



The contribution of the genomes of a termite and a locust to our understanding of insect neuropeptides and neurohormones

Jan A. Veenstra *

INCIA UMR 5287 CNRS, Université de Bordeaux, Pessac, France

Edited by:

Xanthe Vafopoulou, York University, Canada

Reviewed by:

Åsa M. E. Winther, Karolinska Institutet, Sweden
Herman A. Dierick, Baylor College of Medicine, USA

*Correspondence:

Jan A. Veenstra, INCIA UMR 5287 CNRS, Université de Bordeaux, allée Geoffroy St. Hillaire, CS 50023, 33 615 Pessac Cedex, France
e-mail: jan.veenstra@u-bordeaux.fr

The genomes of the migratory locust *Locusta migratoria* and the termite *Zootermopsis nevadensis* were mined for the presence of genes encoding neuropeptides, neurohormones, and their G-protein coupled receptors (GPCRs). Both species have retained a larger number of neuropeptide and neuropeptide GPCRs than the better known holometabolous insect species, while other genes that in holometabolous species appear to have a single transcript produce two different precursors in the locust, the termite or both. Thus, the recently discovered CNMa neuropeptide gene has two transcripts predicted to produce two structurally different CNMa peptides in the termite, while the locust produces two different myosuppressin peptides in the same fashion. Both these species also have a calcitonin gene, which is different from the gene encoding the calcitonin-like insect diuretic hormone. This gene produces two types of calcitonins, calcitonins A and B. It is also present in Lepidoptera and Coleoptera and some Diptera, but absent from mosquitoes and *Drosophila*. However, in holometabolous insect species, only the B transcript is produced. Their putative receptors were also identified. In contrast, *Locusta* has a highly unusual gene that codes for a salivation stimulatory peptide. The *Locusta* genes for neuroparsin and vasopressin are particularly interesting. The neuroparsin gene produces five different transcripts, of which only one codes for the neurohormone identified from the corpora cardiaca. The other four transcripts code for neuroparsin-like proteins, which lack four amino acid residues, and that for that reason we called neoneuroparsins. The number of transcripts for the neoneuroparsins is about 200 times larger than the number of neuroparsin transcripts. The first exon and the putative promoter of the vasopressin genes, of which there are about seven copies in the genome, is very well-conserved, but the remainder of these genes is not. The relevance of these findings is discussed.

Keywords: vasopressin, neuroparsin, neuropeptide, calcitonin, receptor

INTRODUCTION

Neuropeptides and neurohormones are important regulators of physiological processes that may act at the periphery or within the central nervous system. They are evolutionarily very old and consequently orthologs are present in both deuterostomes and protostomes (Mirabeau and Joly, 2013). Their structures are often insufficiently conserved to establish homology. However, most neuropeptides act through G-protein coupled receptors (GPCRs) and due to the co-evolution between the receptors and their ligands (Park et al., 2002), evolutionary relationships between different peptides can be confirmed using the sequences of the transmembrane regions of the GPCRs they activate (Mirabeau and Joly, 2013). Whereas the structures and the immediate functions, i.e., activation of specific GPCRs, are conserved, their physiological effects may be quite different. For example, the neuropeptide SIFamide is expressed in *Drosophila* exclusively in four brain neurons that modulate sexual behavior (Terhzaz et al., 2007), while in ticks the same peptide acts on the salivary glands

and the hindgut (Šimo et al., 2009, 2013; Šimo and Park, 2014). It is this what makes the study of neuropeptides so interesting, as it allows one to see how the endocrine and nervous systems and pathways may evolve over time, thereby ultimately contributing to a better understanding as to the origin of our own nervous and endocrine systems.

After the initial genome sequence of *Drosophila* (Adams et al., 2000) a significant number of arthropod genomes have followed suit, most of them from holometabolous insect species. The three hemimetabolous insect species with a completely sequenced genome are the louse (Kirkness et al., 2010), the pea aphid (Richards et al., 2010) and the as yet unpublished genome of the kissing bug *Rhodnius prolixus*. An impressively complete neuropeptide transcriptome of the brown plant hopper *Nilaparvata lugens* has also recently been published (Tanaka et al., 2014). All four of those species are very specialized feeders and therefore not necessarily representative of this group. The termite *Zootermopsis nevadensis* and the migratory locust *Locusta migratoria* both

belong to more basal insect groups and as their genomes have recently been sequenced (Terrapon et al., 2014; Wang et al., 2014), it seems valuable to look at the neuropeptides they encode. A preliminary list of neuropeptide genes present in the *Zootermopsis* has already been published (Terrapon et al., 2014), but the actual number of such genes is much larger and here we also identify the neuropeptide GPCRs of this species. *Locusta* is probably the insect species from which the largest number of neuropeptides has been biochemically identified, by the Leuven group (Schoofs et al., 1990a,b,c,d,e, 1991a,b, 1992a,b, 1993a,b, 1994; Paemen et al., 1991a,b; Veelaert et al., 1995; Tawfik et al., 1999; Clynen et al., 2001, 2003a,b,c, 2006; Clynen and Schoofs, 2009) and by many others who identified specific neuropeptides, such as the various adipokinetic hormones (Stone et al., 1976; Siegert et al., 1985; Oudejans et al., 1991; Bogerd et al., 1995; Siegert, 1999), the vasopressin-like peptide (Proux et al., 1987), neuroparsins (Girardie et al., 1989), the corticotropin-releasing factor (CRF)-like diuretic hormone (Kay et al., 1991; Lehmberg et al., 1991), ovary maturing parsin (Girardie et al., 1991), FMRamide (Lange et al., 1994; Hill and Orchard, 2007) and periviscerokinin (Predel and Gäde, 2002). Ion transport peptide was identified from a different migratory locust, *Schistocerca gregaria* (Meredith et al., 1996). Obviously, the locust genome is of particular interest as far as neuropeptides are concerned.

In the present paper, we describe the genes coding several neuropeptides from these two insect species, and the GPCRs of these neuropeptides in the termite. We also discuss the discovery of novel neuropeptide genes, which have not been identified previously, i.e., those encoding SMYamide, calcitonin, tryptopyrokinins, and a salivation stimulating peptide, as well as two locust neuropeptide genes that are of particular interest, the neuroparsin and vasopressin genes.

MATERIALS AND METHODS

The *Zootermopsis* genome was downloaded from <http://www.termitegenome.org> and was analyzed by local BLAST (Altschul et al., 1997) on a desktop PC. The *Locusta* genome was searched using the NCBI web interface and interesting contigs were downloaded for a more detailed analysis. Preliminary prediction of gene models or parts thereof was based on homology with known insect neuropeptides and their receptors in combination with likely intron splice sites using Artemis (Rutherford et al., 2000) as described previously (Veenstra et al., 2012). All short sequence read archives (SRAs, both genome and transcriptome data) for these two species that were available at NCBI were downloaded and the SRA toolkit (<http://www.ncbi.nlm.nih.gov/Traces/sra/?view=software>) was used to extract the data in fasta form, which was then used to build searchable BLAST databases using BLAST+ (Camacho et al., 2009). The initially predicted mRNA sequence was used to search all the transcriptome SRAs for that species. Positive reads were collected and used as input for the Trinity program (Haas et al., 2013) in order to produce transcripts that were used to improve and correct the initial predictions using Artemis. This was an effective method to determine N- and/or C-terminal regions of proteins that could not be predicted by homology alone and in cases where there were gaps either within the contigs or within the scaffolds. The Trinity-Artemis cycle was

repeated as long as the coding regions of the predicted mRNA sequences were incomplete and the predicted mRNA's increased in length. No attempt was made to accurately determine the 3'- and 5'- ends of mRNAs; read-throughs from neighboring genes easily give false positives. For some neuropeptides that were previously isolated and identified from *Locusta* no DNA sequences were found in the genome assembly. In those cases we looked in the genomic SRAs in order to find individual reads that might contain sequences potentially coding these peptides but that had not made it into the assembled genome. We then used these sequences to search the genomic SRAs for similar sequences and attempted to assemble those reads into mini-contigs using both edena (Hernandez et al., 2008) and velvet (Zerbino and Birney, 2008), but this was not very successful.

For comparative purposes, a few other insect genomes both published and unpublished were prospected for specific neuropeptide precursors or GPCRs. Preliminary assemblies of the genomes of *Blattella germanica*, *Ladona fulva*, *Ephemera danica*, and *Lutzomyia longipalpis* were downloaded from <https://www.hgsc.bcm.edu/arthropods/> and those of *Phlebotomus papatasi* and *Diaphorina citri* from <https://www.vectorbase.org/> and <http://www.psyllid.org/>, respectively. Previously published annotated neuropeptide GPCRs from *Drosophila*, *Apis*, *Tribolium* (Hauser et al., 2008), *Bombyx* (Yamanaka et al., 2008), a spider mite (Veenstra et al., 2012), and a plant hopper (Tanaka et al., 2014) were added to the curated *Zootermopsis* GPCRs for the construction of a phylogenetic tree. A few deorphanized protostomian neuropeptide GPCRs were appended to this ensemble. Phylogenetic trees were made after using MUSCLE (Edgar, 2004) and/or Clustal Omega (Sievers et al., 2011) for obtaining sequence alignments on the desktop that were manually inspected and corrected using Seaview (Gouy et al., 2010). Seaview was also used for selecting the conserved protein regions used for making the trees. Both PhyML (Guindon and Gascuel, 2003) and FastTree 2 (Price et al., 2010) were used for making trees. As differences between the two were minimal, only the latter was used for the larger ones. Convertase cleavage site predictions were guided by rules previously described (Veenstra, 2000) and signal peptides were analyzed with Signal P 4.0 (Petersen et al., 2011), but the signal peptide of the *Locusta* ion transport peptides were predicted by the hidden Markov model of Signal P 3.0; the neural network models of both Signal P 3.0 and 4.0 suggest that these peptides do not have a signal peptide. The signal anchor of the termite allatostatin CC was predicted by Signal P 3.0 (Bendtsen et al., 2004) and confirmed by the forecast of a transmembrane region by a hidden Markov model (Krogh et al., 2001).

RESULTS

The data for *Zootermopsis* allowed us to predict complete precursor sequences for virtually all known insect neuropeptides, including such recent additions as the CCHamides, ACP, RYamide, trissin, natalisin, and CNMa (Roller et al., 2008; Hansen et al., 2010; Hauser et al., 2010; Ida et al., 2011a,b, 2012; Jiang et al., 2013; Jung et al., 2014). In a number of cases, exons were lacking either completely or partially from the genome assemblies, but all these gaps could be filled and/or corrected with transcriptome data and in a few cases those corrections

were confirmed with the genomic reads. The predicted transcripts for the neuropeptide GPCRs seem similarly very complete. The *Locusta* genome is much bigger, giving rise to much bigger introns, and a couple of genes appear to be represented by different alleles. Furthermore, the *Locusta* transcriptome data is much more limited and for some genes non-existent. Thus, the pyrokinin, and vasopressin genes are known to be expressed exclusively or predominantly in the suboesophageal ganglion (Rémy and Girardie, 1980; Clynen and Schoofs, 2009) and we expect tryptopyrokinins to be expressed also mainly in this ganglion. However, none of these genes is represented in the transcriptome SRAs, suggesting the absence of this ganglion from the transcriptome data. The difficulty to assemble repetitive sequences using a short read sequencing technology also affected the assembly of the *Locusta* genome more than the *Zootermopsis* genome, particularly in the case of the vasopressin genes, present in various copies, as well as some genes that code highly similar or identical multiple copies of the same peptide, such as the calcitonin gene.

The major reason for the less complete *Locusta* sequences is the lack of sufficient transcriptome data; the number of reads was often insufficient for Trinity to build transcripts. The use of individual transcriptome reads and publicly available expressed sequence tags (ESTs) from *Locusta* (Kang et al., 2004; Clynen et al., 2006; Ma et al., 2006; Chen et al., 2010) allowed us to fill some gaps, but many remain. It is for this reason that we did not attempt to predict the neuropeptide GPCRs in this species (in a few cases we used partially constructed GPCRs for phylogenetic tree analysis). Although some genes are thus incompletely present in the *Locusta* genome assembly, most of the neuropeptides previously identified from the migratory locust were found in the genome sequences, if not in the assembly itself, at least in the genomic SRAs. Thus, in spite of the fragmentary nature of the *Locusta* genome (1,397,492 contigs vs. 64,771 for *Zootermopsis*) and the limited amount of transcriptome data, we were able to deduce the complete sequences of a important number of *Locusta* neuropeptide precursors, while for others at least a significant part of their precursor was found.

In a few cases the sequences predicted by the *Locusta* genome are slightly different from those previously reported. As most these differences are easily explained by experimental errors such as contamination of first amino acid residues in Edman degradation and/or the same or very similar molecular masses in mass spectrometry, it seems likely that the genomic sequences are the correct ones. The genomic sequences also allowed us to confirm the sequence of locustamyoinhibin (Schoofs et al., 1994) as allatostatin C.

We found a total of 59 and 63 different transcripts for neuropeptides and neurohormones in the genomes of *Zootermopsis* and *Locusta*, respectively. In the locust the actual number is probably even higher as there are likely to be several transcripts encoding the vasopressin-like peptide, but these could not be reliably identified. As several of these transcripts are predicted to produce more than one peptide the number of neuropeptides produced by these insects is much larger. The predicted precursors of these genes are illustrated in the Supplementary Figures 1, 2, and their sequences are listed in Supplementary Tables 1, 3.

More important than the actual number of different peptides produced is the question as to how many different messages they can transmit. It is plausible that the number of different neuropeptide receptors will give a more accurate answer to that question. In the *Zootermopsis* genome a total of 65 putative neuropeptide GPCRs were identified. Although a large number of them can be deorphanized *in silico* by their close structural similarity to deorphanized insect neuropeptide GPCRs, others are orphans, which may not be neuropeptide receptors. On the other hand some neuropeptides are known to act on receptors that are not GPCRs, such as insulin, eclosion hormone and at least one of the peptides encoded by the NPLP1 gene in *Drosophila* (Chang et al., 2009; Overend et al., 2011; Vogel et al., 2013). The predicted *Zootermopsis* neuropeptide and neurohormone GPCRs are presented in Supplementary Table 2.

On phylogenetic trees (Supplementary Figures 3, 4), most of the GPCRs cluster with related GPCRs from other species and in many cases the receptor can be deorphanized *in silico* due to the fact that for one or more of the same group the ligand has been determined experimentally. It is interesting to see that some receptors are only present in *Zootermopsis*, *Nilaparvata*, and/or *Tetranychus*, suggesting that the particular receptor and perhaps its ligand may have been lost in holometabolous insect species. In other cases, there may be two receptors for the same neuropeptide. For example, in the case of *Zootermopsis* GPCR A51 and *Nilaparvata* GPCRs A38 and A39 it is tempting to speculate that they might be alternative NPF receptors, while *Zootermopsis* GPCR A11 and *Nilaparvata* GPCR A5 might be SIFamide (or SMYamide?) receptors.

ACCESSORY GLAND MYOTROPIN II

An unusual peptide identified in Leuven concerns the accessory gland myotropin II, which was identified from male accessory glands in *Locusta* (Paemen et al., 1991b). The sequence of this peptide is encoded on an EST (Accession # GO240796). The same sequence can also be deduced from genomic and transcriptomic data and shows a precursor that does not look like a neuropeptide precursor as it is missing typical neuropeptide convertase cleavage sites (Figure 1).

