea (Oara),...
Apostichopus japonicus (Ajap), Strongylocentrotus purpuratus (Spur), Homo sapiens (Hsap),
Mus musculus (Mmus) and Drosophila melanogaster (Dmel).

**Figure S2:** Ophionotus victoriae neuropeptide precursor repertoire.

**Figure S3:** Amphiura filiformis neuropeptide precursor repertoire.

**Figure S4:** Ophiopsila aranea neuropeptide precursor repertoire.

**Figure S5:** Partial multiple sequence alignments of echinoderm representatives of A) glycoprotein alpha 2 (GPA2)-type subunits and B) glycoprotein beta 5 (GPB5)-type subunits. Species names: Ophionotus victoriae (Ovic), Asterias rubens (Arub), Strongylocentrotus purpuratus (Spur) and Apostichopus japonicus (Ajap).

**Figure S6:** Partial multiple sequence alignments of echinoderm representatives of large protein hormones. A) insulin/insulin-like growth factor; B) relaxin-like peptide; C) bursicon (bursicon alpha); D) partner of bursicon (bursicon beta). Species names: Ophionotus victoriae (Ovic), Asterias rubens (Arub), Strongylocentrotus purpuratus (Spur) and Apostichopus japonicus (Ajap).

**Figure S7:** Multiple sequence alignment of echinoderm pedal peptides. Species names: Ophionotus victoriae (Ovic), Asterias rubens (Arub), Strongylocentrotus purpuratus (Spur) and Apostichopus japonicus (Ajap).

**Figure S8:** Multiple sequence alignments of echinoderm neuropeptide families. A) F-type SALMFamide alignment; B) L-type SALMFamide alignment; C) AN peptide. Species names: Ophionotus victoriae (Ovic), Asterias rubens (Arub), Strongylocentrotus purpuratus (Spur) and Apostichopus japonicus (Ajap).

**Figure S9:** Multiple sequence alignment of predicted peptides derived from neuropeptide precursor 27 in Ophionotus victoriae (Ovic), Amphiura filiformis (Afil), Ophiopsila aranea (Oara) and Apostichopus japonicus (Ajap).

**Figure S10:** Multiple sequence alignments of neuropeptide precursors used to generate Figure 8.
<table>
<thead>
<tr>
<th>Receptor type</th>
<th>Neuropeptide family</th>
<th>O. victoriae</th>
<th>A. filiformis</th>
<th>O. aranea</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>CCHamide-like 1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>CCHamide-like 2</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>Corazonin</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>Gonadotropin-releasing hormone</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>LeuK</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>Neuropeptide-F/Y 1</td>
<td>1*</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>Neuropeptide-F/Y 2</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>NG peptide / Neuropeptide-S</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>9</td>
<td>Orexin 1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>Thyrotropin-releasing hormone 1</td>
<td>21</td>
<td>14*</td>
<td>17</td>
</tr>
<tr>
<td>11</td>
<td>Vasopressin / Oxytocin</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>12</td>
<td>Kisspeptin</td>
<td>2</td>
<td>1*</td>
<td>1</td>
</tr>
<tr>
<td>13</td>
<td>Melanin-concentrating hormone</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>14</td>
<td>Somatostatin 1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>15</td>
<td>Somatostatin 2</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>16</td>
<td>Bursicon alpha</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>17</td>
<td>Bursicon beta</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>18</td>
<td>Glycoprotein hormone alpha 2.1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>19</td>
<td>Glycoprotein hormone alpha 2.2</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>20</td>
<td>Glycoprotein hormone beta 5.1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>21</td>
<td>Calcitonin</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>22</td>
<td>Pigment-dispersing factor</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>23</td>
<td>AN peptide</td>
<td>2</td>
<td>1/2</td>
<td>2/2</td>
</tr>
<tr>
<td>24</td>
<td>Eclision hormone 1.1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>25</td>
<td>Eclision hormone 1.2</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>26</td>
<td>Eclision hormone 2.1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>27</td>
<td>Eclision hormone 2.2</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>28</td>
<td>Insulin-like peptide</td>
<td>a</td>
<td>a</td>
<td>a</td>
</tr>
<tr>
<td>29</td>
<td>Nucleobindin / Nefastin</td>
<td>b</td>
<td>b</td>
<td>b</td>
</tr>
<tr>
<td>30</td>
<td>Pedal peptide 1</td>
<td>6</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>31</td>
<td>Pedal peptide 2</td>
<td>7</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>32</td>
<td>Pedal peptide 3</td>
<td>8*</td>
<td>1*</td>
<td>1</td>
</tr>
<tr>
<td>33</td>
<td>SALMiforme (L-type)</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>34</td>
<td>SALMiforme (F-type)</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>35</td>
<td>Neuropeptide precursor 18</td>
<td>4</td>
<td>2*</td>
<td>4</td>
</tr>
<tr>
<td>36</td>
<td>Neuropeptide precursor 26</td>
<td>7</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>37</td>
<td>Neuropeptide precursor 27</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

