General and Comparative Endocrinology 278 (2019) 3–11

Contents lists available at ScienceDirect

General and Comparative Endocrinology

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

Research paper

Two Lys-vasopressin-like peptides, EFLamide, and other phasmid neuropeptides

Jan A. Veenstra

INCIA, UMR 5287 CNRS, Université de Bordeaux, allée Geoffroy St Hillaire, CS 50023, 33 615 Pessac Cedex, France

A R T I C L E I N F O

Keywords:
Vasopressin
EFLamide
Leucokinin
Sulfakinin
Gene amplification
Gene loss
Tryptopyrokinin

A B S T R A C T

Phasmid neuropeptide genes were identified in the genomes of two phasmids, *Timema cristinae* and *Clitarchus hookeri*. The two species belong to two sisters groups, the Timematodea and Euphasmatodea respectively. Neuropeptide genes were identified using the BLAST+ program on the genome assemblies and the absence of some neuropeptides was confirmed by the concomitant absence of their G-protein coupled receptors. Both genomes were assembled using short reads and the average coverage of the genome is more than 166 times for both species. This makes it virtually impossible that there would not be a single short read for at least one of the conserved transmembrane regions of a GPCR coded by such a genome. Hence, when not a single read can be found for a specific GPCR, it can be concluded that the particular gene is absent from that species.

Most previously identified insect neuropeptides are used by these two species. Of the three arthropod allatostatin C related peptides, only allatostatins CC and CCC are present. Both species lack leucokinin, while sulfakinin and dilp8 signaling is absent from *Clitarchus*, but present in *Timema*. Interestingly, whereas *Timema* has lost a vasopressin-related peptide, the gene coding such a peptide is amplified in the *Clitarchus* genome. Furthermore, while *Clitarchus* has a specific tryptopyrokinin gene, *Timema* does not and in this species tryptopyrokinin is coded only by the pyrokinin and periviscerokinin genes. Finally, both species have genes coding EFLamide and its GPCR; in phasmids these genes codes for one (*Clitarchus*) or two (*Timema*) EFLamide paracopies.

1. Introduction

Neuropeptides and neurohormones regulate a variety of processes and are often considered the master regulators of physiology. It is reasonable to expect that as the physiology of a species evolves so should its regulation. Insects have adapted to a large variety of different habitats, at least in part by adapting their physiology. This suggests that insects may be an interesting group to study how neuropeptide functions change during evolution. It is remarkable that the very successful holometabolous insect species have lost a significant number of neuropeptide signaling pathways. Thus flies *e.g.*, have lost genes coding allatotropin, calcitonin, a vasopressin-like peptide, elevenin, allatostatin CCC, ACP, baratin and NPF2, while *Drosophila melanogaster* also lacks neuroparsin and the recently discovered agatoxin-like neuropeptide (Sturm et al., 2016) that are present in other flies. Yet, there is little doubt that flies have been very successful from an evolutionary point of view.

We know more about neuropeptides from holometabolous species than from the more basal insect groups, where knowledge is often limited to that obtained from transcriptomes, that do not necessarily reveal all neuropeptide genes of a species; the absence of expression of a neuropeptide gene in a transcriptome does not show that particular neuropeptide to have been lost. Genomes of Polyneoptera are often, but not always, quite big. Thus the genome from the migratory locust is the largest animal genome sequenced to date (Wang et al., 2014). Its neuropeptidome and that from another polyneopteran species with a sequenced genome, the termite *Zootermopsis nevadensis,* has been annotated (Veenstra, 2014). Both these species have a large set of neuropeptides that is remarkably similar to those found in decapod crustaceans (Veenstra, 2016b). However, both these species lack genes coding EFLamide, a peptide that is present in several arthropods, including non-pterygote hexapods (Derst et al., 2016). A receptor for EFLGamide, an analog of this peptide, has recently been deorphanized in *Platynereis dumerilii* and was shown to be an ortholog of the vertebrate TRH receptor, revealing that the protostomian ligands for these receptors are TRH homologs (Bauknecht and Jékely, 2015). The neuropeptide GPCR A45 from *Nilaparvata lugens* (Tanaka et al., 2014) appears to be an ortholog of this gene (Veenstra, 2016a), suggesting that at least some insect species may have this neuropeptide gene. Indeed a recent paper describes a partial gene for such a neuropeptide in the bed

---

E-mail address: jan.veenstra@u-bordeaux.fr.

https://doi.org/10.1016/j.ygcen.2018.04.027
Received 1 February 2018; Received in revised form 23 April 2018; Accepted 25 April 2018

bug, *Cimex lectularius* (Predel et al., 2018). Thus are there good reasons to look at the neuropeptides of other Polyneopteran insect orders, such as the phasmids.

Timema cristinae is a stick insect from the only genus of the suborder Timematodea, the large majority of phasmids belong to the Euphasmatodea. The *Timema* species are native to the Western USA and *T. cristinae* in particular shows interesting color polymorphism. The elucidation of the effects of ecological selection on genome variation in this species has led to construction of a draft genome assembly (Rüdiger et al., 2018). *Clitarchus hookeri* is a stick insect native to New Zealand and belongs to the Euphasmatodea. Its genome assembly (Wu et al., 2017) of 3.8 Gb is almost four times as big as the draft genome from *Timema*. There are also several phasmid transcriptomes, all belonging to the Euphasmatodea, some of those from New Zealand species closely related to *C. hookeri,* but there are also others (*e.g.* Dennis et al., 2015a,b; Shelomi et al., 2014; Shelomi, 2017).

The availability of these two genome assemblies from two quite different phasmid species in combination with a significant number of transcriptomes provides an excellent opportunity to probe phasmid neuropeptidomes.

2. Materials and methods

2.1. DNA sequences

The genome assembly for *C. hookeri* and *T. crisitinae* draft 0.3 were downloaded from respectively NCBI and the Nosil lab web site (http://nosil-lab.group.shef.ac.uk/?page_id=25). In order to be able to confirm the absence of certain genes, short read archives (SRAs) containing the original data used to produce these assemblies were also downloaded from NCBI. These are the following genomic SRAs: for *C. hookeri*: SRR5889584, SRR5889585, SRR5889586, SRR5889587, SRR5889588, SRR5889589, SRR5889590, SRR5889592, SRR5889593, SRR5892339, SRR5892340, SRR5892836 and SRR5892837 and for *T. cristinae*: SRR5192505, SRR5192506, SRR5192507, SRR5192508, SRR5192509 and SRR5192510. The single genomic SRA for *Peruphasma schultei* (SRR1182314) was used to find a partial exon of the EFLamide gene. SRAs were analyzed using the sra toolkit from NCBI (https://www.ncbi.nlm.nih.gov/sra/docs/toolkitsoft/). Individual reads identified from SRAs were assembled with Trinity into larger contigs (Grabherr et al., 2011). For expression analysis trinity transcript assemblies were downloaded from the Dryad digital repository (http://datadryad.org/resource/http://dx.doi.org/10.5061/dryad.kc826) for the following species: *Micrarchus* sp., *Tectarchus salebrosus*, *T. ovobessus, Niveaphasma annulata*, *Asteliaphasma jucundum*, *Argosarchus horridus*, *Spinotectarchus acornutus*, *Acanthoxyla* sp. and *Clitarchus hookeri* (Dunning et al., 2013; Wu et al., 2016). Other phasmid transcriptomes (Shelomi et al., 2014; Shelomi, 2017) were analyzed directly at the NCBI website. For some analyses the original raw transcriptome sequence data were analyzed, to this end the following SRAs were used: SRR921632, SRR1003263, SRR1003264, SRR1003265, SRR1003267, SRR1003269, SRR1002984, SRR1054191, SRR1054192, SRR1054193, SRR2089870, SRR2089871, SRR2089872, SRR2089873, SRR2089874, SRR2089875, SRR2089876, SRR2089877, SRR2089878, SRR2089879, SRR2089880, SRR2089881, SRR2089882, SRR2089883, SRR2089884, SRR2089885, SRR2089886, SRR2089887, SRR2089888, SRR2089889, SRR2089890, SRR2089891, SRR2089892, SRR2089893, SRR2089894, SRR2089895, SRR2089896, SRR2089897, SRR2089898, SRR2089899, SRR2089900, SRR2089901, SRR2089902, SRR2089903, SRR2089904, SRR2089905, SRR2089906, SRR2089907, SRR2089908, SRR2089909, SRR2089910, SRR2089911, SRR2089912, SRR2089913, SRR2089914, SRR2089915, SRR2089916, SRR2089917, SRR2089918, SRR2089919, SRR2089920, SRR2089921, SRR2089922, SRR2089923, SRR2089924, SRR2089925, SRR2089926, SRR6472766, SRR6472767, SRR6472768, SRR6472769, SRR6472770, SRR6472771, SRR6472772, SRR6472773, SRR6472774, SRR6472775, SRR6472776, SRR6472777, SRR6472778, SRR6472779, SRR6472780 and SRR921650.

On discovery of partial EFLamide genes, it seemed of interest to see how widespread this gene is in insects. Genomes for the following species were searched for such a gene: *Orchesella cincta* (Faddeeva-Vakhrusheva et al., 2017) and *Folsomia candida* (Faddeeva-Vakhrusheva et al., 2016), *Locusta migratoria* (Wang et al., 2014), *Bemisia tabaci* (Chen et al., 2016), *Diaphorina citri* (Hunter and Reese, 2014), *Nilaparvata lugens* (Xue et al., 2014), *Oncopeltus fasciatus* (Murali et al., 2015), *Rhodnius prolixus* (Mesquita et al., 2015) and the water strider *Gerris buenoi* (Armisén et al., 2018). Furthermore, the as yet unpublished genomes from the dragonfly *Ladona fulva*, the mayfly *Ephemera danica*, the stink bug *Halyomorpha halys,* the leafhopper *Homalodisca vitripennis* and the psyllid *Pachypsylla venusta* were also used. The latter genomes were produced by the Baylor College of Medicine Human Genome Sequencing Center (https://www.hgsc.bcm.edu).