ADIPOKINETIC HORMONE

Adipokinetic hormone (AKH) was identified from migratory locusts as a hormone from the corpora cardiaca increasing

AVARALAADAVARAVAASKAANAAT
 NAARAVGAHRFAAEDFGALDTAFDG
 PHDAAAMAFRAYDPFRQPTTPRPID

FIGURE 1 | *Locusta* accessory gland myotropin. Deduced amino acid sequence from EST GO240796.1 and genome assembly. Both the genome assembly and the EST have a few more amino acids at the N-terminus, but they differ, probably due to the insertion or deletion of a nucleotide. The sequence shown here is identical between the genome assembly and the EST. It clearly shows that the isolated active peptide (Paemen et al., 1991b), shown in blue, is not flanked by classical neuropeptide convertase cleavage sites.

hemolymph lipid during flight (Beenackers, 1969; Mayer and Candy, 1969). It is made in the glandular cells of the corpora cardiaca and released during flight. After the identification of the decapeptide AKH I (Stone et al., 1976), the octapeptides AKH II and AKH III were identified (Siegert et al., 1985; Oudejans et al., 1991). The closely related peptide ACP (AKH/corazonin-related peptide) was initially also considered to be an AKH (Siegert, 1999), but has now been shown to be a different peptide with its own receptor (Hansen et al., 2010). Unlike AKH it is not expressed in the glandular cells of the corpora cardiaca but by neurons in the nervous system (Hansen et al., 2010; Patel et al., 2014). In the *Locusta* genome the genes encoding AKH I, II, and III as well as ACP were identified. Interestingly, a fourth AKH gene was encountered, which we called AKH IV. Based on the number of transcriptome reads, AKH IV may be even less expressed than AKH III (3 vs. 6 reads), which is much less abundantly expressed than AKHs I and II (586 and 21 reads, respectively). In the *Zootermopsis* genome only a single AKH gene was found, as in many other insect species.

ALLATOSTATIN CC

Allatostatin CC has strong sequence similarity to allatostatin C and based on this was expected to activate the same receptors (Veenstra, 2009), as now has been confirmed for the beetle *Tribolium castaneum* (Audsley et al., 2013). Allatostatin CC does not appear to be a typical neuropeptide and it was suggested that it might function as a paracrine secretion, not only secreted by neurons but also by other cells (Veenstra, 2009). The expression of this peptide by hemocytes (Accession # FP353238.1, FP357967.1, FP358784.1, FP359038.1, FP352222.1, FP352474.1) supports the hypothesis that the peptide is indeed made by non-neuroendocrine cells. In the *Drosophila* species as well as tsetse flies allatostatin CC lacks a signal peptide and instead has a signal anchor, reinforcing the notion that it is released locally and not systemically. In the termite the gene similarly encodes a signal anchor, but in the locust it is predicted to have a classical signal peptide. Therefore, it appears to be a juxtacrine secretion also in termites. A peptide secreted as a paracrine or juxtacrine rather than an endocrine presumably would not need as high an affinity for its receptor as a hormone. This is indeed what was found for *Tribolium* (Audsley et al., 2013).

CALCITONIN AND CALCITONIN-LIKE DIURETIC HORMONE

Anything that stimulates the excretion of water in insects could potentially be used as the basis for designing a novel pesticide and putative insect diuretic hormones have therefore received a lot of attention. One of the identified diuretic hormones has been called the calcitonin-like diuretic hormone but is also known by the abbreviation DH31. This peptide was first identified from the cockroach *Diploptera punctata* (Furuya et al., 2000) and genes coding such a peptide are present in all insect genomes. The *Drosophila* and *Rhodnius* homologs have been shown to activate B-type GPCR's that are related to the vertebrate calcitonin receptors (Johnson et al., 2005; Zandawala et al., 2013). This hormone is generally considered to be the insect calcitonin ortholog. However, the genomes of both *Zootermopsis* and *Locusta* contain another gene that encodes peptides showing even stronger

similarity to calcitonin, as they contain a predicted disulfide bridge.

Analysis of several other arthropod genomes, ESTs and transcriptome shotgun assemblies (TSAs) show this gene to be very generally present in arthropods (Supplementary Figure 5). Comparing the different encoded arthropod calcitonins reveals that they can be separated into two distinct classes, which we have called calcitonin-A and calcitonin-B (Supplementary Figure 6). In the termite this gene produces two different transcripts (Figure 2). The first one, the A-transcript, codes for a single calcitonin-A, while the B-transcript encodes three calcitonin-B peptides. It must be noted that even though both types of calcitonin have a predicted N-terminal disulfide bridge and a C-terminal Pro-amide sequence, the remainder of the sequences show striking differences (Figure 3). The as yet unpublished genomes of the dragonfly *Ladona fulva*, the German cockroach *Blattella germanica* and the may fly *Ephemera danica* have very similar genes, although the number of copies of calcitonin-B on the second transcript is variable. Both types of peptides are also produced in phasmids as TSAs from various species attest.

The *Locusta* calcitonin gene has a gap in its sequence and no calcitonin-A peptide was found in the genome assembly, however in the genomic sequence reads a sequence was found that encodes a calcitonin-A-like peptide, i.e., CYIGGRMGGCDYQDLKQAQGEDQHLNSIDSPGKR.

Although it is notably absent from *Drosophila* and other flies as well as mosquitoes, aphids, *Rhodnius*, and Hymenoptera, it is present in the sand flies *Phlebotomus papatasi* and *Lutzomyia longipalpis* as well as beetles and Lepidoptera. Many of the predicted arthropod calcitonin precursors contain several copies of this neuropeptide, while the *Tribolium* genome has even two

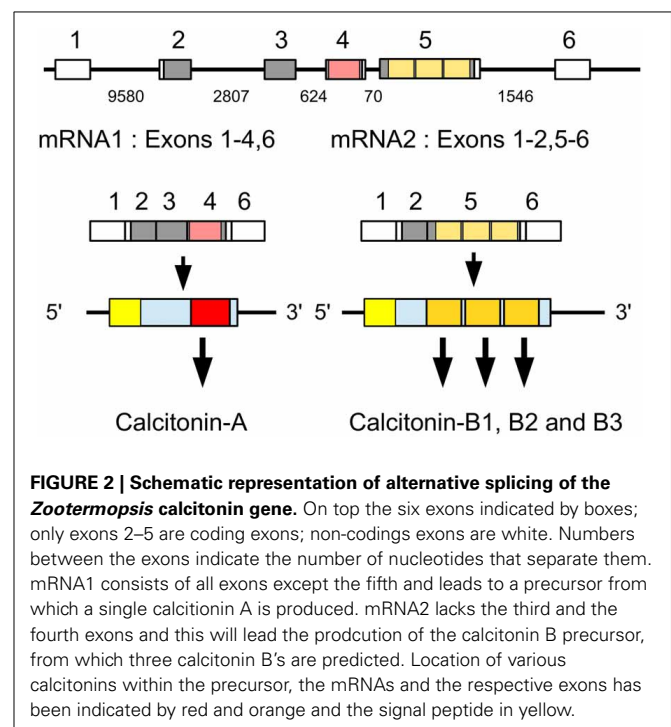


FIGURE 2 | Schematic representation of alternative splicing of the *Zootermopsis calcitonin* gene. On top the six exons indicated by boxes; only exons 2–5 are coding exons; non-coding exons are white. Numbers between the exons indicate the number of nucleotides that separate them. mRNA1 consists of all exons except the fifth and leads to a precursor from which a single calcitonin A is produced. mRNA2 lacks the third and the fourth exons and this will lead the production of the calcitonin B precursor, from which three calcitonin B's are predicted. Location of various calcitonins within the precursor, the mRNAs and the respective exons has been indicated by red and orange and the signal peptide in yellow.

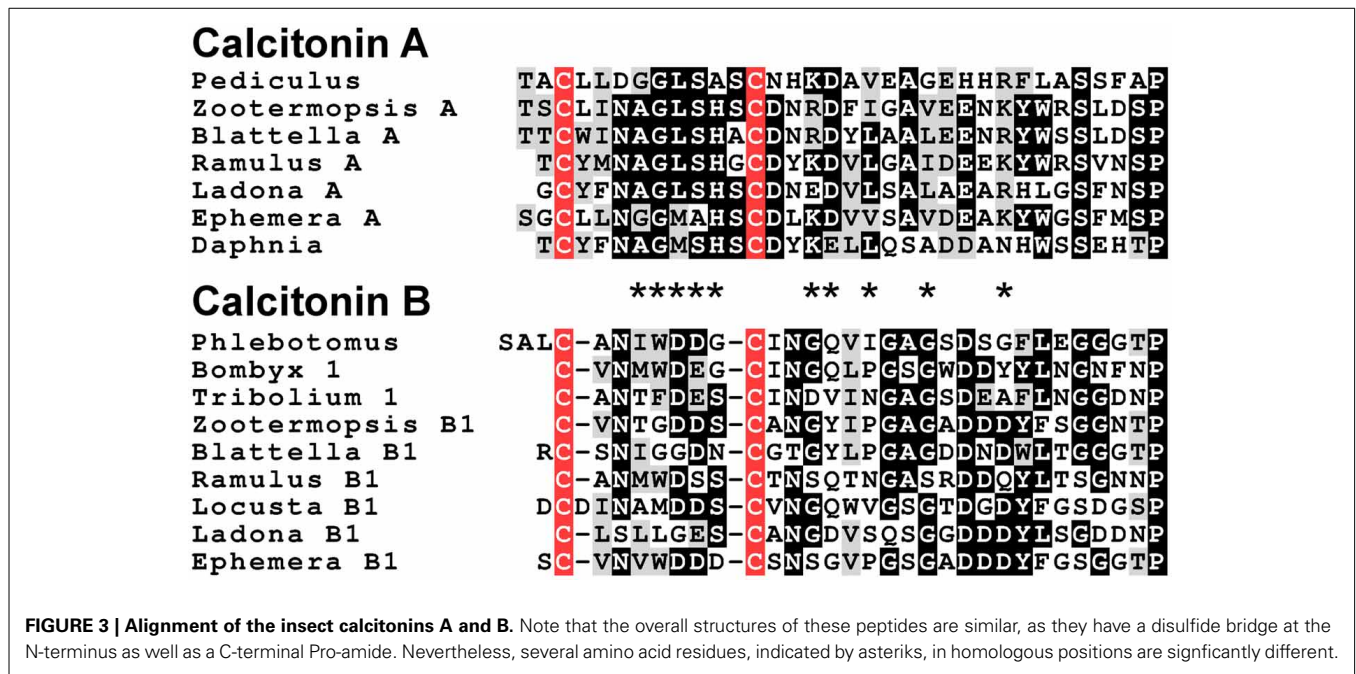


FIGURE 3 | Alignment of the insect calcitonins A and B. Note that the overall structures of these peptides are similar, as they have a disulfide bridge at the N-terminus as well as a C-terminal Pro-amide. Nevertheless, several amino acid residues, indicated by asterisks, in homologous positions are significantly different.

different genes (Supplementary Figure 5). Transcriptome data show calcitonin B to be expressed in the insect midgut.

The B-type GPCRs can be expected to include the receptor(s) for the arthropod calcitonins described here. Of the B-type GPCRs that have not been deorphanized in insects, there is one that appears to be the best candidate. This group of receptors is characterized by the B3 receptor from *Bombyx* (Yamanaka et al., 2008) and the receptor identified as 72 from *Tribolium* (Hauser et al., 2008) and corresponds to cluster A of Cardoso et al. (2014). We thus looked for homologs of this receptor in other arthropod genomes and found such receptors only in those species that also have a calcitonin gene. Given our interest in the *Locusta* receptors, we included two partial homologs from this species. The results suggests that in basal insect groups there are two calcitonin receptors, like in *Locusta*, *Zootermopsis*, and the stick insect *Ramulus artemis* (Supplementary Figure 4). Interestingly, phylogenetic trees for the calcitonin-A peptides correspond completely with one of these receptors, while tree for the calcitonin-B peptides corresponds perfectly with the other (Figure 4).

CAPA PEPTIDES, PYROKININS, PERIVISCEROKININS, AND TRYPTOPYROKININS

The nomenclature of these peptides is not always very clear, in large part because sequences of these peptides are rather similar and peptides have often been named according to the tissue they have been identified from or the gene on which they are encoded. Analysis of their receptors in *Drosophila* has made it clear that pharmacologically one can distinguish three different types of peptides and this has been confirmed in other species (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). The pyrokinins and tryptopyrokinins have C-terminal consensus sequences of FXPRLamide and MWFGPRLamide, respectively, while that of the periviscerokinins is FPRVamide.

The tryptopyrokinins seem to be a relatively recent evolutionary innovation, as they have so far not been found in non-insect arthropods. The close association of their receptors with those for pyrokinins on evolutionary trees for GPCRs suggests the same evolutionary innovation (Paluzzi and O’Donnell, 2012). In a few holometabolous insect species, tryptopyrokinins are known to be expressed specifically by neuroendocrine cells in the labial neuromere of the suboesophageal ganglion. At least in the case of *Drosophila* this is achieved by differential processing of the capa precursor (Wegener et al., 2006) and it has previously been suggested that such a differential processing might also occur in *Bombyx* (Veenstra, 2000).

Insects generally have two genes coding for these peptides, a capa gene coding for periviscerokinins and often a tryptopyrokinin, and a pyrokinin gene that encodes pyrokinins and often a tryptopyrokinin. Both *Zootermopsis* and *Locusta* have a third type of gene that codes for tryptopyrokinins and in the case of the termite a single pyrokinin. Interestingly, the locust has four such genes (Figure 5).

The *Locusta* pyrokinin gene was one of the most difficult genes to analyze. Some of the genomic reads suggest that there may be two alleles represented in the genomic reads while transcriptome reads from the suboesophageal ganglion, where this gene is normally predominantly expressed, are lacking from the expression SRAs. Most, but not all, of the pyrokinins previously identified from this species (Schoofs et al., 1990e, 1991a, 1992b, 1993b; Clynen et al., 2003a,b) were found in individual reads of the *Locusta* genome.

The periviscerokinin transcript is probably also incomplete, the part found encodes only two peptides, periviscerokinin, and TSSLFPHPRamide, both previously identified (Predel and Gäde, 2002; Clynen et al., 2003a,b). From the distribution of these peptides and that of DGAETPGAAASLWFGPRVamide and GLLAFPRVamide (Clynen and Schoofs, 2009), one would expect

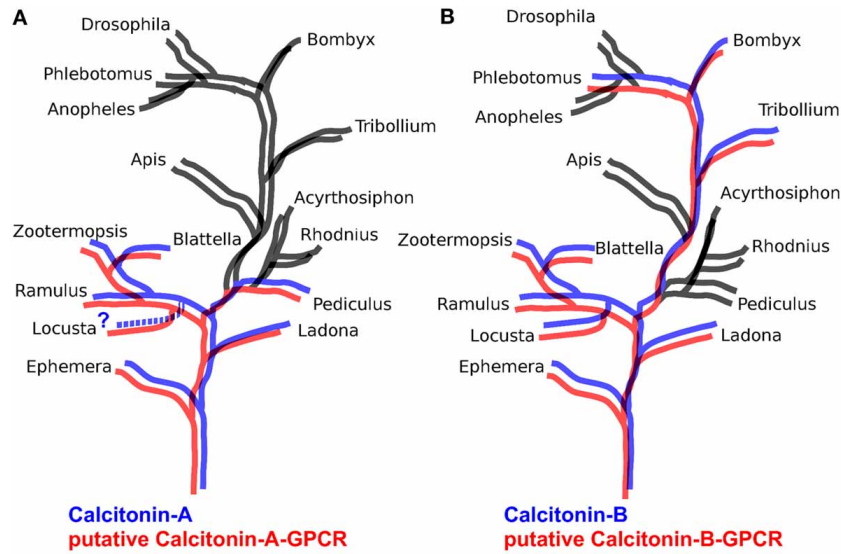


FIGURE 4 | Schematic phylogenetic trees of insects illustrating the presence or absence of Calcitonin A and B in blue and their putative receptors in red. When a peptide or a receptor is absent,

branches are black. Note that in (A) calcitonin A is present in the same species as its putative receptor while in (B) the same holds for calcitonin B.