- **Present**: Neuropeptide is present in all three species.
- **Partial / some mature peptides**: Neuropeptide is present in one or two species with some mature peptides.
- **Absent**: Neuropeptide is absent in all three species.
- **Cannot be determined**: Neuropeptide cannot be determined.

a: Heterodimer of A-chain and B-chain
b: Number of mature peptides unknown
c: Multiple partial precursors
<table>
<thead>
<tr>
<th>Protein</th>
<th>Query Sequence</th>
<th>Matching Sites</th>
</tr>
</thead>
</table>
| Am_clup | PASGSASFGTRKAGQNMLHAFSRK-GMDPSGLSAFSTKR-GEDSALSAFSTKR | NFP-SSDNNEAKDE-GEYPMO-AFASK-RP-DFTGLSAFSTKR-KRPLSGLSAFSTKR-
| Op_filo | P-GMDPSGLSAFSTKR-GEDSALSAFSTKR | NFP-SSDNNEAKDE-GEYPMO-AFASK-RP-DFTGLSAFSTKR-KRPLSGLSAFSTKR-
| Mi_grac | P-GMDPSGLSAFSTKR-GEDSALSAFSTKR | NFP-SSDNNEAKDE-GEYPMO-AFASK-RP-DFTGLSAFSTKR-KRPLSGLSAFSTKR-
| Am_squa | P-GMDPSGLSAFSTKR-GEDSALSAFSTKR | NFP-SSDNNEAKDE-GEYPMO-AFASK-RP-DFTGLSAFSTKR-KRPLSGLSAFSTKR-
| Op_resi | P-GMDPSGLSAFSTKR-GEDSALSAFSTKR | NFP-SSDNNEAKDE-GEYPMO-AFASK-RP-DFTGLSAFSTKR-KRPLSGLSAFSTKR-
| Op_abys | P-GMDPSGLSAFSTKR-GEDSALSAFSTKR | NFP-SSDNNEAKDE-GEYPMO-AFASK-RP-DFTGLSAFSTKR-KRPLSGLSAFSTKR-
| Op_angu | P-GMDPSGLSAFSTKR-GEDSALSAFSTKR | NFP-SSDNNEAKDE-GEYPMO-AFASK-RP-DFTGLSAFSTKR-KRPLSGLSAFSTKR-
| Op_scha | P-GMDPSGLSAFSTKR-GEDSALSAFSTKR | NFP-SSDNNEAKDE-GEYPMO-AFASK-RP-DFTGLSAFSTKR-KRPLSGLSAFSTKR-
| Op_lame | P-GMDPSGLSAFSTKR-GEDSALSAFSTKR | NFP-SSDNNEAKDE-GEYPMO-AFASK-RP-DFTGLSAFSTKR-KRPLSGLSAFSTKR-
| Op_bisp | P-GMDPSGLSAFSTKR-GEDSALSAFSTKR | NFP-SSDNNEAKDE-GEYPMO-AFASK-RP-DFTGLSAFSTKR-KRPLSGLSAFSTKR-