The DNA genome and transcriptome assemblies were made into BLAST+ databases using the BLAST+ program (Camacho et al., 2009) and searched for neuropeptide sequences using the tblastn command of the same program with the published *Locusta* and *Zootermopsis* neuropeptide precursor and/or predicted active neuropeptide sequences (Veenstra, 2014; Hou et al., 2015). Scaffolds containing putative neuropeptide genes were then extracted with the blastdbcmd command and analyzed using Artemis (Rutherford et al, 2000). Putative intron splice sites were predicted manually aided by the locations of known splice sites in the orthologous genes from *Locusta* and/or *Zootermopsis* and, where possible, with transcript data from either the same species, or other phasmids.

Predicted protein sequences were checked for the likely presence of signal peptides using Signal P 4.0 (Petersen et al., 2011) and occasionally Signal P 3.0 as well (Bendtsen et al., 2004), while convertase cleavage sites were identified using previously described rules (Veenstra, 2000).

2.2. Absence of gene in a genome or transcripts in a transcriptome.

Sometimes it is difficult to find a particular neuropeptide gene in a genome because the gene, or part of it, did not get incorporated into the genome. This has happened a few times with the *Timema* draft genome. In those cases an attempt was made to find individual genomic reads in the various genome SRAs from this species and Trinity was used to assemble those reads into a genomic contig. In other cases it may be difficult to detect a gene in a genome and/or its expression in a transcriptome because its primary structure has changed substantially from the consensus sequence making it very difficult to recognize with the BLAST program; this was *e.g.* the case for allatotropin in hymenoptera. The core sequences of some neuropeptides, like leucokinin or EFLamide, are both small and variable. If phasmids were to have a leucokinin gene coding for a single leucokinin copy of a slightly evolved structure, it might have been very difficult to detect in genomes as large as these. When neither evidence for the receptor of a specific neuropeptide could be found, it was concluded that the neuropeptide signaling pathway in question had been lost from a species.

Genes encoding neuropeptide GPCRs in these species have large introns and hence span very long stretches of DNA. It thus seems virtually impossible that there would not be a single genomic short read corresponding to one of the leucokinin GPCR coding exons when both genomes are covered on average 166 times. To obtain convincing evidence for the absence of a particular GPCR from one of these two genomes the tblastn_vdb command from the sra toolkit was used to search all the genome SRAs available for orthologs of each of the individual exons of the orthologous GPCR from the species most closely related. The translated protein sequences of the best matches were then analyzed using the blastp command from the BLAST+ program against a protein GPCR protein database from *Z. nevadensis,* the insect species most closely related to phasmids with a curated set of neuropeptide GPCRs (Veenstra, 2014), to which *Nilaparvata* GPCR A45 (Tanaka et al., 2014), the *Clitarchus* vasopressin and the *Timema* sulfakinin GPCRs had been added. This procedure yields large numbers of matches to the query, and when the query includes a transmembrane region (well conserved sequences of typically about 20 amino acid residues, significantly smaller than the 33 amino acids that are potentially encoded by a short read of 100 nucleotides), the best are unambiguously identified as being part of neuropeptide GPCRs. For example, when looking for the leucokinin receptor in phasmid genomic SRAs, the best matches to the query searches correspond to transmembrane regions of the RYamide GPCR, while the best *Timema* matches for a vasopressin GPCR are for its CCAP GPCR. The same protocol was also used for the expression of the leucokinin receptor in four SRAs from the Malpighian tubules from the phasmid *Carausius morosus* (SRR3211821, SRR3211822, SRR3211823 and SRR3211824; Shelomi, 2017). In the beetle *Tribolium castaneum,* in which allatostatin CCC is lacking, allatostatin C and CC act on the same receptor (Audsley et al., 2013), while in the honeybee, that lacks allatostatin C, both allatostatin CC and CCC act on the same receptor (Urlacher et al., 2016). Thus, for insects there is evidence suggesting that allatostatins C, CC and CCC all act on the same receptor. However, this is not necessarily true for other arthropods, as shown recently for the lobster *Homarus americanus* (Dickinson et al., 2018). It is therefore not possible to use the absence of its likely receptor to confirm the absence of this particular neuropeptide. The allatostatin C, CC and CCC are very well conserved in arthropods (Veenstra, 2016c) but when using various allatostatin C amino acid sequences as queries to search the two genome assemblies and the various SRAs only allatostatins CC and CCC were found.

3. Results and discussion

Two phasmid genome assemblies in combination with several transcriptomes were used to obtain the neuropeptidomes of two stick insects that are not closely related. The genome assemblies analyzed here are relatively large, *Timema*, or even very large, *Clitarchus*, and have been produced using short read illumina sequences. Assemblers using short read sequences have difficulty with polymorphisms and short repeated sequences, such as are typically present in neuropeptide precursors encoding a number of very similar sequences (Richards and Murali, 2015). Thus DNA sequences coding precursors for neuropeptides such as allatostatin B, FMRFamide, and tryptopyrokinin have often gaps and/or may lead to more peptide paracopies in the assembled gene then there are in reality. Genome assemblies produced from small reads also seem to have a stronger tendency to have duplicates of the same gene incorporated into the assemblies. These copies may represent different alleles or may be due to incorrect recombination of chromosome fragments by the assembler (*e.g.* if there are two SNPs separated by more than the short read length, the assembler could then produce four different combinations). Thus, the presence of multiple copies of the same gene in a genome assembly does not necessarily indicate the existence of two genes in the animal while the existence of just two, could be due to the presence of two different alleles.

3.1. Phasmid neuropeptides

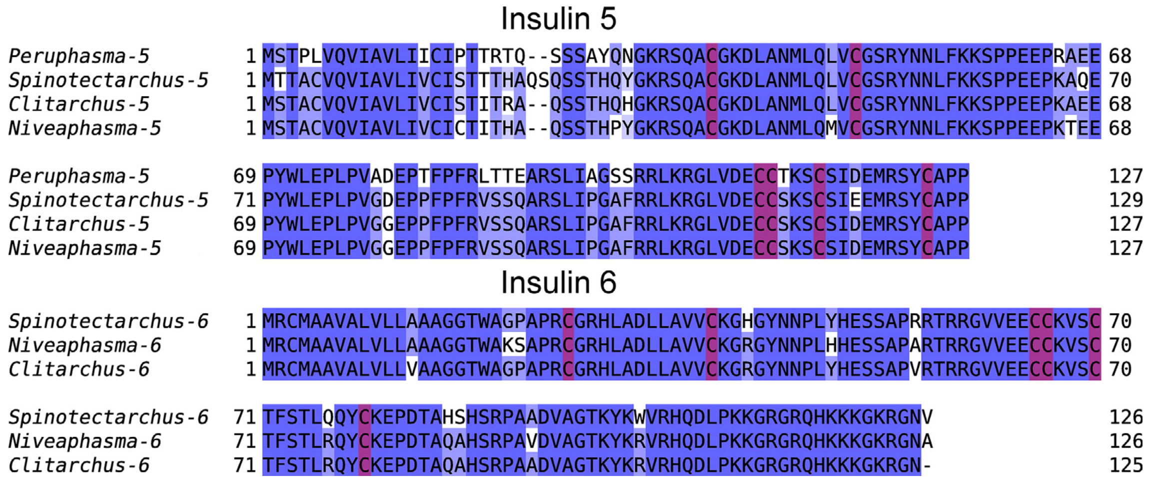
The genome of *Clitarchus* is remarkably good. Nevertheless, the problems associated with the exclusive use of short reads are noticeable in the difficulties in obtaining a complete transcript for the tryptopyrokinin gene and it possibly contributed to a similar problem with the FMRFamide precursor, whereas the exact number of the paracopies on the orcokinin B transcript could well be significantly less (Fig. S1; Table S1).

Finding homologous genes is based on homology searches for similar sequences. When such sequences are not well conserved, as *e.g.* in signal peptides and parts of the precursor that may have little or no physiological function it is difficult to find them if they are present in exons different from those that code the active peptides. This becomes even more of a problem, when intron sizes are large, as is often the case in these two species. The large number of phasmid transcriptomes that is available for species closely related to *C. hookeri* greatly facilitated the deduction of neuropeptide precursors. The transcriptomes of those species revealed neuropeptide precursors that were remarkably similar to those of *C. hookeri,* in some cases only a few silent SNPs were present in the coding regions. The lack of complete precursors for ACP, CCHamide 1, relaxin and periviscerokinin precursors from *Clitarchus* (Fig. S1; Table S1) is due to the absence of transcripts needed to find the less conserved sequences of these precursors. In the case of ACP and CCHamide 1 the absence of transcripts is likely due to low expression of these genes. The available phasmid transcriptomes are mostly made from the midgut or the combination of head and pronotum and thus lack the abdominal ganglia. As in other insect species periviscerokinin is expressed predominantly and relaxin exclusively in the abdominal ganglia it is not surprising that transcripts for these two peptides are lacking from the various transcriptomes.