Zootermopsis

MNTQLLCSQHILIWFLVSA AVNNLCSVNTAMEETRIPGSDYGTVDGQRPSCNK^KSPAADVINLSVEFNMDHKNEKPNGTAE SQPHIKLEDKIMNFIENWSIPAEIKNKT
 TTSLFTGQGDVETLMLQPTSNRMWYAPRL^{GR}D^KSRVNDEQVKQESARGGMWFGPRL^{GR}DKKSSINDEQVKQEFAYGGMWFGPRL^{GR}DKKSNVNDEQVKQEVAYGGMWFG
 PRL^{GR}DKKSNVHDEQVKQEV AHGGMWFGPRL^{GR}EKKS YENEWPKKPRPFWMFLETTNKPREYDFLPQFDQDTVSEIYATAVNNITSELENYSSQLSDDIGIMKYSPSSVR
 FSSQLGHKQKATNVIAYVVPQEVQMTSEDKNK^KPEYDFT^{PR}L^{GR}ELYEKGLTLASKSLRLPQIGHNTGKIYVYFQQD

Locusta 1

MPRGAQLFLLALVTAARVLDAR^AAESSRAGSDAQSSHQDGRGASHDRNEESNELNDENRSDGDAQDAT^{RR}GREMQEFVSHAGGPIGNGEEDQWLALADGSYIPAEAVRELQ
^RGPTLSENTGLWFGPRY^{GR}RTSCGENVPLKWL^{SQ}VE^{KR}AAKQPALWFGPRV^{GR}SLDEEPKGEWRDDDKGLKDGSSQRQD^RSAQPGLWFGPRV^{GR}SDAQVDDMLWFGPRHE
 RSVDTDKQDLYDDEVAMRDQ^RGAKHPGLWFGPR^{GR}

Locusta 2

MESSMDVLPVALLTLAALLFSVET^SEVLHEVNDLNGSNRDSGEQTPSAVLSRS^{KN}NSGAHV^DKWFRPDTGAGEDYWNALFRED^AEDHLSKPDASTVDRL^ENVVPTAET^{PR}LS
 VAQHESRHSEGSDDHNDSEPEVWIGQHS^GSIPEPGTWFGPRIG^RSHTEPGLWFGPRY^{GR}SYPEPGMWFGPRV^{GR}GHPEPGMWFGPRV^{GR}SQAEPGVWFGARIG^RNQPEPGT
 WFGARIG^RSHPEPGMWFGPRV^{GR}SHPEPGTWFGPRIG^RSHSEPELRM

Locusta 3

...RNG^{RR}QPD^TALWFGPRV^{GR}RRMQHIQPEASEFFRPD^VRRSGPEHSLWSDTQLRQGDPEPPLHVRISNPKQNLWFR^{TD}RRQSENILQFGPQV^{GR}NAESALWFGPRV^{GR}S
 YLESSLSDERYVQHRNLEPLGWSNNGIQTTPSMDVANRFP^GDLWSGTDLDR^TKHETALWFGPRIG^RSNPETNLWFGPRV^{GR}SHPETSQSFQYARHDNVETGLWPD^{DF}VQ
 SRPTTLEDVT^{KR}NPERN^WSVNGVGLTKHETTLWFGPRIG^RSNPETNLWFGPRV^{GR}SHPETSQSFQYARHDNVETGLWPD^{DF}VQ
 HETALWFGPRIG^RSNPD^THLWFGPRV^{GR}SQSLASEQLR

Locusta 4

MLRRTHS^LAVAVTVVAISAF^{AV}SA^AAKPREMHSKNVVVQRC^IQTGFSDGNGNIRSV^PETSLWFGPRIG^RSNLETALLFGR^CVGC^SHPGTSQ^{CC}EPLAR^RSGSEDSLGYSSPEPP
 HDNRKINPESGLWFD^{TL}ARRSGSEDSLGYSSPEPPHDNRKINPESGLWFD^{RN}GRQPD^TALWFGPRV^{GR}RRMQHIQPEASEFFRPD^VRRSGPEHSLWSDTQLRQGDPEPPLH
 VRISNPKQNLWFR^{TD}RRQSENILQFGPQV^{GR}NAESALWFGPRV^{GR}SYLESNLSDERYIQHRNLEPLGWSNNGIQTTPSMDV^TNRFPGRDLWSGTDLDR^TKPRYGLD^HG

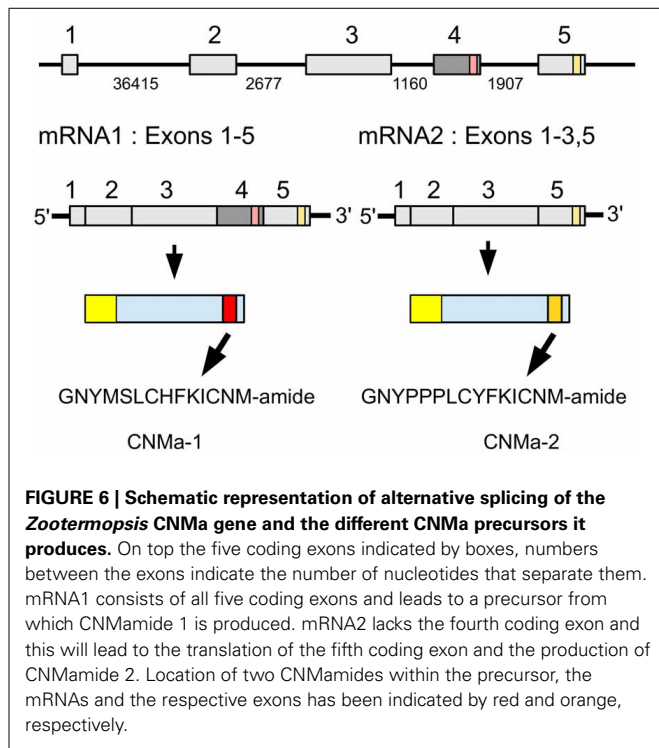
FIGURE 5 | Tryptopyrokinins. The predicted *Zootermopsis* tryptopyrokinin precursor is expected to yield at least four tryptopyrokinins and one pyrokinin. Of the four predicted *Locusta* tryptopyrokinin precursors only one contains a pyrokinin.

to find the latter two also to be expressed by the periviscerokinin gene, but neither genomic nor transcriptomic reads encoding these peptides were found.

CNMa

CNMa was very recently identified as a novel insect neuropeptide (Jung et al., 2014). The *Zootermopsis* CNMa gene produces by alternative splicing two types of mRNA that differ by the inclusion or exclusion of the third coding exon (Figure 6). When the

third coding exon is included in the mature mRNA it should lead to the production of GNYMSLCHF^KICNMamide. In that case the fourth coding exon will not be translated, since the third exon contains an in-frame stop codon near its end. When the third coding is excised from the mature mRNA, the fourth coding exon is predicted to produce the alternative CNMamide, i.e., GNPPPLCY^FKICNMamide. The number of transcriptome reads specific for the first transcript is much smaller (75) than that for the second (931) and it thus seems plausible that the second



peptide is produced in larger amounts than the first. As in several Hymenoptera and *Nilaparvata* (Jung et al., 2014; Tanaka et al., 2014) the termite genome also contains two CNMamide receptor homologs that on phylogenetic trees cluster tightly with the de-orphanized CNMamide receptors and hence are most likely the *Zootermopsis* receptors for these neuropeptides (Supplementary Figure 3). Only a single CNMa peptide was found encoded by the orthologous *Locusta* gene.

CRF-LIKE DIURETIC HORMONE AND OVARY MATURATING PARSIN (OMP)

The insect CRF-like neuropeptides are other putative diuretic hormones and both the *Locusta* peptide and its *Zootermopsis* ortholog have been isolated and sequenced (Kay et al., 1991; Lehmborg et al., 1991; Baldwin et al., 2001). Ovary maturing parsins have been found exclusively in locusts, both *Locusta* and *Schistocerca* (Girardie et al., 1991, 1998). This hormone has been reported to stimulate oocyte growth and induce vitellogenin synthesis (Girardie et al., 1992, 1998; Girardie and Girardie, 1996). These two peptides are produced by the same neuroendocrine cells in the brain (Tamarelle et al., 2000). Recently the cDNA for *Schistocerca* CRF-like diuretic hormone was found to encode also OMP (Badisco et al., 2011; Van Wielendaele et al., 2012) and it is thus not surprising that the *Locusta* transcript similarly encodes both hormones. The diuretic hormone precursor from *Zootermopsis* does not contain a sequence that is recognizable as a homolog of ovary maturing parsin, as is the case with precursors of the CRF-like peptide in all other insect species (Van Wielendaele et al., 2012). RNAi knockdown of the common precursor of these hormones in *Schistocerca* had the opposite effect of what one would expect based on the reported biological activity

of OMP. Inhibition of the production of OMP and the CRF-like diuretic hormone accelerated oocyte growth, while injection of the diuretic hormone inhibited it (Van Wielendaele et al., 2012). As far as the CRF-like peptide goes, these results are consistent, but they seem to contradict previous reports on OMP. It might indicate that ovary maturing parsin does not act like a neuropeptide through a GPCR, but that it has other, indirect effects at the doses injected. More work will be needed to resolve this question.

ECLOSION HORMONE

As in some other species, both the locust and the termite have two genes encoding an eclosion hormone. It seemed interesting to know whether this gene duplication is a relatively recent or a more ancient event. We used the BLAST website at NCBI to find DNA sequences that might code for such hormones and made a phylogenetic tree of the sequences. Results suggest that in some cases, such as *Locusta*, the butterfly *Danaus plexippus* and mosquitoes the duplication is relatively recent, but in the case of the termite, *Tribolium*, *Nilaparvata*, and the psyllid *Diaphorina citri* it looks like the second eclosion hormone gene is quite old (Supplementary Figure 7).

ELEVENIN

Elevenin is a neuropeptide encoded on a mRNA initially identified from the L11 neuron in *Aplysia californica* (Taussig et al., 1984). Although the actual peptide has never been identified, similar neuropeptide precursors are common in mollusks and the peptide has been called elevenin (Veenstra, 2010a). Such neuropeptide precursors have also been deduced from the genome of *Caenorhabditis elegans* (Yamada et al., 2010) and a spidermite (Veenstra et al., 2012) as well as the transcriptome of the plant hopper *Nilaparvata* (Tanaka et al., 2014). Both the *Locusta* and *Zootermopsis* genomes contain an elevenin gene, which turns out to be present in quite a few species (Figure 7). Interestingly, a very large number of *Periplaneta* elevenin ESTs come from the testis of this species (Chen et al., 2013). As there is an intron in the middle of the sequence coding elevenin it is difficult to find it in genomes, particularly in those of holometabola where primary sequences appear more variable. Nevertheless, as the structure of the elevenin gene is well-conserved (Supplementary Figure 8), there is no doubt that the predicted peptides are authentic elevenin orthologs. Although we did find a single Coleopteran sequence from *Dastarcus helophoroides*, we were unable to find it in the *Tribolium* genome.

FMRFamide AND MYOSUPPRESSIN

The termite myosuppressin gene is similar to that of other insect species, but in the locust the RNA produced from this gene is alternatively spliced into two mRNAs that each produce a different myosuppressin (Figure 8). Most of the mRNA (62 vs. 8 transcriptome reads) codes for PDVDHVFLRFamide, the peptide initially identified from *Schistocerca* (Robb et al., 1989) and later also from *Locusta* (Schoofs et al., 1993a; Peeff et al., 1994). The other mRNA codes for EDVGHVFLRFamide, a peptide almost identical to ADVGHVFLRFamide which was identified as a second myosuppressin from *Locusta* (Peeff et al., 1994).

Peripatopsis	TDETTIRPLD	CRKTVF	HFDD	CRGIS	AKSNEF
Lottia	RKLD	CRKTVF	FAPK	CRGVAA	
Aplysia	RPRID	CRKTVF	FAPA	CRGVSA	
Idiosepius	pQGKFELTKKLN	CRRF	IF	FAPRCRGVAA	
Doryteuthis	pQNKFELTKKLN	CRRF	IF	FAPRCRGVAA	
Crassostrea	pQIRARRRR	CRKTVF	FAPRC	CRGVAA	
Capitella	WRID	CRKTVF	FAPK	CRGVAA	
Alvinella	WRVD	CRKTVF	FAPK	CRGVSA	
Caenorhabditis	LD	CRKFS	FAPA	CRGIML	
Ascaris	LD	CRKFS	FAPA	CRGIML	
Pristionchus	LD	CRKFS	FAPA	CRGLLL	
Tetranychus	YPVD	CRKTVF	FAPT	CRGVIA	
Daphnia	RDLD	CRKTVF	FAPM	CRGATI	
Schistocerca	pQKRVD	CRMYF	FAPIC	CRGIMT	
Locusta	pQKRVD	CRMYF	FAPIC	CRGIMTK	
Periplaneta 1	RAPVD	CRKYP	FAPRC	CRGVDPDSG	
Periplaneta 2	pQVID	CRKYP	FAPRC	CRGVDPDSG	
Zootermopsis	VPGLD	CRKTVF	FAPK	CRGVAA	
Pediculus	KRPVD	CRKTVF	FAPIC	CRGVAA	
Nilaparvata	VN	CRKTVF	FAPV	CRGVAA	
Rhodnius 1	pQGENENGT	VV	CRKYP	FAPV	CRGAGA
Rhodnius 2	pQVIGKNGQI	VD	CRKYP	FAPV	CRGAGA
Bombus	ERSESVN	CEAF	PHQM	CRGSM	
Apis	ESVNC	ELYP	FHHT	CRGTMS	
Acromyrmex	VD	CEQEP	YNPA	CRGSQA	
Harpegnathos	VD	CEKYP	YHPS	CRGTMK	
Pogonomyrmex	VD	CEQEP	YHPS	CRGSQA	
Atta	GAVD	CEQEP	YNPA	CRGSQA	
Camponotus	VN	CERNP	YHPS	CRGSQA	
Linepithema	VD	CELYP	YHSS	CRGVMT	
Solenopsis	AN	CEQEP	FHPS	CRGTOA	
Dastarcus	HPNKGRLIN	CRKVVY	IF	CRGVAA	

FIGURE 7 | Deduced amino acid sequences of protostomian elevenins.

Note that the peptide is not only present in insects, but also in velvet worms (*Peripatopsis*), mollusks (*Lottia*, *Aplysia*, *Idiosepius*, *Doryteuthis*, *Crassostrea*), polychaetes (*Capitella*, *Alvinella*), nematodes (*Caenorhabditis*, *Ascaris*, *Pristionchus*), crustaceans (*Daphnia*), and arachnids (*Tetranychus*). Also note that the sequences of the peptide in holometabolous insect species is less conserved.