| Cl_cana | P-GMDPSGLSAFSTKR-GEDSALSAFSTKR | NFP-SSDNNEAKDE-GEYPMO-AFASK-RP-DFTGLSAFSTKR-KRPLSGLSAFSTKR-
| Op_exim_1 | P-GMDPSGLSAFSTKR-GEDSALSAFSTKR | NFP-SSDNNEAKDE-GEYPMO-AFASK-RP-DFTGLSAFSTKR-KRPLSGLSAFSTKR-
| Op_lloid | P-GMDPSGLSAFSTKR-GEDSALSAFSTKR | NFP-SSDNNEAKDE-GEYPMO-AFASK-RP-DFTGLSAFSTKR-KRPLSGLSAFSTKR-
| Op_prol | P-GMDPSGLSAFSTKR-GEDSALSAFSTKR | NFP-SSDNNEAKDE-GEYPMO-AFASK-RP-DFTGLSAFSTKR-KRPLSGLSAFSTKR-
| As_tubi | P-GMDPSGLSAFSTKR-GEDSALSAFSTKR | NFP-SSDNNEAKDE-GEYPMO-AFASK-RP-DFTGLSAFSTKR-KRPLSGLSAFSTKR-
| Op_brev | P-GMDPSGLSAFSTKR-GEDSALSAFSTKR | NFP-SSDNNEAKDE-GEYPMO-AFASK-RP-DFTGLSAFSTKR-KRPLSGLSAFSTKR-
| Ba_hero | P-GMDPSGLSAFSTKR-GEDSALSAFSTKR | NFP-SSDNNEAKDE-GEYPMO-AFASK-RP-DFTGLSAFSTKR-KRPLSGLSAFSTKR-
| Op_appr | P-GMDPSGLSAFSTKR-GEDSALSAFSTKR | NFP-SSDNNEAKDE-GEYPMO-AFASK-RP-DFTGLSAFSTKR-KRPLSGLSAFSTKR-
| Op_vivi | P-GMDPSGLSAFSTKR-GEDSALSAFSTKR | NFP-SSDNNEAKDE-GEYPMO-AFASK-RP-DFTGLSAFSTKR-KRPLSGLSAFSTKR-
| Op_wend | P-GMDPSGLSAFSTKR-GEDSALSAFSTKR | NFP-SSDNNEAKDE-GEYPMO-AFASK-RP-DFTGLSAFSTKR-KRPLSGLSAFSTKR-
| Op_plic | P-GMDPSGLSAFSTKR-GEDSALSAFSTKR | NFP-SSDNNEAKDE-GEYPMO-AFASK-RP-DFTGLSAFSTKR-KRPLSGLSAFSTKR-
| Op_perf | P-GMDPSGLSAFSTKR-GEDSALSAFSTKR | NFP-SSDNNEAKDE-GEYPMO-AFASK-RP-DFTGLSAFSTKR-KRPLSGLSAFSTKR-
| Po_exim_1 | P-GMDPSGLSAFSTKR-GEDSALSAFSTKR | NFP-SSDNNEAKDE-GEYPMO-AFASK-RP-DFTGLSAFSTKR-KRPLSGLSAFSTKR-
| As_love | P-GMDPSGLSAFSTKR-GEDSALSAFSTKR | NFP-SSDNNEAKDE-GEYPMO-AFASK-RP-DFTGLSAFSTKR-KRPLSGLSAFSTKR-
| Op_prol | P-GMDPSGLSAFSTKR-GEDSALSAFSTKR | NFP-SSDNNEAKDE-GEYPMO-AFASK-RP-DFTGLSAFSTKR-KRPLSGLSAFSTKR-
| Op_lyma | P-GMDPSGLSAFSTKR-GEDSALSAFSTKR | NFP-SSDNNEAKDE-GEYPMO-AFASK-RP-DFTGLSAFSTKR-KRPLSGLSAFSTKR-