The *Timema* genome is a draft assembly and as such still contains a significant number of small and larger gaps (see *e.g.* the CG31096 GPCR ortholog in supplementary data). It also has a very large number of ambiguous nucleotides, probably in large part due to the use of a large number of individuals used for DNA extraction. Those ambiguous nucleotides may well represent SNPs. There are also several exons for neuropeptide genes lacking in the assembly that, in some instances, can be found in the single transcript SRA for this species (SRR921650). Nevertheless, the data are useful to show the presence of neuropeptide genes and in many cases one can deduce the complete sequences of the neuropeptides themselves (Fig. S2; Table S1). However, as the amount of transcript data is much more limited than in the case of *Clitarchus,* it is often impossible to find the exons coding poorly conserved regions of the precursors. The separation of the Timematodea from the Euphasmatodea is estimated to have happened about 100 million years ago (Mya), while the common ancestor of the New Zealand species closely related to *C. hookeri* may have lived about 20–25 Mya (Buckley et al., 2010). It is thus not surprising that the sequence differences in neuropeptide precursors are much more pronounced between *Clitarchus* and *Timema* than within the New Zealand species. For example, whereas the AKH sequence for the former is a decapeptide, *Timema* has an octapeptide and whereas *Timema* is predicted to have the classical corazonin sequence, in *Clitarchus* two amino acid residues have changed (Table 1). Unlike *Clitarchus, Timema* has two eclosion hormone genes (Table 1) and while SIFamide is well conserved between these two stick insects, the *Timema* SMYamide, ANYRKLPFNGGMYamide, is significantly different from the predicted *Clitarchus* peptide (Table 1). Most neuropeptide precursors do not merit comments, as, at least

Table 1
Salient features of the neuropeptidomes and corresponding receptors of *Timema* and *Clitarchus.*

Gene	*Timema*	*Clitarchus*
AKH mature peptide	pQVNFSPSWamide	pQLTFTPNWGTamide
Corazonin mature peptide	pQTFQYSRGWTNamide	pQTFQYSNGWTPamide
Dilp8 GPCR ortholog	present	absent
Eclosion Hormone	2 genes	1 gene
EFLamide	2 paracopies	1 paracopy
Leucokinin	not found	not found
Leucokinin GPCR	absent	absent
Sulfakinin	present	absent
Sulfakinin GPCR	present	absent
Tryptopyrokinin gene	absent	present
Vasopressin	absent	≥3 genes
Vasopressin GPCR	absent	1 gene



```
                           Insulin 5
Peruphasma-5         1 MSTPLVQVIAVLIICIPTTRTQ--SSSAYQNGKRSQACGKDLANMLQLVCGSRYNNLFKKSPPEEPRAEE 68
Spinotectarchus-5    1 MTTACVQVIAVLIVCISTTTHAQSQSSTHQYGKRSQACGKDLANMLQLVCGSRYNNLFKKSPPEEPKAQE 70
Clitarchus-5         1 MSTACVQVIAVLIVCISTITRA--QSSTHQHGKRSQACGKDLANMLQLVCGSRYNNLFKKSPPEEPKAEE 68
Niveaphasma-5        1 MSTACVQVIAVLIVCICTITHA--QSSTHPYGKRSQACGKDLANMLQMVCGSRYNNLFKKSPPEEPKTEE 68

Peruphasma-5        69 PYWLEPLPVADEPTFPFRLTTEARSLIAGSSRRLKRGLVDECCTKSCSIDEMRSYCAPP           127
Spinotectarchus-5   71 PYWLEPLPVGDEPPFPFRVSSQARSLIPGAFRRLKRGLVDECCSKSCSIEEMRSYCAPP           129
Clitarchus-5        69 PYWLEPLPVGGEPPFPFRVSSQARSLIPGAFRRLKRGLVDECCSKSCSIDEMRSYCAPP           127
Niveaphasma-5       69 PYWLEPLPVGGEPPFPFRVSSQARSLIPGAFRRLKRGLVDECCSKSCSIDEMRSYCAPP           127
                           Insulin 6
Spinotectarchus-6    1 MRCMAAVALVLLAAAGGTWAGPAPRCGRHLADLLAVVCKGHGYNNPLYHESSAPRRTRRGVVEECCKVSC 70
Niveaphasma-6        1 MRCMAAVALVLLAAAGGTWAKSAPRCGRHLADLLAVVCKGRGYNNPLHHESSAPARTRRGVVEECCKVSC 70
Clitarchus-6         1 MRCMAAVALVLLVAAGGTWAGPAPRCGRHLADLLAVVCKGRGYNNPLYHESSAPVRTRRGVVEECCKVSC 70

Spinotectarchus-6   71 TFSTLQQYCKEPDTAHSHSRPAADVAGTKYKWVRHQDLPKKGRGRQHKKKGKRGNV              126
Niveaphasma-6       71 TFSTLRQYCKEPDTAQAHSRPAVDVAGTKYKRVRHQDLPKKGRGRQHKKKGKRGNA              126
Clitarchus-6        71 TFSTLRQYCKEPDTAQAHSRPAADVAGTKYKRVRHQDLPKKGRGRQHKKKGKRGN-              125
```

Fig. 1. Sequence alignments of two insulin precursors for which the orthologous hormones could be inferred from different phasmid species. Note that in both cases the sequences are very well conserved, even in species as distantly related as *Peruphasma* and *Clitarchus*. The cysteine residues have been colored purple.

superficially, they seem very much like those in other insect species. However, there are both interesting differences between these two species, as well as between phasmids and the majority of other insect species. These are summarized in Table 1.

3.2. Insulins

Both phasmid species have several insulin genes that are similar in structure to the *Drosophila* insulin like peptides (dilps). Most of those that are likely to act on the classical tyrosine kinase receptor seem to have arisen by amplification after the separation of the Euphasmatodean lineage from the Timemodea, as suggested by sequence similarity between the different predicted Euphasmatodean insulins (Fig. 1). Both species also have an insulin that is homologous to *Drosophila* insulin-like peptide 7 (dilp7) that has been named relaxin (Figs. S1 and S2), as it likely acts on a GPCR homologous to a vertebrate relaxin GPCR and *Drosophila* CG34411 (*cf* Veenstra et al., 2012). Homologous GPCRs were found in both species. The termite *Zootermopsis* has a second relaxin GPCR, the one that is homologous to *Drosophila* CG31096 that codes for the receptor of dilp8 (Meissner et al., 2016;Vallejo et al., 2015). The amino acid sequence of dilp8 is very poorly conserved and it has been suggested that it might be the insect ortholog of the crustacean androgenic hormone (Veenstra, 2016b). However, whereas an ortholog of the *Drosophila* dilp8 GPCR is present in the *Timema* genome (Figs. S3 and S4), it is absent from *Clitarchus*. It thus seems likely that the dilp8 signaling pathway has been lost in the latter species.

3.3. Leucokinin and Allatostatin C are lacking in both species and sulfakinin in C. hookeri.

Two common insect neuropeptide genes, leucokinin and sulfakinin, could not be found in the *Clitarchus* assembly and the leucokinin gene could neither be found in the *Timema* genome. However, the latter species does have both a sulfakinin gene and a sulfakinin GPCR. Evidence for leucokinin genes could neither be found in the individual SRAs and the same was true for a leucokinin GPCR. Given the very high coverage of the two genomes and the strong conservation of GPCR transmembrane regions, it must be concluded that both species lost the leucokinin signaling pathway. In other insect species leucokinin stimulates diuresis by the Malpighian tubules which express the leucokinin receptor. However, in the four SRAs from the Malpighian tubules of the phasmid *Carausius morosus* no spot could be found that corresponds to a leucokinin GPCR, although twelve spots were found for a diuretic hormone receptor ortholog. This suggests that leucokinin may be universally absent from phasmids. The absence of sulfakinin from *Clitarchus* was similarly confirmed by the absence of a sulfakinin GPCRs in this species, again from both the genome assembly and the genome SRAs. Of the three reported Arthropod allatostatin C genes, both allatostatins CC and CCC are present, but allatostatin C is absent. The latter was confirmed by the BLAST program on the genomic reads from both species.

3.4. Tryptopyrokinins

Tryptopyrokinins are N-terminally extended pyrokinins that have a typical tryptophan residue. Together with the periviscerokinins these are three different types of related peptides that each have their own dedicated GPCRs, at least in those species where this has been investigated (*e.g.* Iversen et al., 2002; Rosenkilde et al., 2003; Cazzamali et al., 2005; Homma et al.,2006; Paluzzi et al., 2010; Paluzzi and O'Donnell, 2012). Most insect species that have been studied have both a pyrokinin and a periviscerokinin gene, each of which may in addition to pyrokinins or periviscerokinins also produce a tryptopyrokinin. Two neuroendocrine cells of the third neuromere of the suboesophageal ganglion often produce (almost) exclusively tryptopyrokinins from either a pyrokinin or periviscerokinin precursor (see *e.g.* Wegener et al., 2006; Neupert et al., 2010). The exact mechanism of how this is done remains unresolved (differential degrading/inactivation of the pyrokinins or periviscerokinins produced from the same precursor in these cells). Interestingly, in two Polyneopteran species, *Locusta* and *Zootermopsis*, specific tryptopyrokinin genes were found that produce precursors coding for multiple tryptopyrokinin paracopies (Veenstra, 2014; Redeker et al., 2017). The short illumina sequences did not allow for the deduction of the complete sequence of such a tryptopyrokinin gene from *Clitarchus*, nevertheless numerous genomic reads encoding two such peptides show that such a gene is present in this species. Interestingly, several of the *Clitarchus* pyrokinins are slightly modified in that the typical Phe and Leu residues of the C-terminal core sequence are replaced by Tyr and Met residues respectively. Surprisingly, although the phylogenetic position of *Timema* (Misof et al., 2014) suggests that its ancestor had such a gene, it must have been lost. Searching for tryptopyrokinin coding genome reads only yields reads that code for the parts of the periviscerokinin and pyrokinin genes, each of which code for a single tryptopyrokinin paracopy (Table S2).