Both species also have a gene encoding FMRFamides. In the termite, it encodes six copies ending in NFIRFamide and four copies in NFVRFamide plus a number of other peptides. The *Locusta* FMRFamide transcript could not be reconstructed completely, but the part that was found codes for one FIRFamide and four FLRFamides, including GSERNFLRFamide. The latter peptide has a molecular mass very similar to LWENLRFamide, the sequence proposed for a peptide isolated from the gut of *Locusta* (Hill and Orchard, 2007) as well as GQERNFLRFamide, identified from the ventral nerve cord (Lange et al., 1994). Six transcriptome reads show that this sequence is part of a gene that is expressed.

INSULIN- AND RELAXIN-LIKE PEPTIDES

A single insulin gene was found in the migratory locust and six such genes were identified in the termite. One of the termite genes codes a peptide homologous to the *Drosophila* insulin-like peptide (dilp) 7. This peptide is significantly different from other insect insulin-like peptides and is predicted to bind a GPCR activated by relaxin-like hormones, rather than the classical insulin tyrosine kinase receptor (Veenstra et al., 2012). We have therefore called it relaxin-like, rather than insulin-like. It has two different transcripts that are predicted to yield very similar peptides.

NEUROPARSIN

The *Zootermopsis* genome contains a neuroparsin gene with three coding exons that lead to the production of a single mRNA. The *Locusta* neuroparsin gene produces five different mRNAs (Figure 9). Transcript 5 is the one that encodes neuroparsin as it has been isolated from the corpora cardiaca

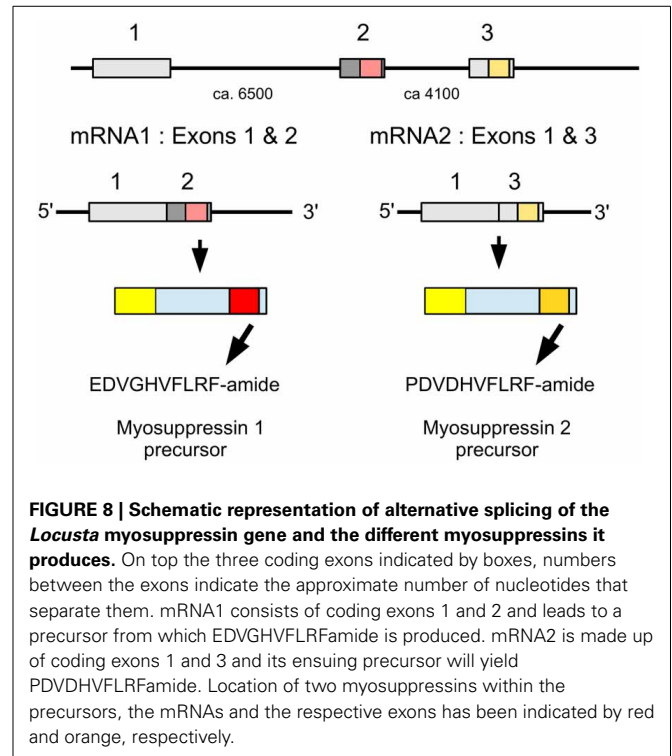
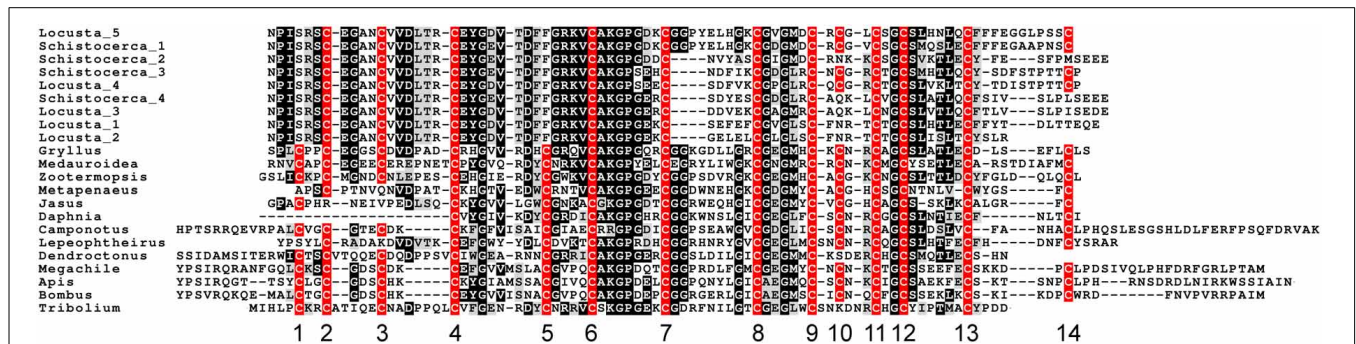
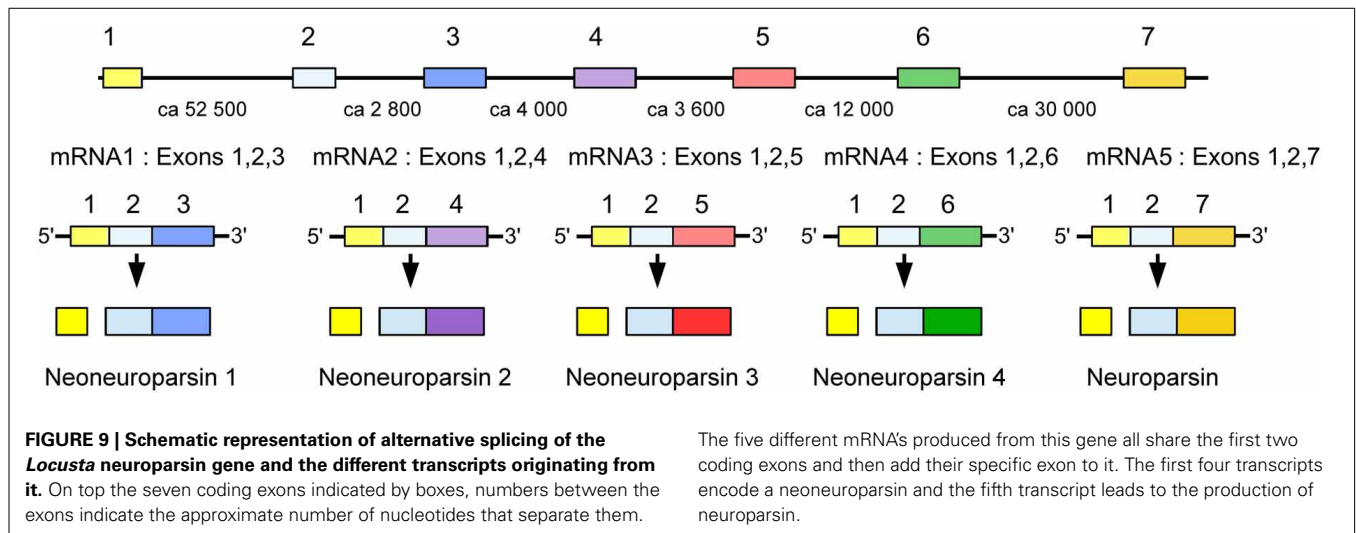


FIGURE 8 | Schematic representation of alternative splicing of the *Locusta* mysuppressin gene and the different mysuppressins it produces. On top the three coding exons indicated by boxes, numbers between the exons indicate the approximate number of nucleotides that separate them. mRNA1 consists of coding exons 1 and 2 and leads to a precursor from which EDVGHVFLRFamide is produced. mRNA2 is made up of coding exons 1 and 3 and its ensuing precursor will yield PDVDHVFLRFamide. Location of two mysuppressins within the precursors, the mRNAs and the respective exons has been indicated by red and orange, respectively.

(Girardie et al., 1989; Lagueux et al., 1992). The presence of five transcripts is similar to the situation in *Schistocerca gregaria* where four different neuroparsin mRNAs have been identified (Janssen et al., 2001; Claeys et al., 2003). The *Locusta* neuroparsin 5 transcript and the *Schistocerca* neuroparsin 1 transcript code for the neuroparsins that were isolated from the corpora cardiaca.

It is of interest to note that the proteins predicted from the *Locusta* neuroparsin transcripts 1–4 are lacking a four amino acid sequence in the middle of the molecule (Figure 10) and this is also the case for the other neuroparsin transcripts identified from *Schistocerca* (Janssen et al., 2001; Claeys et al., 2003). Although it is clear from earlier work that the neuroparsin primary sequences are very variable, notably in holometabolous insect species (Veenstra, 2010b), the neuroparsin core appears much better conserved (Figure 10). It is for this reason that we prefer to give the proteins predicted from these novel transcripts a different name and call them neoneuroparsins to distinguish them from the hormones identified from the corpora cardiaca.

Analysis of the expression of the various *Locusta* neuroparsin transcripts reveals that transcripts 1, 2, and 3 have much higher expression levels than those of 4 and 5. In all the *Locusta* transcriptome SRAs combined, there are 520,499,475 reads, of which 115,938 correspond to exons 1 and/or 2 (the exons that are common to all neuroparsin transcripts), 127,664 correspond to exons 3, 4 or 5, 116 to exon 6 and 203 to exon 7. In the SRA specific for the nervous system of *Locusta* (SRR167712) there are 28,946,371 individual reads, with 11,936 for exons 1 and/or 2, 12,637 for exons 3, 4 or 5, 33 for exon 6 and 66 for exon 7. These numbers show that a very large majority of neuroparsin transcripts is coding for one of the predicted neoneuroparsins 1, 2 or 3, whereas



residues is more variable near the N- and C-terminal regions of these proteins, it is very constant in the core of the protein between cysteine residues 4 and 13. However, in the neoneuropepsin four residues are lacking between cysteines 7 and 8, suggesting that the three dimensional structures of these molecules may be different from that of the neuropepsins.

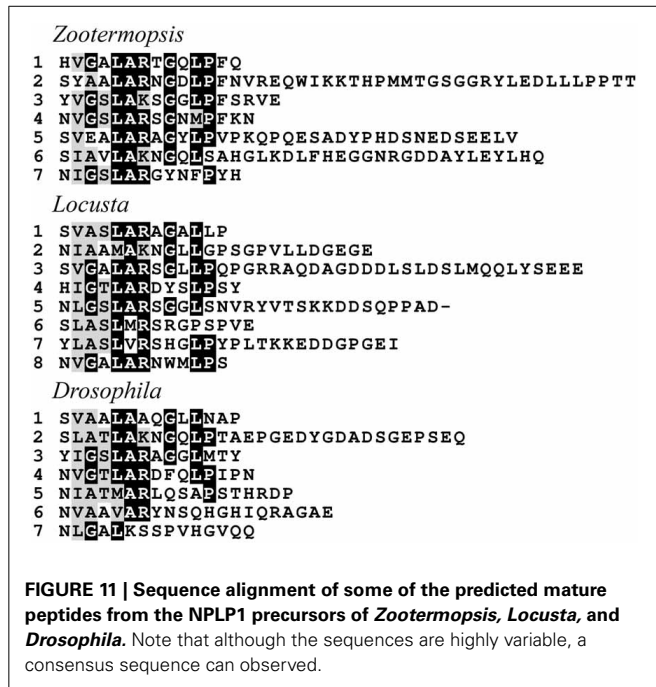
only a tiny minority codes for neuropepsin as identified from the corpora cardiaca.

NEUROPEPTIDE-LIKE PRECURSORS

Using mass spectrometry four proteins were identified as precursors for peptides in the *Drosophila* CNS (Baggerman et al., 2002). Only one of them, neuropeptide-like precursor 1 (NPLP1), has typical neuropeptide convertase cleavage sites. NPLP1 orthologs were found in both the termite and the locust. In *Drosophila* the structure of these peptides seems very variable between the different peptides. However, when the different paracopies from *Zootermopsis* are aligned, a conserved N-terminal consensus sequence becomes visible (Figure 11). Similar sequences are recognizable in the precursors of *Locusta* and *Drosophila* NPLP1. *Drosophila* NPLP1-VQQ, the only NPLP1 derived peptide known to have biological activity (Overend et al., 2011), conforms only partially to this consensus sequence (Figure 11), so it is plausible that some of these peptides have a different active site.

Although obviously any peptide present in the brain can be called a neuropeptide, it is perhaps more likely that the

other three precursors do not produce neuropeptides that activate GPCRs. It is for this reason that NPLP2, NPLP3, and NPLP4 are not considered here. A somewhat similar case can be made for the ITGQGNRIF precursor identified in a similar fashion from the honeybee brain (Hummon et al., 2006). In that case the observed peptide is indeed produced by cleavage at a typical convertase cleavage site and hence may well be produced by peptidergic neurons, but the part of the precursor that is best conserved appears to be a large protein of unknown function. Orthologs of the honey bee ITGQGNRIF precursor are present in both genomes (Supplementary Figure 9), as are orthologs of the honeybee NVPIYQEPFR precursor (Supplementary Figure 10). Interestingly, the *Blattella* homolog of the latter contains a sequence that is almost identical to baratin, a neuropeptide isolated from the brain of the cockroach *Leucophaea maderae* (Nässel et al., 2000). Immunoreactivity to baratin and its *Bombyx* ortholog has been described from neurons and neuroendocrine cells in *Leucophaea* and *Bombyx* (Nässel et al., 2000; Mitsumasa et al., 2009).



NPF AND sNPF

The two species each have two genes coding for an NPF-like peptide. As suggested by a phylogenetic tree based on the predicted NPF sequences, this duplication of the NPF gene has an ancient origin and appears to predate the separation of the Chelicerates from the Mandibulata (Supplementary Figure 11), as was previously suggested (Nuss et al., 2010). Transcriptome data show that the Carabid beetle *Pogonus chalceus* also has these two NPF genes, although we could only find the NPF 1 gene in *Tribolium*. The termite NPF 1 gene is similar to the *Bombyx* NPF1 and the *Daphnia* NPF gene in that it has a second transcript (NPF1b), where an optional exon gets added between the two others (Roller et al., 2008; Dirksen et al., 2011). This optional exon appears to be generally present in insects, because it was found in the NPF 1 genes and/or transcripts from cockroaches, phasmids, beetles, and Lepidoptera. The *Locusta* NPF 1 gene of *Locusta* also has this optional exon (Supplementary Figure 12). The termite has a third NPF gene, which may have become a pseudo gene. Although it has clear sequence similarity to NPF, including the typical intron site in the sequence coding the C-terminal and there are transcriptome reads corresponding to this sequence, it looks like the latter do not form a viable mRNA. A peptide very similar to the predicted structure of *Zootermopsis* NPF 1a has been identified from another termite species and it also has an NPF 1b transcript. Expression data show both transcripts of the NPF 1 gene in the central nervous system and the fore- and mid-gut (Nuss et al., 2010). Recent work showing that *Schistocerca* NPF is important in the regulation of feeding and male reproduction (Van Wielendaele et al., 2013a,b,c) concerns the ortholog of *Locusta* NPF1a.

Both the termite and the locust have one gene coding for sNPF, which in the locust codes for two paracopies and in the termite for one. The *Locusta* transcript is very similar to the one recently

described from *Schistocerca* (Dillen et al., 2014). Although the first identified sNPF from insects was isolated from cockroach midgut (Veenstra and Lambrou, 1995), the sNPF gene is not expressed in the midgut of *Schistocerca* (Dillen et al., 2014), and thus likely neither in *Locusta*. This suggests that sNPF isolated from the cockroach midgut was present in the visceral nervous system, like in *Drosophila* (Veenstra et al., 2008).

ORCOKININS AND ORCOMYOTROPINS

Orcokininins were initially isolated from *Orconectes limosus*, a crayfish (Stangier et al., 1992) and similar peptides have also been identified from cockroaches and locusts (Pascual et al., 2004; Hofer et al., 2005). Orcokinin genes are generally present in insects. A more detailed analysis of this gene in *Rhodnius prolixus* showed its mRNA to be alternatively spliced (Sterkel et al., 2012). The newly discovered mRNA was called the B-transcript and the peptides produced from it the B-orcokininins; similar transcripts are generally present in insects and may be specific for the midgut, at least in *Rhodnius* and *Drosophila* (Sterkel et al., 2012; Veenstra and Ida, 2014). Orcokinin-B is an evolutionary old peptide that can be found in the precursors predicted from TSA from species as diverse as *Glomeris postulata* (Myriapoda, Accession # GAKW01029956), *Calanus finmarchinus* (Maxillopoda, Accession # GAXK01131805) or *Speleonectes cf. tulumensis* (Remipedia, Accession # JL137063.1).

The orcomyotropins were similarly discovered from *Orconectes limosus* (Dirksen et al., 2000). Sequencing of the genome of *Daphnia pulex* revealed that the orcokininins and the orcomyotropins are encoded by the same gene and this led to the discovery that orcomyotropin-like peptides are also encoded by at least some insect orcokinin genes (Dirksen et al., 2011). This was also found in *Locusta* and *Zootermopsis*, which have both an orcokinin gene encoding both types of orcokininins well as orcomyotropin-like peptides. Interestingly, a peptide isolated using a tachykinin antiserum from the cockroach *Leucophaea maderae* is the only insect orcomyotropin physically identified from insects (Muren and Nässel, 1997); as it has a C-terminal Arg residue it may represent an incompletely processed peptide (Figure 12). Some of the peptides predicted from orcokinin genes, such as the *Locusta* peptides NLDGLGGHLLRQT and SGLDSLGFGEQ predicted from the orcokinin A transcript or SLDGIGGGNLVG from *Tribolium*, share sequence similarity with both the orcomyotropins and the orcokinin B's.

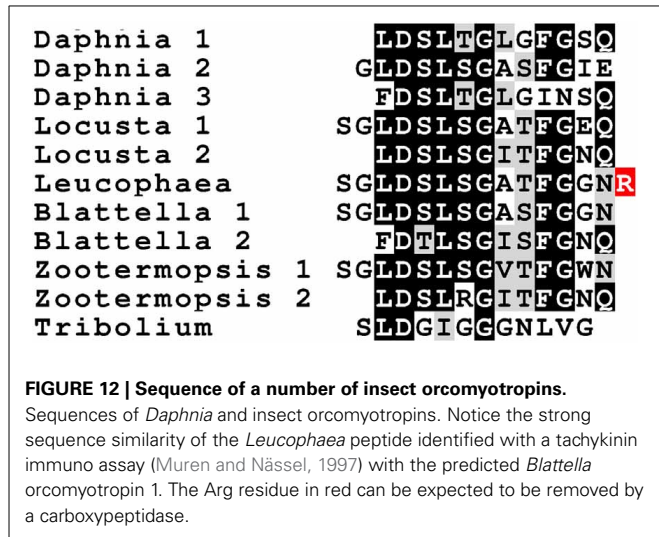
In both the locust and the termite, two different orcokinin transcripts were found, with the B-transcripts encoding large numbers of orcokinin B's and the A-transcripts encoding smaller numbers of orcomyotropins and orcokinin A's.

SALIVARY GLAND SALIVATION STIMULATING PEPTIDE

The salivary gland salivation stimulating peptide (SGSSP) was identified from the salivary glands of *Locusta* by its ability to stimulate the production of cAMP in the same tissue and its sequence reported to be EVGDLFKEWLQGNMN (Veelaert et al., 1995). Its likely precursor contains a number of identical copies of almost the same peptide (Figure 13); the 7th residue in the Edman degradation may have been incorrectly assigned as Lys while it should have been Gln, while in the majority of copies encoded by the

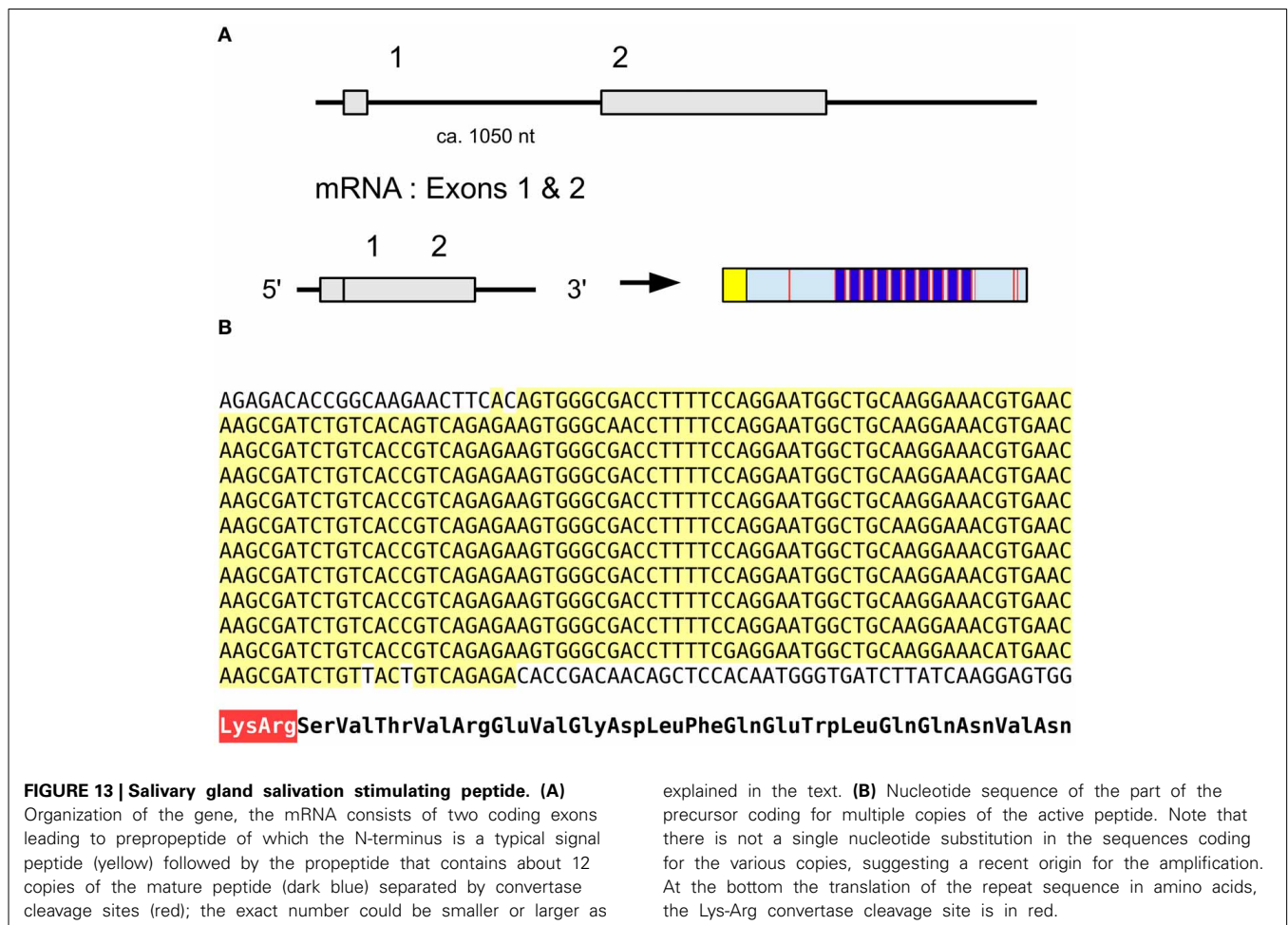
precursor the penultimate amino acid residue is Val rather than Met. The predicted precursor shown here is what we believe the best consensus sequence based on the transcriptome and genome data (Figure 13). This is the type of repetitive DNA sequence

that is hard if not impossible to resolve with the short paired-end sequences as used for the *Locusta* genome. It is therefore not impossible that a duplication present in the genome assembly that we removed, is indeed real and/or that the actual number of copies of this peptide encoded on this precursor is either larger or smaller. It is worth noting that not only the deduced amino acid sequences for the various paracopies are identical, but so are the DNA sequences encoding them. This suggests a very recent origin for the multiplication of the number of paracopies encoded by this gene. This is an unusual peptide as homologs of this peptide have never been found in other species and even with the large amounts of genomic and transcriptomic data available today we were unable to find a homolog in other species.



SIFamide AND SMYamide

SIFamide is a peptide initially identified from flies by using a bioassay on *Locusta* (Janssen et al., 1996) and it is now known to be present in many different arthropods. In both the termite and the locust there is a second gene encoding a SIFamide-like peptide, which we have called SMYamide to stress its similarity to SIFamide, even though the predicted C-terminal in *Locusta* is actually AMYamide (Figure 14). An SMYamide gene is also present in the preliminary assembly of the cockroach *Blattella germanica*, but was not found in several other arthropod genomes



that are currently assembled. An independent duplication of the SIFamide gene was previously reported for the silkworm, where the gene has been called IMFamide (Roller et al., 2008). A comparison of the genomic sequences shows that the duplication in Lepidoptera is independent from that in cockroaches, locusts, and termites. The expression of the SMYamide gene appears to about 10 percent of that of the SIFamide genes (2 vs. 17 transcriptome reads in *Locusta* and 447 vs. 5210 in *Zootermopsis*).

SULFAKININ

Although a *Locusta* sulfakinin has been identified (Schoofs et al., 1990d) the part of the gene coding this neuropeptide is not in the current genome assembly. However, a single transcriptome read was found to encode it and part of the sulfakinin gene could then be reconstructed using the genomic reads. It shows that the gene is similar to other insect sulfakinin genes in encoding two different sulfakinins, but the second copy does not adhere to the C-terminal consensus sequence, as the Met residue has been replaced by a Phe residue.

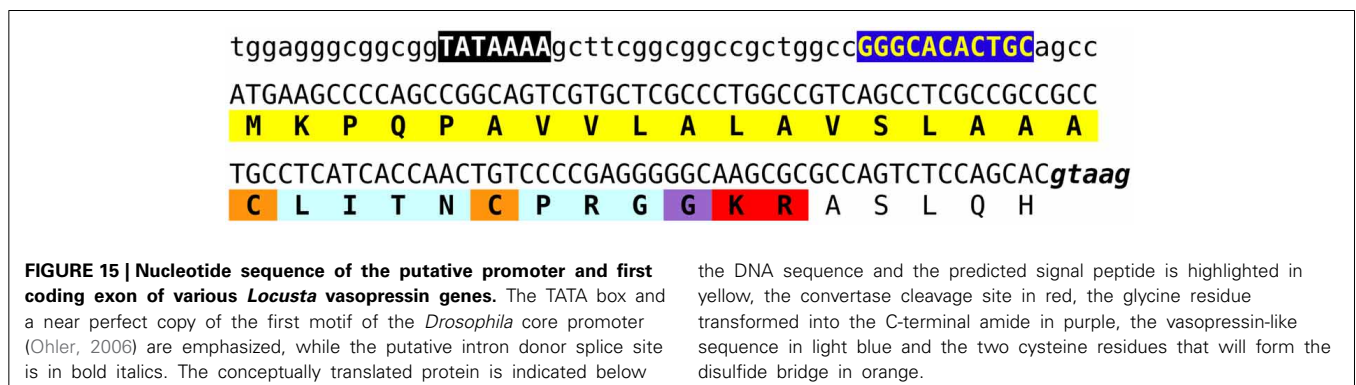
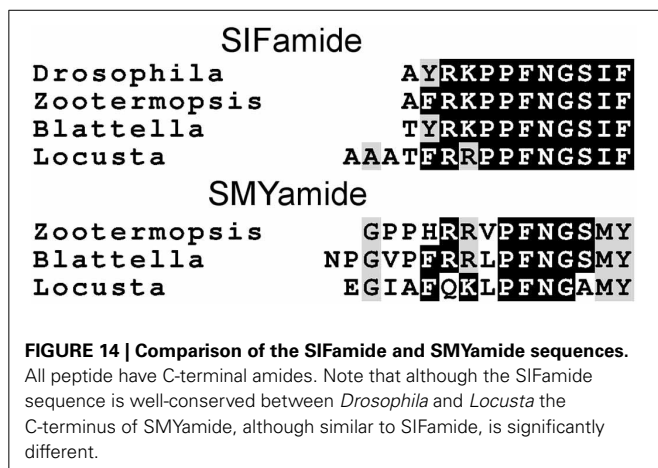
VASOPRESSIN-LIKE PEPTIDES

In the *Zootermopsis* genome a single vasopressin gene was encountered, but the *Locusta* genome seems to have 6–8 such genes. This estimate is based on the number of genomic reads that encode the CLINTCPRGGKR sequence present in various *Locusta* genomic SRAs. Interestingly the first exon of the *Locusta* gene as

well as the putative promoter sequence upstream of it, including a TATA box and a near perfect match of motif 1 of the *Drosophila* core promotor described by Ohler (2006), are very well-conserved (Figure 15). Three sequences in the genome assembly are completely identical to this consensus sequence. Of the other eight highly similar sequences in the assembly we detected one that has a single silent nucleotide substitution, another that lacks the last G in the GTAAG splice donor site, and a third one that has several nucleotide substitutions, some of which are predicted to lead to a different, but just as functional, signal peptide. In the latter sequence the amino acid immediately after the Lys-Arg convertase cleavage site is predicted to be a Asp residue rather than the Ala in the other sequences. Finally, there are four incomplete sequences and one that misses a piece in the DNA coding the prepropeptide and cannot produce a vasopressin-like peptide. Whereas the first exon and the DNA sequence immediately preceding it, are very well-conserved between the different genes, this is not the case for the other exons or the intron following the first coding exon. Although it is possible to identify some putative second exons, we were unable to identify the third. Those exons seem to be much less conserved than the first one. It is not clear whether the current genome assembly of the various vasopressin genes is correct, as the short length of the reads makes it difficult to assemble repetitive sequences or multiple copies of a similar gene.

DISCUSSION

Both the locust and the termite have a virtually complete set of insect neuropeptide genes, the only ones that seem to be lacking are a dilp 7 homolog in *Locusta* and dilp 8 homologs in both species. Nevertheless, as explained below we think the dilp 8 homologs may actually be present but escaped detection. However, this does not necessarily mean that these two species have all insect neuropeptides. For example, in *Nilaparvata* a GPCR was found that is clearly related to the vertebrate TRH receptors (Tanaka et al., 2014), but such a GPCR was not found in the *Zootermopsis* genome, and hence it may be that the termite does not have a gene coding its ligand, although it is also possible that this particular GPCR is encoded by the small part of the genomic DNA that is lacking from the genome assembly. These two species not only have what seems to be an almost complete set of neuropeptide genes, but they also appear to use these genes more intensively (two different CNMA transcripts in termites, two myosuppressin transcripts in the locust) and have some extra



the DNA sequence and the predicted signal peptide is highlighted in yellow, the convertase cleavage site in red, the glycine residue transformed into the C-terminal amide in purple, the vasopressin-like sequence in light blue and the two cysteine residues that will form the disulfide bridge in orange.

neuropeptide genes, the tryptopyrokinin and SMYamide genes. Furthermore, on the phylogenetic GPCR trees, the termite has more often than other species two receptors on each leaf. Some of these genes merit special discussion.