3.5. EFLamide

In arthropods EFLamide was first described from the spider mite, *Tetranychus urticae,* where it is predicted to produce both EFLamides and EFLGGPamides from two alternatively spliced mRNAs (Veenstra et al., 2012). The gene is homologous to the PXFVamide from the mollusc *Lottia gigantea*, the FVamide gene from the polychaete worm *Capitella teleta* and the *Platynereis* FVamide/EFLGamide gene (Veenstra, 2010, 2011; Conzelmann et al., 2013). All of these genes appear to have alternatively spliced mRNAs. Although, alternative splicing of this gene has also been reported from spiders and a scorpion (Veenstra, 2016a), only a single mRNA has been found in myriapods, decapods or *Daphnia pulex* and several apterygote hexapods (Veenstra, 2016a,b; Derst et al., 2016).

Identification of a *Platynereis* GPCR that is homologous the vertebrate TRH receptor as the receptor for this peptide in (Bauknecht and Jékely, 2015) made it possible to identify one of the *Nilaparvata* GPCRs (Tanaka et al., 2014) as a likely EFLamide receptor (Veenstra, 2016a). This strongly suggested that some insect species might have such a similar neuropeptide gene and this was recently confirmed with the description of a partial EFLamide precursor from the bed bug *Cimex lectularius* (Predel et al., 2018). The presence of DNA sequences coding partial GPCRs orthologous with *Nilaparvata* GPCR A45 in both phasmid genomes suggested that these insects also have such genes. Of the various *Clitarchus* genomic DNA sequences that potentially encode EFLGKR sequences only one yielded a plausible EFLamide precursor exon. It starts with a phase 2 intron acceptor site and could be the last coding exon of an EFLamide neuropeptide gene. It predicts a mature SLGSEFLamide that is preceded by a classical convertase cleavage site. A homologous sequence was found in the *Timema* genome, where the predicted EFLamide precursor exon codes for two EFLamide paracopies, DLGSEFLamide and SLGLEFLamide. This exon also starts with a phase 2 intron (Fig. 2).

Attempts to recover the remainder of the precursors failed. There is a limited number of individual reads that corresponds to the *Clitarchus* exon and four (SRR5889596.6374502, SRR5889596.3495238, SRR5889598.572299, SRR5889596.25123060) of them cover the intron acceptor splice site. However, the three different exons spliced on to the EFLamide coding exon that can be identified in the genome (the fourth has only a few nucleotides), all have all in frame stop codons and can thus not be part of an EFLamide precursor mRNA. The absence of an exon coding for the signal peptide from the transcriptome data and more specifically the presence of non-sense exons spliced on the EFLamide coding exon might, superficially, suggest that the EFLamide coding exon forms part of pseudogene. However, similar arguments for the *Drosophila* RYamide gene, where very large numbers of transcripts did not yield a single transcript read coding for a signal peptide but instead many non-sense exons (Veenstra and Khammassi, 2017) were invalidated by the isolation of the RYamide peptides (Ida et al., 2011).

Analysis of other phasmid transcriptomes yielded a partial transcript from *Ramulus artemis* (GAWE01031730.1), that shows significant homology to the *Clitarchus* exon (Fig. 2). The only transcript read from *Timema* (SRR921650.3697533) shows the putative intron acceptor site unused, and could represent an mRNA molecule from the opposite strand. Identification of EFLamide encoding exons from various insect genomes yielded similar putative exons, all except one starting with a phase 2 intron acceptor site; the mayfly *Ephemera danica* is so far the only insect predicted to produce multiple EFLamide paracopies (Fig. S5).

3.6. The vasopressin-like peptide was lost in Timema but Clitarchus has two such peptides

Perhaps the most intriguing find is the presence of several genes encoding a Lys-vasopressin-like neuropeptide in *Clitarchus* (Fig. 3; Table S1), while such a gene and the one coding its receptor are absent from *Timema*. The absence of a vasopressin-like peptide in *Timema* was confirmed in the same fashion as for the leucokinin and sulfakinin signaling pathways. In the current *Clitarchus* genome assembly there are several scaffolds that contain either some or all three coding exons of a vasopressin-like neuropeptide gene. There are five instances in the assembly in which all three coding exons are arranged in a logical order to produce a complete precursor. Of those, four are quite similar between them and those are predicted to produce CLITNCPKGamide, while the other one is quite distinct and is predicted to yield CLIVNCPKGamide (Fig. 3). However, as explained above, it is not uncommon to find the same gene incorporated into the genome assembly in different locations when using short reads for assembly. For example, there are two and a half AKH genes in the assembly, but in that case the coding sequences of the AKH precursor are completely identical, suggesting a very recent origin of this duplication, *i. e.* during the genome assembly. This is not the case for the various vasopressin-like genes in the *Clitarchus* assembly, and furthermore the numbers of the genome reads suggest that there are likely to be about four different genes, but the actual number could be different. Partial transcripts confirm the expression of at least two of those genes in *Clitarchus*, and partial transcripts similarly show the expression of at least two vasopressin genes in *Techtarchus ovobessus, Niveaphasma annulata* and *Asteliaphasma jucundum*. On the contrary, only a single gene was found that encodes an vasopressin GPCR ortholog. Interestingly, such receptors appear abundantly expressed in the midgut of a variety of phasmids, as the transcriptomes from the midguts of these species contain complete transcripts for this GPCR (Fig. 4).

In *Clitarchus* there are several genes encoding a vasopressin-like neuropeptide and there is one gene for an vasopressin GPCR, while *Timema* does not have such genes. There are other species, *Homo sapiens* for one, that have more than one gene coding vasopressin/oxytocin-like peptides. In those species, there are usually two receptors and the two peptides that have structures that are significantly different. Although only two amino acid residues differ between vasopressin and oxytocin,

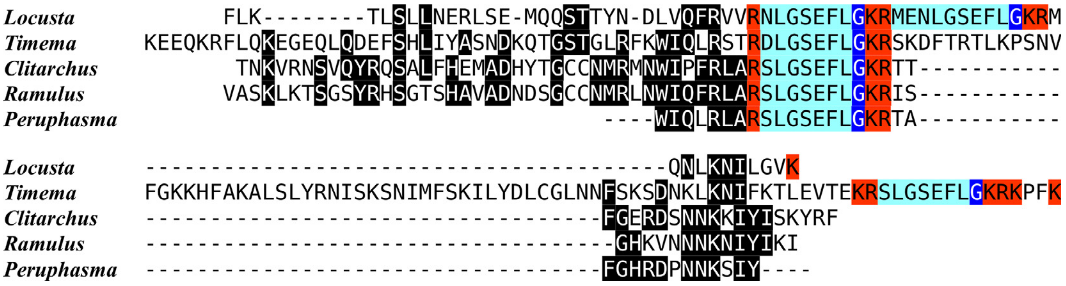
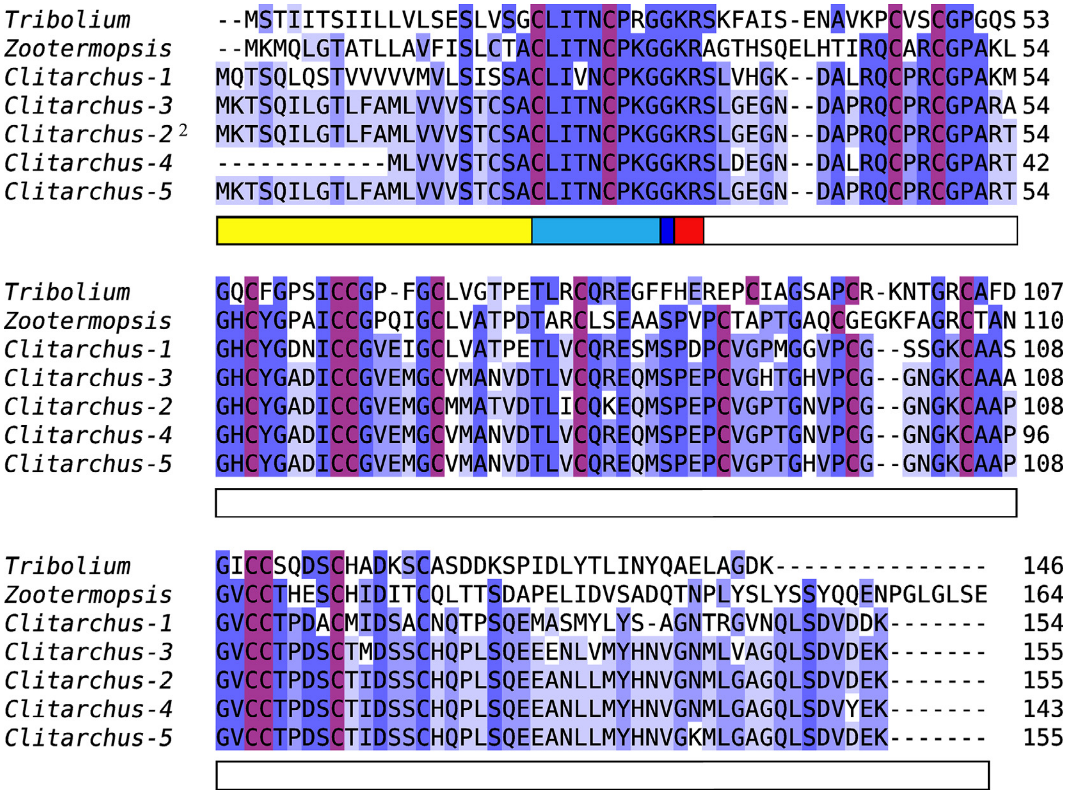