CALCITONIN

The presence of calcitonin genes in insects has not been previously reported. The calcitonin-like insect diuretic hormone that activates a calcitonin-like GPCR was generally assumed to be the insect calcitonin ortholog. We here show that insects not only have an authentic calcitonin gene, but also that in the more basal insect groups it produces two different types of calcitonin-like peptides. We were also able to identify candidate receptors for the different calcitonins. Although definitive proof that these GPCRs function as calcitonin receptors is lacking, the evidence is compelling. Like calcitonin B, several of its putative receptors are expressed in the midgut, as shown both by ESTs and TSAs data as well as the GPCR expression data from the silk worm (Yamanaka et al., 2008). This peptide thus seems to be an important midgut peptide and it is plausible that calcitonin A precursors, which yield only a single peptide per prepropeptide, is made predominantly in the central nervous system.

Processing of the insect calcitonin-B precursors seems different from that of other insect neuropeptide precursors, including the calcitonin-A precursor. Although this peptide has so far only been identified *in silico*, given its well-conserved structure and that of its precursors, it must be assumed that these precursors are cleaved by proteases at Lys-Arg base pairs. Common neuropeptide convertases, however, don't usually cleave next to the cysteine residue of a disulfide bridge (Devi, 1991; Rholam et al., 1995; Veenstra, 2000). As suggested by the structures of the precursors of calcitonin-like diuretic hormones that, like calcitonins, have a C-terminal Pro-amide, common neuropeptide convertases also seem to have difficulty cleaving immediately behind the dipeptide Pro-Gly [all the arthropod precursors of the calcitonin-like diuretic hormone have several dibasic amino acid residues following the Pro-Gly in the precursor, just as found here for the *Locusta* and *Zootermopsis* DH31 precursors (Supplementary Figures 1, 2)]. Yet, in the insect calcitonin B precursors the common convertase cleavage site is very close to a Cys residue and the Pro-Gly is followed by a single Lys-Arg pair. Indeed, in some species (*Tribolium*, *Bombyx*) the cleavage seems to happen at the sequence Pro-Gly-Lys-Arg-Cys, i.e., a convertase site that is both immediately after a Pro-Gly sequence *and* next to a Cys residue. This implies processing by a different convertase.

INSULIN- AND RELAXIN-LIKE PEPTIDES

Insects have often multiple insulin genes. In *Drosophila* there are eight, in the silkworm almost fifty (Mizoguchi and Okamoto, 2013). The migratory locust appears to be an exception with only one insulin gene found, but the termite has six such genes. The structural variability of the peptides predicted from these genes is large, yet many species have only a single classical insulin tyrosine kinase receptor. It is thus obvious that several of these peptides must act on the same receptor. Curiously, while in the locust only a single insulin gene was identified, three tyrosine kinase receptors were reported from its genome (Wang et al., 2014).

The *Drosophila* genes CG34411 and CG31096 code for GPCRs predicted to have relaxin-like ligands. Neither of these receptors is activated by *drosophila* dilp 2 (Van Hiel et al., 2014). Dilp 7 has a structure that is much better preserved than the other insulin-like peptides and has unambiguous homologs in mollusks and even some deuterostomes, while genes encoding such a peptide are only present in those genomes that also have a homolog of the *Drosophila* gene CG34411. This suggests that dilp 7 is the ligand for this receptor (Veenstra et al., 2012) and it thus seems appropriate to call it a relaxin-like peptide. The ligand for GC31096 may well be the recently identified dilp 8 (Colombani et al., 2012; Garelli et al., 2012), as it not only has a relaxin-like structure, but is also expressed in the ovary (Chintapalli et al., 2007). As the structure of dilp 8 is poorly conserved it is very difficult to identify in insect genomes. Nevertheless, as termites have its putative receptor (Supplementary Figure 3), we think it is plausible it also has a dilp 8 homolog. We did not find a dilp 7 homolog in the *Locusta* genome, nor evidence for its putative receptor, but there are *Locusta* genomic contigs of a GPCR gene that is an ortholog of *Drosophila* GC31096 and we thus think the migratory locust may well have a dilp 8 ortholog.

The dilp 7 and 8 homologs should perhaps be called relaxin-like peptides, as this reflects better their relationship to the vertebrate hormones. The other insulin-like peptides on the other hand are functionally similar to the insulin-like growth factors (Mizoguchi and Okamoto, 2013). Interestingly, in many species, including the migratory locust, such peptides are also expressed in the fat body (Lagueux et al., 1990; Okamoto et al., 2009a,b), where, like the vertebrate IGFs there is no convertase to remove the C-peptide. Although the absence of removal of the C-peptide has only been confirmed biochemically in the silk worm (Okamoto et al., 2009a), it seems likely that the same is true for other insect species (Mizoguchi and Okamoto, 2013). In those species that have multiple genes encoding insulin-like peptides, only one of them is expressed in the fat body. Although in *Drosophila* the insulin-like peptide expressed in the fat body is not expressed in endocrine cells, other genes, like dilp 3 and dilp 5 are expressed in both peripheral tissues and endocrine cells (Brogiolo et al., 2001; Chintapalli et al., 2007; Veenstra et al., 2008). It has been shown that an increase in the expression of dilp 3 by midgut muscle is associated with an increase in gut size (O'Brien et al., 2011), suggesting that independently of whether the insulin-like peptide is produced by endocrine cells in the brain or peripheral cells it activates the same receptor, even as its structure is likely different from that of the insulin-like peptides made by neuroendocrine cells.

NEUROPARSIN AND NEONEUROPARSINS

The data on the expression of the *Locusta* neuroparsin gene are truly intriguing. The number of neuroparsin transcripts seems to be similar to the number of transcripts of other neuropeptides. For example, there are 79 transcriptome reads for the periviscerokinin gene, a gene that appears well-expressed in the ventral nerve cord (Predel and Gäde, 2002) and 154 reads for the myosuppressin gene. Although 154 is a higher number than 66 neuroparsin specific reads, this is due to the much longer nucleotide sequence used to probe the SRAs (279 nucleotides for

the myosuppressin gene, 368 for the periviscerokinin gene but only 144 for exon 7 of the neuroparsin gene). Thus, it is the number of neoneuroparsin transcripts that is extremely high, while the number of neuroparsin transcripts seems normal. This means that it is unlikely that the neoneuroparsins are produced by classical neuroendocrine cells in the nervous system, of which the total numbers are perhaps 10–20 times those of the neuroparsin cells. This implies that neoneuroparsins are expressed by a large number of neurons, perhaps all.

Transcripts from the neuroparsin gene are also produced by the fatbody as well as ovary and testes, both in *Locusta* and *Schistocerca* (Lagueux et al., 1992; Claeys et al., 2003, 2005, 2006b). From the *Locusta* data it is impossible to ascertain whether or not the fat body produces only neoneuroparsins, or whether it also makes neuroparsin, but the data from *Schistocerca* suggest that although neuroparsin is expressed in the fat body, neoneuroparsin expression levels are about 500-fold higher (Claeys et al., 2005). As maturation of these proteins does not depend on specific processing enzymes typically present in neuroendocrine cells, they are expected to be secreted efficiently through the constitutive pathway. The question is what are these neoneuroparsins doing and why are they made in such large quantities? The four amino acid sequence (Gly-Gly-Pro-Tyr) that is lacking in the neoneuroparsins (Figure 10) can be expected to be important for correct protein folding as both Gly and Pro significantly affect the secondary structure of proteins. This suggests that the structure of the neoneuroparsins is different from that of neuroparsin and this difference may affect their binding to the receptor. It is difficult to understand why the neoneuroparsins would bind better to this receptor than neuroparsin. On the other hand, why would evolution favor the appearance of molecules that bind less efficiently to the receptor if the insect then needs to make these molecules in large quantities? After all, transcriptome estimates suggest that as many as 1 out of 2500 proteins made by the locust could be a neoneuroparsin. Could it be that the neoneuroparsins do indeed bind to the receptor but are unable to stimulate it and hence act essentially as neuroparsin antagonists? For as unusual as such an explanation seems, it could explain why they are made in such large amounts. In *Locusta* neuroparsin has an anti-juvenile hormone effect (Girardie et al., 1987), while in *Schistocerca* juvenile hormone and ecdysone increase transcripts for the neoneuroparsins, but not for neuroparsin (Claeys et al., 2006a), thus suggesting that the structural difference may be associated with a functional difference between the neoneuroparsins and neuroparsin. Unfortunately, a neuroparsin receptor has not yet been identified and hence the relative binding and stimulating activities of neuroparsin and the neoneuroparsins cannot be determined. The function of neuroparsin remains obscure, but it could be an important differentiation hormone (Veenstra, 2010b), an idea consistent with their effects on neurons in culture (Vanhems et al., 1990). If this were true, it might well have an important role in the differentiation between gregarious and solitary forms as also suggested by their expression in *Schistocerca* (Claeys et al., 2006a). Although the metabolic cost of producing so many neoneuroparsins seems high, the value of polymorphism for species survival is clear.

TRYPYROPYROKININS AND SALIVATION

If the existence of specific tryptopyrokinin genes was not unexpected, the question still arises as to why should the locust need four such genes? In *Drosophila* and *Bombyx*, the tryptopyrokinin receptor is strongly expressed in the salivary gland (Chintapalli et al., 2007; Yamanaka et al., 2008). Although this is not the case in *Rhodnius* (Paluzzi and O'Donnell, 2012), it seems nevertheless possible that in locusts tryptopyrokinins are important regulators of salivation. Apart from four tryptopyrokinin genes the locust also has evolved what looks like an entirely new neuropeptide gene (see below) encoding multiple copies of a neuropeptide that increases salivation by means of the stimulation of intracellular cAMP (Veelaert et al., 1995). The tryptopyrokinins usually are released as hormones into the hemolymph, but the salivation peptide is present in the salivary glands and can thus be expected to be released directly in the gland. It thus appears that the locust may have five neuropeptide genes controlling salivation in addition to salivation control by aminergic motoneurons in the suboesophageal ganglion and perhaps still other neuropeptides (see e.g., Ali et al., 1993). Even if our suggestion that the tryptopyrokinins may regulate salivation in the migratory locust were to be incorrect, the salivation peptide still raises the question as to why this locust needs to reinforce control of its salivary glands by what appears to be a novel neuropeptide. We believe the answer may be that this is not an ordinary but a migratory locust. Once millions of locusts descend on a field to feed, they all have great interest in starting to feed fast and furiously; for if not, there will be nothing left. The physiological changes from flying to feeding full speed are dramatic, but if the insect does not manage to make the transition fast, it will starve. Perhaps this explains why this species has four tryptopyrokinin genes and still evolved an entirely new salivation neuropeptide.

VASOPRESSIN

It is unfortunate that we were unable to determine the sequence(s) of the last part of the precursor of the vasopressin-like peptide. It is known from humans that numerous mutations of the neurophysin part of the vasopressin precursor will lead to autosomal dominant familial neurohypophyseal diabetes insipidus, as due to the incorrect folding of the vasopressin precursor in the endoplasmic reticulum the cells producing vasopressin die (Christensen et al., 2004). Obviously, human life span is much larger than that of a locust, and if neurons only die after a year, the locust would be dead anyway, but it still raises interesting questions with regard to the synthesis of this neuropeptide. When the vasopressin-like peptide was isolated from *Locusta*, it was followed by both RIA and a bioassay that measures an increase in amaranth secretion by the Malpighian tubules (Proux et al., 1987). However, when the identified peptide was synthesized in its monomeric form, i.e., in the same form as vasopressin, it had no biological activity. It was only active, when it was synthesized as an anti-parallel dimer. This was very surprising as all vasopressin-related peptides are known as monomers and a dimer had never been found. Proteins and peptides containing disulfide bridges need a mechanism to avoid the formation of inappropriate disulfide bridges like those occurring in vasopressin dimers. It seems a reasonable hypothesis, that neurophysins may help the proper

folding of the vasopressin precursor and once they are no longer made correctly, “errors” in the form of dimers may occur. As will be discussed below, the cells making the vasopressin-like peptide in *Locusta* need to make it in large quantities. One way to speed up production is to condense the gene; the short distance between the TATA box and the transcription start site of the various vasopressin genes (Figure 15) suggests that this has occurred. Another way would be to get rid of the neurophysin part of the precursor, which might then facilitate the formation of dimers. It is for this reason that it will be very interesting to see the complete sequence(s) of this precursor.

The function of this peptide in insects is poorly understood. Although it was initially proposed that the peptide had a diuretic function (Proux et al., 1987), it turned out that a vasopressin-like peptide from *Locusta* in any conformation (monomer, parallel dimer or anti-parallel dimer) failed to stimulate fluid secretion by the Malpighian tubules (Coast et al., 1993). The current hypothesis is that the peptide may act indirectly through the release of a brain diuretic hormone (Aikins et al., 2008). This peptide has received less attention than other insect neuropeptides because it is absent from *Drosophila*. Consequently, we still know very little about its physiology. In the pond snail the orthologous peptide acts on the *vas deferens* (Van Kesteren et al., 1995), while in *Schistocerca* there is sexual dimorphism of the axons in the terminal ganglion (Tyrer et al., 1993). Others have shown that ovariectomy of female crickets leads to a decrease in neurosecretory material in these neurons that is associated with changes at cellular level suggesting an increase in peptide production (Dürnberger et al., 1978). Together these data may indicate a reproductive function for this peptide.

AMPLIFICATION OF NEUROPEPTIDE GENES

Some neuropeptide genes are amplified, while others are not. In the locust, the insulin gene is not amplified, whereas it is amplified in many other insect species. On the other hand, the *Locusta* vasopressin gene is amplified, while in other species it is not. Thus, it seems an appropriate occasion to discuss this phenomenon. It is interesting to note that often the same genes are amplified in different species. In insects, these are usually the AKH and insulin genes. The latter shows a high degree of amplification. There are six genes in *Drosophila*, many more in the silk worm (e.g., Mizoguchi and Okamoto, 2013) and five in the termite genome. However, whereas in *Bombyx* and *Drosophila* numerous insulin genes are expressed by a limited number of brain neuroendocrine cells, only a single insulin gene is expressed in the fat body of both species. We think this is a clue as to why the insulin gene is amplified. Whereas the fat body consists of many cells and a small amount of insulin made by each of these cells will suffice to attain a physiologically relevant hemolymph concentration of this peptide, this is a much more difficult task to achieve by a limited number of neuroendocrine cells in the brain. It seems that two copies of the insulin gene (one on each chromosome) is simply insufficient to keep up with the large amounts of mRNA needed to produce the quantities of hormone that need to be secreted. When it is a matter of a small neuropeptide, the number of paracopies on the gene can be amplified. For example, leucokinin appears to be a diuretic peptide that in *Drosophila* is produced

from a gene encoding a single copy (Terhzaz et al., 1999), while mosquitoes that after a blood meal increase diuresis, make three copies from the same gene (Veenstra et al., 1997). It is obviously more difficult to produce multiple copies of a larger neuropeptide, probably even more so when it contains several disulfide bridges that need to be formed correctly on entry of its precursor into the endoplasmic reticulum. The hypothesis that it is a matter of quantity that needs to be produced is strongly reinforced by the observation that the *Locusta* vasopressin gene is amplified. In other insect genomes only a single copy of this gene has been found, as is the case for the termite as well as several other insects for which preliminary genome assemblies are available. So why is vasopressin so different in *Locusta*? In *Locusta* the neurons producing this peptide seem to have acquired an endocrine function with the peptide being released as a hormone from neurohemal terminals located on the thoracic peripheral nerves. In other locusts, such as *Schistocerca gregaria*, these cells are not only much smaller but they neither have these neuroendocrine release sites (Tyrer et al., 1993). Interestingly, in some crickets these neurons appear to have independently acquired a neuroendocrine function; in those animals the neurohemal release site is located on the ventral side of the brain (Weinbörmair et al., 1975). It is thus clear that in *Locusta* the amount of peptide that needs to be made to achieve physiological effects is much larger than in *Schistocerca*, because in *Locusta* the peptide is released as a hormone into the hemolymph whereas in *Schistocerca* the peptide is released only within the central nervous system. Thus, the correlation between quantity of peptide released and number of genes holds very well and it is hard to escape the notion that this is the reason why this particular gene got amplified. We would expect that crickets too may well have more than one vasopressin gene, whereas *Schistocerca* should need only one. Another solution to the problem would have been to increase the number of neurons producing this peptide, but it is probably much more difficult for a spontaneous mutation to alter the developmental program of neurogenesis in such a precise fashion than to duplicate a gene. Indeed, the number of endocrine cells in the corpora cardiaca producing AKH seems much more variable, from hundreds in locusts and cockroaches to about 20 in *Drosophila*.