Fig. 2. Partial sequences of phasmid EFLamide precursors. The EFLamide precursor from the common ancestor of locusts and phasmids likely coded for multiple EFLamide copies, as the mayfly gene still does (Fig. S5). Comparing the sequences of these precursors, it would seem that the phasmids and the locust lost different paracopies from the ancestral gene. Sequences deduced from genome assemblies of L. migratoria, T. cristinae and C. hookeri, an R. artemis transcriptome contig (GAWE01031730.1) or a single genome read from P. schultei (SRR1182314.8015249.1). Amino acid residues in red are predicted to be cleaved by convertase and subsequently removed by a carboxypeptidase, those in dark blue to be transformed in C-terminal amides and those in black indicate other residues that are conserved between different species. For additional insect partial EFLamide precursors see Fig. S5.



```
Tribolium        --MSTIITSIILLVLSESLVSGCLITNCPRGGKRSKFAIS-ENAVKPCVSCGPGQS 53
Zootermopsis     --MKMQLGTATLLAVFISLCTACLITNCPKGGKRAGTHSQELHTIRQCARCGPAKL 54
Clitarchus-1     MQTSQLQSTVVVVVMVLSISSACLIVNCPKGGKRSLVHGK--DALRQCPRCGPAKM 54
Clitarchus-3     MKTSQILGTLFAMLVVVSTCSACLITNCPKGGKRSLGEGN--DAPRQCPRCGPARA 54
Clitarchus-2[2]  MKTSQILGTLFAMLVVVSTCSACLITNCPKGGKRSLGEGN--DAPRQCPRCGPART 54
Clitarchus-4     ------------MLVVVSTCSACLITNCPKGGKRSLDEGN--DALRQCPRCGPART 42
Clitarchus-5     MKTSQILGTLFAMLVVVSTCSACLITNCPKGGKRSLGEGN--DAPRQCPRCGPART 54

Tribolium        GQCFGPSICCGP-FGCLVGTPETLRCQREGFFHEREPCIAGSAPCR-KNTGRCAFD 107
Zootermopsis     GHCYGPAICCGPQIGCLVATPDTARCLSEAASPVPCTAPTGAQCGEGKFAGRCTAN 110
Clitarchus-1     GHCYGDNICCGVEIGCLVATPETLVCQRESMSPDPCVGPMGGVPCG--SSGKCAAS 108
Clitarchus-3     GHCYGADICCGVEMGCVMANVDTLVCQREQMSPEPCVGHTGHVPCG--GNGKCAAA 108
Clitarchus-2     GHCYGADICCGVEMGCMMATVDTLICQKEQMSPEPCVGPTGNVPCG--GNGKCAAP 108
Clitarchus-4     GHCYGADICCGVEMGCVMANVDTLVCQREQMSPEPCVGPTGNVPCG--GNGKCAAP 96
Clitarchus-5     GHCYGADICCGVEMGCVMANVDTLVCQREQMSPEPCVGPTGHVPCG--GNGKCAAP 108

Tribolium        GICCSQDSCHADKSCASDDKSPIDLYTLINYQAELAGDK---------------   146
Zootermopsis     GVCCTHESCHIDITCQLTTSDAPELIDVSADQTNPLYSLYSSYQQENPGLGLSE   164
Clitarchus-1     GVCCTPDACMIDSACNQTPSQEMASMYLYS-AGNTRGVNQLSDVDDK-------   154
Clitarchus-3     GVCCTPDSCTMDSSCHQPLSQEEENLVMYHNVGNMLVAGQLSDVDEK-------   155
Clitarchus-2     GVCCTPDSCTIDSSCHQPLSQEEANLLMYHNVGNMLGAGQLSDVDEK-------   155
Clitarchus-4     GVCCTPDSCTIDSSCHQPLSQEEANLLMYHNVGNMLGAGQLSDVYEK-------   143
Clitarchus-5     GVCCTPDSCTIDSSCHQPLSQEEANLLMYHNVGKMLGAGQLSDVDEK-------   155
```

Fig. 3. Sequence alignment of vasopressin-like neuropeptide precursors from Tribolium castaneum, Zootermopsis nevadensis and Clitarchus hookeri. Note that the Clitarchus 1 precursor is significantly different from the other four Clitarchus precursors; this is particularly noticeable in the signal peptide, but also in the mature vasopressin-like peptide. Location of the signal peptide in the precursors is indicated in yellow, the sequence of the vasopressin-like peptide in light blue, the Gly residue predicted to be transformed into a C-terminal amide in dark blue and the convertase cleavage site in red. Cysteine residues are in purple and other conserved amino acid residues are in blue. Sequences are those described here and previously (Aikins et al., 2008; Stafflinger et al., 2008; Veenstra, 2014).

these involve Leu/Arg and Phe/Ile changes, both of which, but particularly the first one, have large effects on receptor binding. This insures that at physiological concentrations the vasopressin and oxytocin receptors are specific for their respective ligands. However, whereas those two neuropeptides have acquired different physiological functions, this is unlikely to be the case for the two vasopressin *Clitarchus* isoforms. On the one hand, the amino acid change between Thr and Val is structurally much less important than the replacement of positively charged amino acid for a neutral one and on the other hand *Clitarchus* has only one GPCR that can be expected to bind these peptides.

After *L. migratoria, C. hookeri* is the second insect species for which we know that this gene is amplified. In *Locusta* the two neuroendocrine cells producing this peptide release it into the hemolymph (Rémy and Girardie, 1980), while in *Schistocerca gregaria*, another migratory locust, anatomy of these neurons shows that the peptide is released exclusively within the central nervous system (Tyrer et al., 1993). It is, therefore, likely that much larger quantities of peptide need to be produced in *Locusta* and this may explain why these neuroendocrine cells are so much bigger in *Locusta* than in *Schistocerca*. If the function of a particular cell is the production and export of a single protein in large quantities, increasing cell size may be insufficient and amplification of the gene coding such a protein is likely favored by selection in order to insure sufficient production of mRNA (*cf* Veenstra, 2014). This suggests that in *Clitarchus* these cells also release the vasopressin-like peptide as a hormone. That this peptide is not released exclusively within the central nervous system of the Euphasmatodea is also suggested by the presence of complete vasopressin GPCR transcripts in the midgut transcriptomes of various phasmids. Obviously, if it were exclusively released within the central nervous system, it would not make sense to express its receptor in the midgut in quantities sufficient to allow assembly of complete transcripts from RNAseq experiments.

The differences in the primary amino acid sequences of *Clitarchus* vasopressin precursors coding for CLITNCPKGamide and CLIVNCPKGamide respectively (Fig. 3) are significant, and comparable to, if not larger than, those between orthologous insulin precursors from *Peruphasma* and *Clitarchus* (Fig. 1). It thus seems that the initial amplification of the vasopressin-like gene in Euphasmatodea is ancient and likely widespread. Furthermore, the presence of complete transcripts for the vasopressin GPCR in various Euphasmatoid midgut transcriptomes suggests that its ligand is commonly released into the hemolymph. Wherever the peptide may be released in *Clitarchus,* the data strongly suggest an increased use of vasopressin-like peptides in one phasmid species and its loss in another. It is tempting to speculate that in the common ancestor of phasmids the role of this peptide was rather limited, such that it got lost in some of its descendants and that it got reemployed for a different or related function in others.

The function of the vasopressin related peptides in insects is a topic of dispute. The first such peptide identified from insects was the antiparallel dimere of an arg-vasopressin-like peptide from *Locusta* as a hormone that increased amaranth secretion by the Malpighian tubules. Interestingly, the monomere was without effect on amaranth excretion (Proux et al., 1987). Although this suggested that the antiparallel dimere stimulates fluid secretion by the Malpighian tubules as a diuretic hormone, this was shown to be incorrect (Coast et al., 1993). The hypothesis that it may stimulate the release of a diuretic hormone has been proposed to explain this apparent discrepancy (Aikins et al., 2008). In other protostomians vasopressin-related peptides may be involved in the regulation of reproduction (van Kesteren et al., 1995; Garrison et al., 2012). It is of interest to note in this respect that many years ago the cells that we now know to produce these peptides were shown to experience important cytological changes after ovariectomy in a cricket (Dürnberger et al., 1978), while in the phasmid *Carausius*

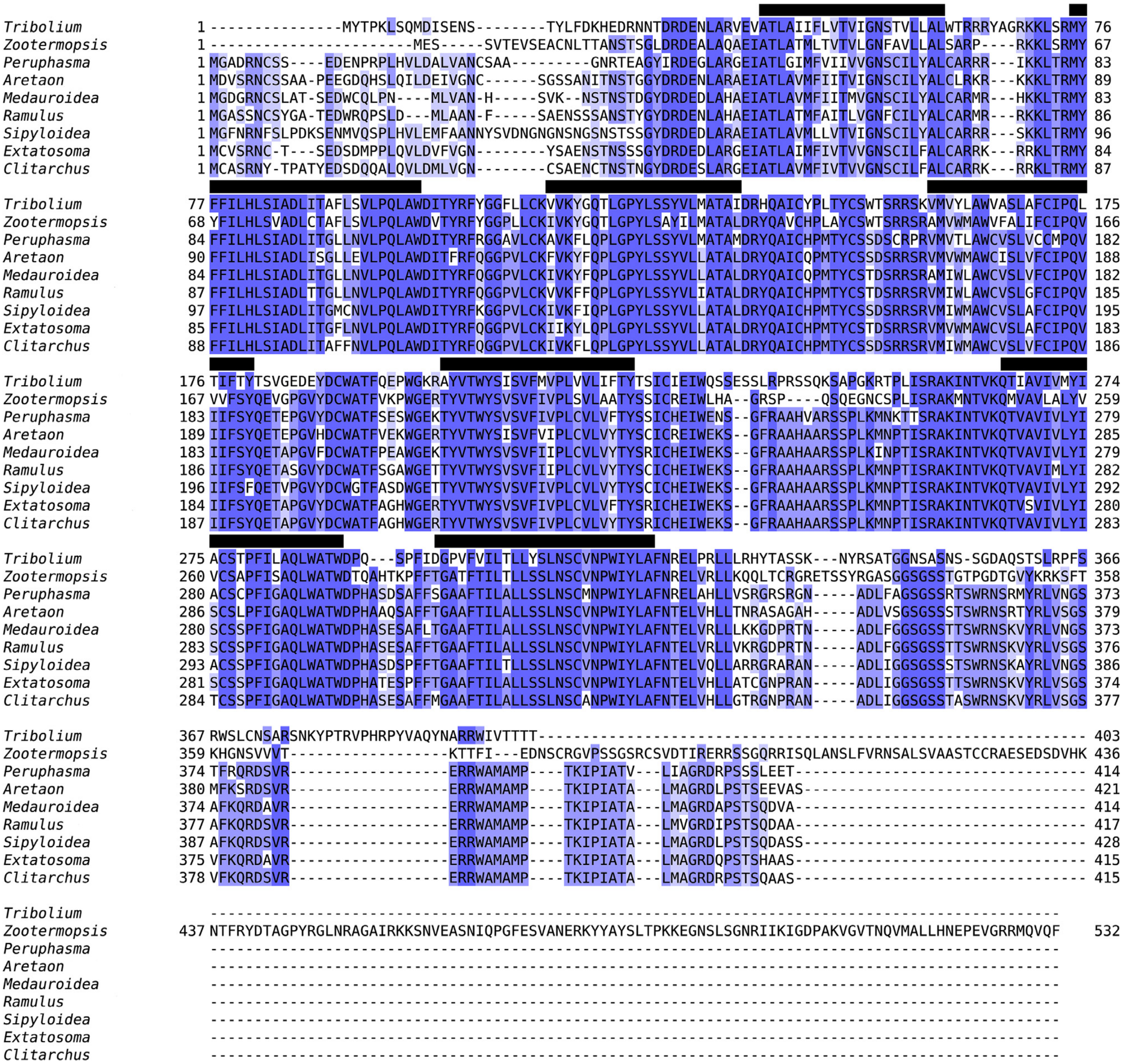


Fig. 4. Sequence alignment of phasmid vasopressin receptors with those from Tribolium and Zootermopsis. The majority of these sequences come from midgut transcriptomes, Extatosoma tiaratum (GAWG01068688.1), Medauroidea extradentata (GAWD01027027.1), Ramulus artemis (GAWE01092087.1), Aretaon asperrimus (GAWC01043798.1) and Sipyloidea sipylus (GAWF01069286.1). The sequence from Peruphasma schultei was obtained with Trinity using reads from transcriptome SRAs from this species, while the Clitarchus hookeri sequence was deduced from its genome by homology to the other phasmid GPCRs. Sequences for Tribolium and Zootermopsis have been previously published (Aikins et al., 2008; Stafflinger et al., 2008; Veenstra, 2014). Black bars over the sequences indicate the locations of the transmembrane regions.

morosus such cells increase in size in reproducing females and have been suggested to change their physiological activity in relation to pigmentation (Raabe, 1966). It will be interesting to elucidate the function of these peptides in insects.

Acknowledgements

I am most grateful for the constructive criticism of two reviewers which allowed me to improve the manuscript and eliminate some errors.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.ygcen.2018.04.027.

References

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

Armisén, D., Rajakumar, R., Friedrich, M., Benoit, J.B., Robertson, H.M., Panfilio, K.A., Seung, J.A., Poelchau, M., Chao, H., Dinh, H., Doddapaneni, H., Shannon Dugan, S., Richard A. Gibbs, R.A., Hughes, D.S.T., Han, Y., Lee, S.L., Murali, S.C., Muzny, D.M., Qu, J., Worley, K.C., Munoz-Torres, M., Abouheif, A., Bonneton, F., Chen, T., Childers, C., Cridge, A.G., Crumière, A.J.J., Amelie Decaras, A., Didion, E.M., Elizabeth Duncan, E., Elpidina, E.N., Favé, M.J., Finet, C., Jacobs, C.G.C., Jarvela, A., Jennings, E.J., Jones, J.W., Lesoway, M.P., Lovegrove, M., Martynov, A., Oppert, B., Lillico-Ouachour, A., Rajakumar, A., Refki, P.N., Rosendale, A.J., Santos, M.E., Toubiana, W., van der Zee, M., Jentzsch, I.M.V., Lowman, A.V., Viala, S., Richards, S., Khila, A., 2018. The genome of the water strider Gerris buenoi reveals expansions of gene repertoires associated with adaptations to life on the water. bioRxiv http://dx.doi.org/10.1101/242230.

Audsley, N., Vandersmissen, H.P., Weaver, R., Dani, P., Matthews, J., Down, R., Vuerinckx, K., Kim, Y.J., Vanden Broeck, J., 2013. Characterisation and tissue distribution of the PISCF allatostatin receptor in the red flour beetle, Tribolium castaneum. Insect Biochem. Mol. Biol. 43, 65–74.

Bendtsen, J.D., Nielsen, H., von Heijne, G., Brunak, S., 2004. Improved prediction of signal peptides: SignalP 3.0. J. Mol. Biol. 340, 783–795.

Bauknecht, P., Jékely, G., 2015. Large-scale combinatorial deorphanization of *Platynereis* Neuropeptide GPCRs. Cell Rep. 12, 684–693.

Buckley, T.R., Attanayake, D., Nylander, J.A.A., Bradler, S., 2010. The phylogenetic placement and biogeographical origins of the New Zealand stick insects (Phasmatodea). Syst. Entomol. 35, 207–225.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., Madden, T.L., 2009. BLAST+: architecture and applications. BMC Bioinf. 10, 421.

Cazzamali, G., Torp, M., Hauser, F., Williamson, M., Grimmelikhuijzen, C.J.P., 2005. The *Drosophila* gene CG9918 codes for a pyrokinin-1 receptor. Biochem. Biophys. Res. Commun. 335, 14–19.

Chen, W., Hasegawa, D.K., Kaur, N., Kliot, A., Pinheiro, P.V., Luan, J., Stensmyr, M.C., Zheng, Y., Liu, W., Sun, H., Xu, Y., Luo, Y., Kruse, A., Yang, X., Kontsedalov, S., Lebedev, G., Fisher, T.W., Nelson, D.R., Hunter, W.B., Brown, J.K., Jander, G., Cilia, M., Douglas, A.E., Ghanim, M., Simmons, A.M., Wintermantel, W.M., Ling, K.S., Fei, Z., 2016. The draft genome of whitefly *Bemisia tabaci* MEAM1, a global crop pest, provides novel insights into virus transmission, host adaptation, and insecticide resistance. BMC Biol. 14, 110.

Coast, G.M., Rayne, R.C., Hayes, T.K., Mallet, A.I., Thompson, K.S., Bacon, J.P., 1993. A comparison of the effects of two putative diuretic hormones from *Locusta migratoria* on isolated locust Malpighian tubules. J. Exp. Biol. 175, 1–14.

Conzelmann, M., Williams, E.A., Krug, K., Franz-Wachtel, M., Macek, B., Jékely, G., 2013. The neuropeptide complement of the marine annelid *Platynereis dumerilii*. BMC Genomics 14, 906.

Dennis, A.B., Dunning, L.T., Sinclair, B.J., Buckley, T.R., 2015a. Parallel molecular routes to cold adaptation in eight genera of New Zealand stick insects. Sci. Rep. 5, 13965.

Dennis, A.B., Dunning, L.T., Sinclair, B.J., Buckley, T.R., 2015b. Data from: parallel molecular routes to cold adaptation in eight genera of New Zealand stick insects. Dryad Digital Repository. http://dx.doi.org/10.5061/dryad.kc826.

Derst, C., Dircksen, H., Meusemann, K., Zhou, X., Liu, S., Predel, R., 2016. Evolution of neuropeptides in non-pterygote hexapods. BMC Evol. Biol. 16, 51.

Dickinson, P.S., Armstrong, M.K., Dickinson, E.S., Fernandez, R., Miller, A., Pong, S., Powers, B., Pupo Wiss, A., Stanhope, M.E., Walsh, P.J., Wiwatpanit, T., Christie, A.E., 2018. Three members of a peptide family are differentially distributed and elicit differential state-dependent responses in a pattern generator-effector system. i. J. Neurophysiol n press.

Dunning, L.T., Dennis, A.B., Park, D., Sinclair, B.J., Newcomb, R.D., Buckley, T.R., 2013. Identification of cold-responsive genes in a New Zealand alpine stick insect using RNA-Seq. Comp. Biochem. Physiol. D8, 24–31.