SIFamide AND SMYamide

Both the locust and termite have a second gene that codes for a SIFamide related peptide. The same phenomenon has been described previously for *Bombyx mori* (Roller et al., 2008) and may well be generally occurring in Lepidoptera as such a gene is also present in the genomes of several other Lepidoptera. The SIFamide neurons are very large and their axons innervate many regions of the neuropile in insect species as diverse as *Drosophila* and *Schistocerca* (Terhzaz et al., 2007; Gellerer et al., 2014). Thus, these neurons likely produce and release large quantities of neuropeptides. However, as they release their peptide exclusively within the nervous system, a single gene could be sufficient, as is the case for example for the *Schistocerca* vasopressin neurons. Furthermore, whereas the primary sequence of the active peptide predicted from the various *Locusta* vasopressin genes is perfectly conserved, that is not the case for the SIFamide offshoot. Since the C-terminal of SIFamide has been very well-conserved during

evolution, this suggests that neither SMYamide nor IMFamide has the same affinity for the SIFamide receptor as SIFamide itself. Intriguingly, both *Locusta* and *Zootermopsis* not only have a GPCR that is likely the SIFamide receptor, but also a second GPCR that seems to be closely related to this putative SIFamide receptor (*Zootermopsis* A11 in Supplementary Figure 3), raising the question whether that receptor might be more specific for the predicted SMYamides. However, this second receptor also has a homolog in *Nilaparvata* (Tanaka et al., 2014; see Supplementary Figure 3, *Nilaparvata*_A11), but SMYamide has so far not been found in this species. We hypothesize that the duplication of the SIFamide is not maintained because of a need to produce more peptide, but rather to produce a different one.

In *Drosophila* SIFamide is produced by four very large interneurons that appear important for behavior. Sexual behavior in males is notably perturbed in flies that lack these neurons or which production of the peptide has been abolished by selective expression of SIFamide specific RNAi (Terhzaz et al., 2007). In *Bombyx mori* (Roller et al., 2008), the SIFamide-related peptide IMFamide is expressed not only in neurons homologous to the *Drosophila* SIFamide neurons, but also in a more ventrally located bilateral pair of neurons (Roller et al., 2008). Recent work on *Schistocerca* shows similar SIFamide immunoreactive neurons in this species (Gellerer et al., 2014). Given the strong similarity between SIFamide and SMYamide, it is likely that SIFamide antisera cross-react with SMYamide and it will be interesting to see which, if any, of the SIFamide immunoreactive neurons produce SMYamide in the locust.

DE NOVO EVOLUTION OF NEUROPEPTIDES

As indicated in the introduction, most neuropeptides and their receptors are evolutionarily very old and many have homologs in both deuterostomes and protostomes. The sequence of a neuropeptide GPCR is generally sufficient to predict its likely ligand due to the coevolution between receptors and ligands. This easily leads to the generalization that neuropeptides and receptors *always* coevolve. Indeed, there are very few examples of instances where one can make the argument that a neuropeptide evolved *de novo*.

The locust genome contains three genes that code unusual biologically active peptides, ovary maturing peptide, accessory gland myotropin II and SGSSP, the salivation peptide. None of these peptides have been found in any other insect orders and, with the exception of ovary maturing parsin, which is also made by another migratory locust, *Schistocerca gregaria*, not even in other insect species. This suggests a recent origin of these peptides. The accessory gland myotropin II was isolated from the male accessory gland. When males copulate with a female they not only transfer sperm, but also material made by the accessory glands, containing many substances that change female physiology. In particular they may induce ovulation and make the female reluctant to mate again. It is known that the accessory gland proteins evolve very rapidly (e.g., Wolfner, 2002; Panhuis et al., 2006), and it is therefore less surprising that this particular peptide has not been found in other species. Nothing is known about its mode of action and as it is derived from a protein that lacks the

typical neuropeptide convertase cleavage sites it is unlikely to be a classical neuropeptide, although it might act through a neuropeptide GPCR.

Ovary maturing parsin on the other hand is a neuropeptide. It was isolated from the corpora cardiaca and brain of the migratory locust and it has impressive effects on reproductive physiology of *Locusta* (Girardie and Girardie, 1996; Girardie et al., 1998). In *Locusta* it is produced from the same gene that produces the CRF-like diuretic hormone and the same is true for *Schistocerca* (Van Wielendaele et al., 2012). No clear homologs of this peptide have been identified in other insect species and immunoreactivity with antiserum against this hormone is only detected in locusts (Richard et al., 1994). However, it cannot be excluded that it is the function rather than the structure of this peptide that has been conserved and it may yet be determined that parts of the CRF-like diuretic hormone precursor in other insect species have similar effects. Nevertheless, experiments will be needed to address the apparent contradiction between the results of injections of this peptide and eliminating its production by RNAi silencing.

The most interesting *Locusta* peptide with regard to neuropeptide evolution is the salivation peptide. It looks like a real neuropeptide, as its precursor has the typical convertase cleavage sites and it stimulates the production of cyclic AMP, which usually occurs after interaction with a GPCR. On the other hand, it looks like this peptide has effectively evolved *de novo*, as similar peptides have not been found elsewhere and at the very least the amplification of the number of paracopies of the neuropeptide must be of a very recent origin as nucleotide mutations in the sequences coding them have yet to appear. Although it remains to be proven that this peptide acts on a GPCR, this seems the most likely explanation for an increase in intracellular cyclic AMP. It is hard to imagine that it would not act on a preexisting GPCR. Identification of the mechanism of action of this peptide seems thus of much interest.

There are other examples of what appear to be *de novo* evolution of ligands for existing neuropeptide receptors. On the one hand there are the examples of the tachykinins in salivary glands of the mosquito *Aedes aegypti* (Champagne and Ribeiro, 1994) and on the other hand there are the male accessory gland peptides in various insects that act on neuropeptide receptors in the female. The famous *Drosophila* sex peptide (Kubli, 2003) and the so called head peptide from the mosquito *Aedes aegypti* (Brown et al., 1994; Naccarati et al., 2012) both act through neuropeptide receptors, those for allatostatin B and short NPF, respectively (Kim et al., 2010; Poels et al., 2010; Yamanaka et al., 2010; Liesch et al., 2013). While the origin for both of them is obscure, in the case of the mosquito the DNA sequences (Stracker et al., 2002) suggest not only that this peptide has a recent origin, as it is lacking in other mosquitoes, but also that its origin is independent from the sNPF neuropeptide gene. Such “accidental” evolution of new ligand receptor combinations may be more common than generally realized. We will only find it when it has happened in a clade that has become a success during evolution. A possible example of such a successful *de novo* association between a ligand and a receptor is provided by prostaglandin D and the DP2 receptor (Hirai et al., 2001).

CONCLUSION

If there is a message from these genomes, it is probably that we will need more physiology. Genome analysis mostly identifies neuropeptide genes that we know already. Of course, there are interesting details to discover, such as the alternative splicing of the myosuppressin gene in *Locusta* or the CNMa gene in *Zootermopsis*, or the calcitonin gene and the putative identification of its receptors. The most interesting findings of the *Locusta* genome concern the salivation peptide, vasopressin, neuroparsins, and the tryptopyrokinins, but these findings pose more questions than can be solved by bioinformatics alone. Real physiology will be needed to answer these questions. The importance of physiology is nicely illustrated by the salivation peptide. If this peptide had not been previously shown to stimulate salivation we would not have a clue as to its function and, consequently, it would hardly be perceived as interesting. The physiological significance of the neuroparsins, tryptopyrokinins, and the vasopressin-like peptide similarly awaits physiological experiments.

ACKNOWLEDGMENTS

I thank the two reviewers and in particular associate editor Xanthe Vafopoulou for their helpful suggestions to improve this manuscript. I am also much indebted to Pierre Rouzé for gently pointing out that I missed the calcitonin gene in the spider mite genome. Analysis of the preliminary genome assemblies of the sandfly *Lutzomyia longipalpis*, the dragonfly *Ladona fulva*, the German cockroach *Blattella germanica* and the mayfly *Ephemera danica* was only possible because of data collected at and made available by the Baylor College of Medicine Human Genome Sequencing Center. The genome of the sand fly *Phlebotomus papatasi* was determined by the Washington University Genome Sequencing Center and the sequencing of *Diaphorina citri* was organized by Wayne B. Hunter of the USDA-ARS. I am also grateful to M. Shelomi and B. R. Johnson (University of California) for making their transcriptome data on the phasmid midgut public before publication, as that was very helpful.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://www.frontiersin.org/journal/10.3389/fphys.2014.00454/abstract>

REFERENCES

- Adams, M. D., Celniker, S. E., Holt, R. A., Evans, C. A., Gocayne, J. D., Amanatides, P. G., et al. (2000). The genome sequence of *Drosophila melanogaster*. *Science* 287, 2185–2195. doi: 10.1126/science.287.5461.2185
- Aikins, M. J., Schooley, D. A., Begum, K., Detheux, M., Beeman, R. W., and 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. doi: 10.1016/j.ibmb.2008.04.006
- Ali, D. W., Orchard, I., and Lange, A. B. (1993). The aminergic control of locust (*Locusta migratoria*) salivary glands: evidence for serotonergic and dopaminergic innervation. *J. Insect Physiol.* 39, 623–632. doi: 10.1016/0022-1910(93)90067-2
- Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., et al. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402. doi: 10.1093/nar/25.17.3389
- Audsley, N., Vandersmissen, H. P., Weaver, R., Dani, P., Matthews, J., Down, R., et al. (2013). Characterisation and tissue distribution of the PISCF allatostatin receptor in the red flour beetle, *Tribolium castaneum*. *Insect Biochem. Mol. Biol.* 43, 65–74. doi: 10.1016/j.ibmb.2012.09.007
- Badisco, L., Huybrechts, J., Simonet, G., Verlinden, H., Marchal, E., Huybrechts, R., et al. (2011). Transcriptome analysis of the desert locust central nervous system: production and annotation of a *Schistocerca gregaria* EST database. *PLoS ONE* 6:e17274. doi: 10.1371/journal.pone.0017274
- Baggerman, G., Cerstiaens, A., de Loof, A., and Schoofs, L. (2002). Peptidomics of the larval *Drosophila melanogaster* central nervous system. *J. Biol. Chem.* 277, 40368–40374. doi: 10.1074/jbc.M206257200
- Baldwin, D. C., Schegg, K. M., Furuya, K., Lehmborg, E., and Schooley, D. A. (2001). Isolation and identification of a diuretic hormone from *Zootermopsis nevadensis*. *Peptides* 22, 147–152. doi: 10.1016/S0196-9781(00)00371-5
- Beenakkers, A. M. Th. (1969). The influence of corpus allatum and corpus cardiacum on lipid metabolism in *Locusta*. *Gen. Comp. Endocrinol.* 13, 492. doi: 10.1016/S0196-9781(00)00371-5
- Bendtsen, J. D., Nielsen, H., von Heijne, G., and Brunak, S. (2004). Improved prediction of signal peptides: SignalP 3.0. *J. Mol. Biol.* 340, 783–795. doi: 10.1016/j.jmb.2004.05.028
- Bogerd, J., Kooiman, F. P., Pijnenburg, M. A., Hekking, L. H., Oudejans, R. C., and Van der Horst, D. J. (1995). Molecular cloning of three distinct cDNAs, each encoding a different adipokinetic hormone precursor, of the migratory locust, *Locusta migratoria*. Differential expression of the distinct adipokinetic hormone precursor genes during flight activity. *J. Biol. Chem.* 270, 23038–23043. doi: 10.1074/jbc.270.39.23038
- Broggiolo, W., Stocker, H., Ikeya, T., Rintelen, F., Fernandez, R., and Hafen, E. (2001). An evolutionarily conserved function of the *Drosophila* insulin receptor and insulin-like peptides in growth control. *Curr. Biol.* 11, 213–221. doi: 10.1016/S0960-9822(01)00068-9
- Brown, M. R., Klowden, M. J., Crim, J. W., Young, L., Shrouder, L. A., and Lea, A. O. (1994). Endogenous regulation of mosquito host-seeking behavior by a neuropeptide. *J. Insect Physiol.* 40, 399–406. doi: 10.1016/0022-1910(94)90158-9
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009). BLAST+: architecture and applications. *BMC Bioinformatics* 10:421. doi: 10.1186/1471-2105-10-421
- Cardoso, J. C. R., Félix, R. C., and Power, D. M. (2014). Nematode and arthropod genomes provide new insights into the evolution of class 2 B1 GPCRs. *PLoS ONE* 9:e92220. doi: 10.1371/journal.pone.0092220
- Cazzamali, G., Torp, M., Hauser, F., Williamson, M., and Grimmelikhuijzen, C. J. P. (2005). The *Drosophila* gene CG9918 codes for a pyrokinin-1 receptor. *Biochem. Biophys. Res. Commun.* 335, 14–19. doi: 10.1016/j.bbrc.2005.07.038
- Champagne, D. E., and Ribeiro, J. M. C. (1994). Sialokinin I and II: vasodilatory tachykinins from the yellow fever mosquito *Aedes aegypti*. *Proc. Natl. Acad. Sci. U.S.A.* 91, 138–142.
- Chang, J. C., Yang, R. B., Adams, M. E., and Lu, K. H. (2009). Receptor guanylyl cyclases in Inka cells targeted by eclosion hormone. *Proc. Natl. Acad. Sci. U.S.A.* 106, 13371–13376. doi: 10.1073/pnas.0812593106
- Chen, S., Yang, P., Jiang, F., Wei, Y., Ma, Z., and Kang, L. (2010). *De novo* analysis of transcriptome dynamics in the migratory locust during the development of phase traits. *PLoS ONE* 5:e15633. doi: 10.1371/journal.pone.0015633
- Chen, W., Jiang, G. F., Sun, S. H., Lu, Y., Ma, F., and Li, B. (2013). Identification of differentially expressed genes in American cockroach ovaries and testes by suppression subtractive hybridization and the prediction of its miRNAs. *Mol. Genet. Genomics* 288, 627–638. doi: 10.1007/s00438-013-0777-1
- Chintapalli, V. R., Wang, J., and Dow, J. A. (2007). Using FlyAtlas to identify better *Drosophila melanogaster* models of human disease. *Nat. Genet.* 39, 715–720. doi: 10.1038/ng2049
- Christensen, J. H., Siggaard, C., Corydon, T. J., deSanctis, L., Kovacs, L., Robertson, G. L., et al. (2004). Six novel mutations in the arginine vasopressin gene in 15 kindreds with autosomal dominant familial neurohypophyseal diabetes insipidus give further insight into the pathogenesis. *Eur. J. Hum. Genet.* 12, 44–51. doi: 10.1038/sj.ejhg.5201086
- Claeys, I., Breugelmanns, B., Simonet, G., Franssens, V., Van Soest, S., and Vanden Broeck, J. (2006a). Regulation of *Schistocerca gregaria* neuroparsin transcript levels by juvenile hormone and 20-hydroxyecdysone. *Arch. Insect Biochem. Physiol.* 62, 107–115. doi: 10.1002/arch.20127