Dürnberger, H., Pohlhammer, K., Weinbörmair, G., 1978. The paramedial neurosecretory cells of the suboesophageal ganglion of the cricket, *Teleogryllus commodus* (Walk.). Cell Tissue Res. 187, 489–494.

Faddeeva-Vakhrusheva, A., Derks, M.F., Anvar, S.Y., Agamennone, V., Suring, W., Smit, S., van Straalen, N.M., Roelofs, D., 2016. Gene family evolution reflects adaptation to soil environmental stressors in the genome of the collembolan *Orchesella cincta*. Genome Biol. Evol. 8, 2106–2117.

Faddeeva-Vakhrusheva, A., Kraaijeveld, K., Derks, M.F.L., Anvar, S.Y., Agamennone, V., Suring, W., Kampfraath, A.A., Ellers, J., Le Ngoc, G., van Gestel, C.A.M., Mariën, J., Smit, S., van Straalen, N.M., Roelofs, D., 2017. Coping with living in the soil: the genome of the parthenogenetic springtail *Folsomia candida*. BMC Genomics 18, 493.

Garrison, J.L., Macosko, E.Z., Bernstein, S., Pokala, N., Albrecht, D.R., Bargmann, C.I., 2012. Oxytocin/vasopressin-related peptides have an ancient role in reproductive behavior. Science 338, 540–543.

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

Homma, T., Watanabe, K., Tsurumaru, S., Kataoka, H., Imai, K., Kamba, M., Niimi, T., Yamashita, O., Yaginuma, T., 2006. G protein-coupled receptor for diapause hormone, an inducer of *Bombyx* embryonic diapause. Biochem. Biophys. Res. Commun. 344, 386–393.

Hou, L., Jiang, F., Yang, P., Wang, X., Kang, L., 2015. Molecular characterization and expression profiles of neuropeptide precursors in the migratory locust. Insect Biochem. Mol. Biol. 63, 63–71.

Hunter, W.B., Reese, J., International Psyllid Genome Consortium, 2014. The Asian citrus psyllid genome (Diaphorina citri, Hemiptera). The Asian citrus psyllid genome (*Diaphorina citri*, Hemiptera). J. Citrus Pathol. 1, 143.

Ida, T., Takahashi, T., Tominaga, H., Sato, T., Kume, K., Ozaki, M., Hiraguchi, T., Maeda, T., Shiotani, H., Terajima, S., Sano, H., Mori, K., Yoshida, M., Miyazato, M., Kato, J., Murakami, N., Kangawa, K., Kojima, M., 2011. Identification of the novel bioactive peptides dRYamide-1 and dRYamide-2, ligands for a neuropeptide Y-like receptor in *Drosophila*. Biochem. Biophys. Res. Commun. 410, 872–877.

Iversen, A., Cazzamali, G., Williamson, M., Hauser, F., Grimmelikhuijzen, C.J.P., 2002. Molecular cloning and functional expression of a *Drosophila* receptor for the neuropeptides capa-1 and -2. Biochem. Biophys. Res. Commun. 299, 628–633.

Meissner, G.W., Luo, S.D., Dias, B.G., Texada, M.J., Baker, B.S., 2016. Sex-specific regulation of Lgr3 in *Drosophila* neurons. Proc. Natl. Acad. Sci. U.S.A. 113, E1256–E1265.

Mesquita, R.D., Vionette-Amaral, R.J., Lowenberger, C., Rivera-Pomar, R., Monteiro, F.A., Minx, P., Spieth, J., Carvalho, A.B., Panzera, F., Lawson, D., Torres, A.Q., Ribeiro, J.M., Sorgine, M.H., Waterhouse Montague, R.M.M.J., Abad-Franch, F., Alves-Bezerra, M., Amaral, L.R., Araujo, H.M., Araujo, R.N., Aravind, L., Atella, G.C., Azambuja, P., Berni, M., Bittencourt-Cunha, P.R., Braz, G.R., Calderón-Fernández, G., Carareto, C.M., Christensen, M.B., Costa, I.R., Costa, S.G., Dansa, M., Daumas-Filho, C.R., De-Paula, I.F., Dias, F.A., Dimopoulos, G., Emrich, S.J., Esponda-Behrens, N., Fampa, P., Fernandez-Medina, R.D., da Fonseca, R.N., Fontenele, M., Fronick, C., Fulton, L.A., Gandara, A.C., Garcia, E.S., Genta, F.A., Giraldo-Calderón, G.I., Gomes, B., Gondim, K.C., Granzotto, A., Guarneri, A.A., Guigó, R., Harry, M., Hughes, D.S., Jablonka, W., Jacquin-Joly, E., Juárez, M.P., Koerich, L.B., Lange, A.B., Latorre-Estivalis, J.M., Lavore, A., Lawrence, G.G., Lazoski, C., Lazzari, C.R., Lopes, R.R., Lorenzo, M.G., Lugon, M.D., Majerowicz, D., Marcet, P.L., Mariotti, M., Masuda, H., Megy, K., Melo, A.C., Missirlis, F., Mota, T., Noriega, F.G., Nouzova, M., Nunes, R.D., Oliveira, R.L., Oliveira-Silveira, G., Ons, S., Orchard, I., Pagola, L., Paiva-Silva, G.O., Pascual, A., Pavan, M.G., Pedrini, N., Peixoto, A.A., Pereira, M.H., Pike, A., Polycarpo, C., Prosdocimi, F., Ribeiro-Rodrigues, R., Robertson, H.M., Salerno, A.P., Salmon, D., Santesmasses, D., Schama, R., Seabra-Junior, E.S., Silva-Cardoso, L., Silva-Neto, M.A., Souza-Gomes, M., Sterkel, M., Taracena, M.L., Tojo, M., Tu, Z.J., Tubio, J.M., Ursic-Bedoya, R., Venancio, T.M., Walter-Nuno, A.B., Wilson, D., Warren, W.C., Wilson, R.K., Huebner, E., Dotson, E.M., Oliveira, P.L., 2015. Genome of *Rhodnius prolixus*, an insect vector of Chagas disease, reveals unique adaptations to hematophagy and parasite infection. Proc. Natl. Acad. Sci. U.S.A 112, 14936–14941.

Misof, B., Liu, S., Meusemann, K., Peters, R.S., Donath, A., Mayer, C., Frandsen, P.B., Ware, J., Flouri, T., Beutel, R.G., Niehuis, O., Petersen, M., Izquierdo-Carrasco, F., Wappler, T., Rust, J., Aberer, A.J., Aspöck, U., Aspöck, H., Bartel, D., Blanke, A., Berger, S., Böhm, A., Buckley, T.R., Calcott, B., Chen, J., Friedrich, F., Fukui, M., Fujita, M., Greve, C., Grobe, P., Gu, S., Huang, Y., Jermiin, L.S., Kawahara, A.Y., Krogmann, L., Kubiak, M., Lanfear, R., Letsch, H., Li, Y., Li, Z., Li, J., Lu, H., Machida, R., Mashimo, Y., Kapli, P., McKenna, D.D., Meng, G., Nakagaki, Y., Navarrete-Heredia, J.L., Ott, M., Ou, Y., Pass, G., Podsiadlowski, L., Pohl, H., von Reumont, B.M., Schütte, K., Sekiya, K., Shimizu, S., Slipinski, A., Stamatakis, A., Song, W., Su, X., Szucsich, N.U., Tan, M., Tan, X., Tang, M., Tang, J., Timelthaler, G., Tomizuka, S., Trautwein, M., Tong, X., Uchifune, T., Walzl, M.G., Wiegmann, B.M., Wilbrandt, J., Wipfler, B., Wong, T.K., Wu, Q., Wu, G., Xie, Y., Yang, S., Yang, Q., Yeates, D.K., Yoshizawa, K., Zhang, Q., Zhang, R., Zhang, W., Zhang, Y., Zhao, J., Zhou, C., Zhou, L., Ziesmann, T., Zou, S., Li, Y., Xu, X., Zhang, Y., Yang, H., Wang, J., Wang, J., Kjer, K.M., Zhou, X., 2014. Phylogenomics resolves the timing and pattern of insect evolution. Science 346, 763–767.

Murali, S.C., Bandaranaike, D., Bellair, M., Blankenburg, K., Chao, H., Dinh, H., Doddapaneni, H., Downs, B., Dugan-Rocha, S., Elkadiri, S., Gnanaolivu, R. D., Hernandez, B., Javaid, M., Jayaseelan, J. C., Lee, S., Li, M., Ming, W., Munidasa, M., Muniz, J., Nguyen, L., Ongeri, F., Osuji, N., Pu, L.-L., Puazo, M., Qu, C., Quiroz, J., Raj, R., Weissenberger, G., Xin, Y., Zou, X., Han, Y., Richards, Stephen, Worley, Kim C., Muzny, Donna M., Gibbs, Richard A., Koelzer. S., Panfilio, K. A., 2015. Ag Data Commons. doi: 10.15482/USDA.ADC/1173238.

Neupert, S., Russell, W.K., Russell, D.H., Predel, R., 2010. Two capa-genes are expressed in the neuroendocrine system of *Rhodnius prolixus*. Peptides 31, 408–411.

Paluzzi, J.P., O'Donnell, M.J., 2012. Identification, spatial expression analysis and functional characterization of a pyrokinin-1 receptor in the Chagas' disease vector, *Rhodnius prolixus*. Mol. Cell. Endocrinol. 363, 36–45.

Paluzzi, J.P., Park, Y., Nachman, R.J., Orchard, I., 2010. Isolation, expression analysis, and functional characterization of the first antidiuretic hormone receptor in insects. Proc. Natl. Acad. Sci. U.S.A. 107, 10290–10295.

Petersen, T.N., Brunak, S., von Heijne, G., Nielsen, H., 2011. Signal P4.0: Discriminating signal peptides from transmembrane regions.

Predel, R., Neupert, S., Derst, C., Reinhardt, K., Wegener, C., 2018. Neuropeptidomics of the bed bug *Cimex lectularius*. J. Proteome Res. 17, 440–454.

Proux, J.P., Miller, C.A., Li, J.P., Carney, R.L., Girardie, A., Delaage, M., Schooley, D.A., 1987. Identification of an arginine vasopressin-like diuretic hormone from *Locusta migratoria*. Biochem. Biophys. Res. Commun. 149, 180–186.

Raabe, M., 1966. Recherches sur la neurosécrétion dans la chaîne nerveuse ventrale du Phasme, *Clitumnus extradentatus*: Liaison entre l'activité des cellules B1 et la pigmentation. C.R. Acad. Sc. Paris 263, 408–411.

Redeker, J., Bläser, M., Neupert, S., Predel, R., 2017. Identification and distribution of products from novel tryptopyrokinin genes in the locust, *Locusta migratoria*. Biochem. Biophys. Res. Commun. 486, 70–75.

Rémy, C., Girardie, J., 1980. Anatomical organization of two vasopressin-neurophysin-like neurosecretory cells throughout the central nervous system of the migratory locust. Gen. Comp. Endocrinol. 40, 27–35.

Richards, S., Murali, S.C., 2015. Best practices in insect genome sequencing: what works and what doesn't. COIS 7, 1–7.

Rosenkilde, C., Cazzamali, G., Williamson, M., Hauser, F., Søndergaard, L., DeLotto, R., Grimmelikhuijzen, C.J.P., 2003. Molecular cloning, functional expression, and gene silencing of two *Drosophila* receptors for the *Drosophila* neuropeptide pyrokinin-2. Biochem. Biophys. Res. Commun. 309, 485–494.

Rüdiger, R., Muschick, M., Lindtke, D., Villoutreix, R., Comeault, A.A., Farkas, T.E., Lucek, K., Hellen, E., Soria-Carrasco, V., Dennis, S.R., de Carvalho, C.F., Safran, R.J., Sandoval, C.P., Feder, J.L., Gries, R., Crespi, B.J., Gries, G., Gompert, Z., Nosil, P., 2018. Transitions between phases of genomic differentiation during stick-insect speciation. Evol. Nature Ecol in press.

Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M.A., Barrell, B., 2000. Artemis: sequence visualization and annotation. Bioinformatics 16, 944–945.

Shelomi, M., 2017. De novo transcriptome analysis of the excretory tubules of *Carausius morosus* (Phasmatodea) and possible functions of the midgut 'appendices'. PLoS One 12, e0174984. http://dx.doi.org/10.1371/journal.pone.0174984.

Shelomi, M., Jasper, W.C., Atallah, J., Kimsey, L.S., Johnson, B.R., 2014. Differential expression of endogenous plant cell wall degrading enzyme genes in the stick insect (Phasmatodea) midgut. BMC Genomics 15, 917.

Shelomi, M., 2017. De novo transcriptome analysis of the excretory tubules of Carausius morosus (Phasmatodea) and possible functions of the midgut 'appendices'. PLoS One 12, e0174984.

Stafflinger, E., Hansen, K.K., Hauser, F., Schneider, M., Cazzamali, G., Williamson, M., Grimmelikhuijzen, C.J.P., 2008. Cloning and identification of an oxytocin/vasopressin-like receptor and its ligand from insects. Proc. Natl. Acad. Sci. U.S.A. 105, 3262–3267.

Sturm, S., Ramesh, D., Brockmann, A., Neupert, S., Predel, R., 2016. Agatoxin-like peptides in the neuroendocrine system of the honey bee and other insects. J. Proteomics 132, 77–84.

Tanaka, Y., Suetsugu, Y., Yamamoto, K., Noda, H., Shinoda, T., 2014. Transcriptome analysis of neuropeptides and G-protein coupled receptors (GPCRs) for neuropeptides in the brown planthopper *Nilaparvata lugens*. Peptides 53, 125–133.

Tyrer, N.M., Davis, N.T., Arbas, E.A., Thompson, K.S., Bacon, J.P., 1993. Morphology of the vasopressin-like immunoreactive (VPLI)neurons in many species of grasshopper. J. Comp. Neurol. 329, 385–401.

Urlacher, E., Soustelle, L., Parmentier, M.L., Verlinden, H., Gherardi, M.J., Fourmy, D., Mercer, A.R., Devaud, J.M., Massou, I., 2016. Honey bee allatostatins target galanin/somatostatin-like receptors and modulate learning: a conserved function? PLoS One. 11, e0146248.

Vallejo, D.M., Juarez-Carreño, S., Bolivar, J., Morante, J., Dominguez, M., 2015. A brain circuit that synchronizes growth and maturation revealed through Dilp8 binding to Lgr3. Science 350 aac6767.

van Kesteren, R.E., Smit, A.B., De Lange, R.P., Kits, K.S., Van Golen, F.A., van der Schors, R.C., De With, N.D., Burke, J.F., Geraerts, W.P., 1995. Structural and functional evolution of the vasopressin/oxytocin superfamily: vasopressin-related conopressin is the only member present in *Lymnaea*, and is involved in the control of sexual behavior. J. Neurosci. 15, 5989–5998.

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., 2010. Neurohormones and neuropeptides encoded by the genome of *Lottia gigantea*, with reference to other mollusks and insects. Gen. Comp. Endocrinol. 167, 86–103.

Veenstra, J.A., 2011. Neuropeptide evolution: neurohormones and neuropeptides predicted from the genomes of *Capitella teleta* and *Helobdella robusta*. Gen. Comp. Endocrinol. 171, 160–175.

Veenstra, J.A., 2014. The contribution of the genomes of a termite and a locust to our understanding of insect neuropeptides and neurohormones. Front. Physiol. 5, 454.

Veenstra, J.A., 2016a. Neuropeptide evolution: chelicerate neurohormone and neuropeptide genes may reflect one or more whole genome duplications. Gen. Comp. Endocrinol. 229, 41–55.

Veenstra, J.A., 2016b. Similarities between decapod and insect neuropeptidomes. PeerJ 4, e2043.

Veenstra, J.A., 2016c. Allatostatins C, double C and triple C, the result of a local gene triplication in an ancestral arthropod. Gen. Comp. Endocrinol. 230–231, 153–157.

Veenstra, J.A., Khammassi, H., 2017. Rudimentary expression of RYamide in *Drosophila melanogaster* relative to other *Drosophila* species points to a functional decline of this neuropeptide gene. Insect Biochem. Mol. Biol. 83, 68–79.

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.

Wang, X., Fang, X., Yang, P., Jiang, X., Jiang, F., Zhao, D., Li, B., Cui, F., Wei, J., Ma, C., Wang, Y., He, J., Luo, Y., Wang, Z., Guo, X., Guo, W., Wang, X., Zhang, Y., Yang, M., Hao, S., Chen, B., Ma, Z., Yu, D., Xiong, Z., Zhu, Y., Fan, D., Han, L., Wang, B., Chen, Y., Wang, J., Yang, L., Zhao, W., Feng, Y., Chen, G., Lian, J., Li, Q., Huang, Z., Yao, X., Lv, N., Zhang, G., Li, Y., Wang, J., Wang, J., Zhu, B., Kang, L., 2014. The locust genome provides insight into swarm formation and long-distance flight. Nat. Commun. 5, 2957.

Wegener, C., Reinl, T., Jänsch, L., Predel, R., 2006. Direct mass spectrometric peptide profiling and fragmentation of larval peptide hormone release sites in *Drosophila melanogaster* reveals tagma-specific peptide expression and differential processing. J. Neurochem. 96, 1362–1374.

Wu, C., Crowhurst, R.N., Dennis, A.B., Twort, V.G., Liu, S., Newcomb, R.D., Ross, H.A., Buckley, T.R., 2016. *De novo* transcriptome analysis of the common New Zealand stick insect *Clitarchus hookeri* (Phasmatodea) reveals tenes involved in olfaction, digestion and sexual reproduction. PLoS One 11, e0157783.

Wu, C., Twort, V.G., Crowhurst, R.N., Newcomb, R.D., Buckley, T.R., 2017. Assembling large genomes: analysis of the stick insect (*Clitarchus hookeri*) genome reveals a high repeat content and sex-biased genes associated with reproduction. BMC Genomics 18, 884.

Xue, J., Zhou, X., Zhang, C.X., Yu, L.L., Fan, H.W., Wang, Z., Xu, H.J., Xi, Y., Zhu, Z.R., Zhou, W.W., Pan, P.L., Li, B.L., Colbourne, J.K., Noda, H., Suetsugu, Y., Kobayashi, T., Zheng, Y., Liu, S., Zhang, R., Liu, Y., Luo, Y.D., Fang, D.M., Chen, Y., Zhan, D.L., Lv, X.D., Cai, Y., Wang, Z.B., Huang, H.J., Cheng, R.L., Zhang, X.C., Lou, Y.H., Yu, B., Zhuo, J.C., Ye, Y.X., Zhang, W.Q., Shen, Z.C., Yang, H.M., Wang, J., Wang, J., Bao, Y.Y., Cheng, J.A., 2014. Genomes of the rice pest brown planthopper and its endosymbionts reveal complex complementary contributions for host adaptation. Genome Biol. 15, 521.