- Claeys, I., Breugelmanns, B., Simonet, G., Van Soest, S., Sas, F., de Loof, A., et al. (2006b). Neuroparsin transcripts as molecular markers in the process of desert locust (*Schistocerca gregaria*) phase transition. *Biochem. Biophys. Res. Commun.* 341, 599–606. doi: 10.1016/j.bbrc.2006.01.011
- Claeys, I., Simonet, G., Breugelmanns, B., Van Soest, S., Franssens, V., Sas, F., et al. (2005). Quantitative real-time RT-PCR analysis in desert locusts reveals phase dependent differences in neuroparsin transcript levels. *Insect Mol. Biol.* 14, 415–422. doi: 10.1111/j.1365-2583.2005.00572.x
- Claeys, I., Simonet, G., Van Loy, T., de Loof, A., and Vanden Broeck, J. (2003). cDNA cloning and transcript distribution of two novel members of the neuroparsin family in the desert locust, *Schistocerca gregaria*. *Insect Mol. Biol.* 12, 473–481. doi: 10.1046/j.1365-2583.2003.00431.x
- Clynen, E., Baggerman, G., Huybrechts, J., Vanden Bosch, L., de Loof, A., and Schoofs, L. (2003a). Peptidomics of the locust corpora allata: identification of novel pyrokinins (-FXPRLamides). *Peptides* 24, 1493–1500. doi: 10.1016/j.peptides.2003.10.006
- Clynen, E., Baggerman, G., Veelaert, D., Cerstiaens, A., Van der Horst, D., Harthoorn, L., et al. (2001). Peptidomics of the pars intercerebralis-corpora cardiaca complex of the migratory locust, *Locusta migratoria*. *Eur. J. Biochem.* 268, 1929–1939. doi: 10.1046/j.1432-1327.2001.02067.x
- Clynen, E., Huybrechts, J., Baggerman, G., Van Doorn, J., Van Der Horst, D., de Loof, A., et al. (2003c). Identification of a glycogenolysis-inhibiting peptide from the corpora cardiaca of locusts. *Endocrinology* 144, 3441–3448. doi: 10.1210/en.2002-0107
- Clynen, E., Huybrechts, J., de Loof, A., and Schoofs, L. (2003b). Mass spectrometric analysis of the perisymphatic organs in locusts: identification of novel periviscerokinins. *Biochem. Biophys. Res. Commun.* 300, 422–428. doi: 10.1016/S0006-291X(02)02869-3
- Clynen, E., Huybrechts, J., Verleyen, P., de Loof, A., and Schoofs, L. (2006). Annotation of novel neuropeptide precursors in the migratory locust based on transcript screening of a public EST database and mass spectrometry. *BMC Genomics* 7:201. doi: 10.1186/1471-2164-7-201
- Clynen, E., and Schoofs, L. (2009). Peptidomic survey of the locust neuroendocrine system. *Insect Biochem. Mol. Biol.* 39, 491–507. doi: 10.1046/j.1432-1327.2001.02067.x
- Coast, G. M., Rayne, R. C., Hayes, T. K., Mallet, A. I., Thompson, K. S., and 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
- Colombani, J., Andersen, D. S., and Léopold, P. (2012). Secreted peptide Dilp8 coordinates *Drosophila* tissue growth with developmental timing. *Science* 336, 582–585. doi: 10.1126/science.1216689
- Devi, L. (1991). Consensus sequence for processing of peptide precursors at monobasic sites. *FEBS Lett.* 280, 189–194. doi: 10.1016/0014-5793(91)80290-J
- Dillen, S., Verdonck, R., Zels, S., Van Wielendaele, P., and Vanden Broeck, J. (2014). Identification of the short neuropeptide F precursor in the desert locust: evidence for an inhibitory role of sNPF in the control of feeding. *Peptides* 53, 134–139. doi: 10.1016/j.peptides.2013.09.018
- Dirksen, H., Burdzik, S., Sauter, A., and Keller, R. (2000). Two orckinins and the novel octapeptide orcomytropin in the hindgut of the crayfish *Orconectes limosus*: identified myostimulatory neuropeptides originating together in neurones of the terminal abdominal ganglion. *J. Exp. Biol.* 203, 2807–2818.
- Dirksen, H., Neupert, S., Predel, R., Verleyen, P., Huybrechts, J., Strauss, J., et al. (2011). Genomics, transcriptomics, and peptidomics of *Daphnia pulex* neuropeptides and protein hormones. *J. Proteome Res.* 10, 4478–4504. doi: 10.1021/pr200284e
- Dürnberger, H., Pohlhammer, K., and Weinbörmair, G. (1978). The paramedial neurosecretory cells of the suboesophageal ganglion of the cricket, *Teleogryllus commodus* (Walk.). *Cell Tissue Res.* 187, 489–494. doi: 10.1007/BF00229613
- Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797. doi: 10.1186/1471-2105-5-113
- Furuya, K., Milchak, R. J., Schegg, K. M., Zhang, J., Tobe, S. S., Coast, G. M., et al. (2000). Cockroach diuretic hormones: characterization of a calcitonin-like peptide in insects. *Proc. Natl. Acad. Sci. U.S.A.* 97, 6469–6474. doi: 10.1073/pnas.97.12.6469
- Garelli, A., Gontijo, A. M., Míguela, V., Caparros, E., and Dominguez, M. (2012). Imaginal discs secrete insulin-like peptide 8 to mediate plasticity of growth and maturation. *Science* 336, 579–582. doi: 10.1126/science.1216735
- Gellerer, A., Franke, A., Neupert, S., Predel, R., Xin, Z., Liu, S., et al. (2014). Identification and distribution of SIFamide in the nervous system of the desert locust *Schistocerca gregaria*. *J. Comp. Neurol.* 523, 108–125. doi: 10.1002/cne.23671
- Girardie, J., Bourème, D., Couillaud, F., Tamarelle, M., and Girardie, A. (1987). Anti-juvenile hormone effect of neuroparsin A, a neuroprotein isolated from locust corpora cardiaca. *Insect Biochem.* 17, 977–983.
- Girardie, J., and Girardie, A. (1996). Lom OMP, a putative ecdysiotropic factor for the ovary in *Locusta migratoria*. *J. Insect Physiol.* 42, 215–221. doi: 10.1016/0022-1910(95)00097-6
- Girardie, J., Girardie, A., Huet, J. C., and Pernollet, J. C. (1989). Amino acid sequence of locust neuroparsins. *FEBS Lett.* 245, 4–8. doi: 10.1016/0014-5793(89)80179-6
- Girardie, J., Huet, J. C., Atay-Kadir, Z., Ettaouil, S., Delbecq, J. P., Fournier, B., et al. (1998). Isolation, sequence determination, physical and physiological characterization of the neuroparsins and ovary maturing parsins of *Schistocerca gregaria*. *Insect Biochem. Mol. Biol.* 28, 641–650. doi: 10.1016/S0965-1748(98)00053-8
- Girardie, J., Olivier, R., and Girardie, A. (1992). Time-dependent variations in the activity of a novel ovary maturing neurohormone from the nervous corpora cardiaca during oogenesis in the locust, *Locusta migratoria migratorioides*. *J. Insect Physiol.* 38, 215–221. doi: 10.1016/0022-1910(92)90069-P
- Girardie, J., Richard, O., Huet, J. C., Nespoulous, C., Van Dorselaer, A., and Pernollet, J. C. (1991). Physical characterization and sequence identification of the ovary maturing parsin. A new neurohormone purified from the nervous corpora cardiaca of the African locust (*Locusta migratoria migratorioides*). *Eur. J. Biochem.* 202, 1121–1126. doi: 10.1111/j.1432-1033.1991.tb16479.x
- Gouy, M., Guindon, S., and Gascuel, O. (2010). SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol. Biol. Evol.* 27, 221–224. doi: 10.1093/molbev/msp259
- Guindon, S., and Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* 52, 696–670. doi: 10.1080/10635150390235520
- Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., et al. (2013). *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* 8, 1494–1512. doi: 10.1038/nprot.2013.084
- Hansen, K. K., Stafflinger, E., Schneider, M., Hauser, F., Cazzamali, G., Williamson, M., et al. (2010). Discovery of a novel insect neuropeptide signaling system closely related to the insect adipokinetic hormone and corazonin hormonal systems. *J. Biol. Chem.* 285, 10736–10747. doi: 10.1074/jbc.M109.045369
- Hauser, F., Cazzamali, G., Williamson, M., Park, Y., Li, B., Tanaka, Y., et al. (2008). A genome-wide inventory of neurohormone GPCRs in the red flour beetle *Tribolium castaneum*. *Front. Neuroendocrinol.* 29:142–165. doi: 10.1016/j.yfrne.2007.10.003
- Hauser, F., Neupert, S., Williamson, M., Predel, R., Tanaka, Y., and Grimmelikhuijzen, C. J. P. (2010). Genomics and peptidomics of neuropeptides and protein hormones present in the parasitic wasp *Nasonia vitripennis*. *J. Proteome Res.* 9, 5296–5310. doi: 10.1021/pr100570j
- Hernandez, D., François, P., Farinelli, L., Osteràs, M., and Schrenzel, J. (2008). *De novo* bacterial genome sequencing: millions of very short reads assembled on a desktop computer. *Genome Res.* 18, 802–809. doi: 10.1101/gr.072033.107
- Hill, S. R., and Orchard, I. (2007). Isolation and sequencing of two FMRFamide-related peptides from the gut of *Locusta migratoria* L. *Peptides* 28, 1490–1497. doi: 10.1016/j.peptides.2007.06.015
- Hirai, H., Tanaka, K., Yoshie, O., Ogawa, K., Kenmotsu, K., Takamori, Y., et al. (2001). Prostaglandin D2 selectively induces chemotaxis in T helper type 2 cells, eosinophils, and basophils via seven-transmembrane receptor CRTH2. *J. Exp. Med.* 193, 255–261. doi: 10.1084/jem.193.2.255
- Hofer, S., Dirksen, H., Tollbäck, P., and Homberg, U. (2005). Novel insect orckinins: characterization and neuronal distribution in the brains of selected dicondylan insects. *J. Comp. Neurol.* 490, 57–71. doi: 10.1002/cne.20650
- Homma, T., Watanabe, K., Tsurumaru, S., Kataoka, H., Imai, K., Kamba, M., et al. (2006). G protein-coupled receptor for diapause hormone, an inducer of *Bombyx* embryonic diapause. *Biochem. Biophys. Res. Commun.* 344, 386–393. doi: 10.1016/j.bbrc.2006.03.085
- Hummon, A. B., Richmond, T. A., Verleyen, P., Baggerman, G., Huybrechts, J., Ewing, M. A., et al. (2006). From the genome to the proteome: uncovering peptides in the *Apis* brain. *Science* 314, 647–649. doi: 10.1126/science.1124128

- Ida, T., Takahashi, T., Tominaga, H., Sato, T., Kume, K., Ozaki, M., et al. (2011a). 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. doi: 10.1016/j.bbrc.2011.06.081
- Ida, T., Takahashi, T., Tominaga, H., Sato, T., Kume, K., Yoshizawa-Kumagaye, K., et al. (2011b). Identification of the endogenous cysteine-rich peptide trissin, a ligand for an orphan G protein-coupled receptor in *Drosophila*. *Biochem. Biophys. Res. Commun.* 414, 44–48. doi: 10.1016/j.bbrc.2011.09.018
- Ida, T., Takahashi, T., Tominaga, H., Sato, T., Sano, H., Kume, K., et al. (2012). Isolation of the bioactive peptides CCHamide-1 and CCHamide-2 from *Drosophila* and their putative role in appetite regulation as ligands for G protein-coupled receptors. *Front. Endocrinol. (Lausanne)*. 3:177. doi: 10.3389/fendo.2012.00177
- Iversen, A., Cazzamali, G., Williamson, M., Hauser, F., and 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. doi: 10.1016/S0006-291X(02)02709-2
- Janssen, I., Schoofs, L., Spittaels, K., Neven, H., Vanden Broeck, J., Devreese, B., et al. (1996). Isolation of NEB-LFamide, a novel myotropic neuropeptide from the grey fleshfly. *Mol. Cell. Endocrinol.* 117, 157–165. doi: 10.1016/0303-7207(95)03746-2
- Janssen, T., Claeys, I., Simonet, G., de Loof, A., Girardie, J., and Vanden Broeck, J. (2001). cDNA cloning and transcript distribution of two different neuropeptin precursors in the desert locust, *Schistocerca gregaria*. *Insect Mol. Biol.* 10, 183–189. doi: 10.1046/j.1365-2583.2001.00257.x
- Jiang, H., Lkhagva, A., Dauberová, I., Chae, H. S., Simo, L., Jung, S. H., et al. (2013). Natalisin, a tachykinin-like signaling system, regulates sexual activity and fecundity in insects. *Proc. Natl. Acad. Sci. U.S.A.* 110, 3526–3534. doi: 10.1073/pnas.1310676110
- Johnson, E. C., Shafer, O. T., Trigg, J. S., Park, J., Schooley, D. A., Dow, J. A. T., et al. (2005). A novel diuretic hormone receptor in *Drosophila*: evidence for conservation of CGRP signaling. *J. Exp. Biol.* 208, 1239–1246. doi: 10.1242/jeb.01529
- Jung, S. H., Lee, J. H., Chae, H. S., Seong, J. Y., Park, Y., Park, Z. Y., et al. (2014). Identification of a novel insect neuropeptide, CNMa and its receptor. *FEBS Lett.* 588, 2037–2041. doi: 10.1016/j.febslet.2014.04.028
- Kang, L., Chen, X., Zhou, Y., Liu, B., Zheng, W., Li, R., et al. (2004). The analysis of large-scale gene expression correlated to the phase changes of the migratory locust. *Proc. Natl. Acad. Sci. U.S.A.*, 101, 17611–17615. doi: 10.1073/pnas.0407753101
- Kay, I., Wheeler, C. H., Coast, G. M., Totty, N. F., Cusinato, O., Patel, M., et al. (1991). Characterization of a diuretic peptide from *Locusta migratoria*. *Biol. Chem. Hoppe-Seyler*. 372, 929–934. doi: 10.1515/bchm3.1991.372.2.929
- Kim, Y.-J., Bartalska, K., Audsley, N., Yamanaka, N., Yapici, N., Lee, J. Y., et al. (2010). MIPs are ancestral ligands for the sex peptide receptor. *Proc. Natl. Acad. Sci. U.S.A.* 107, 6520–6525. doi: 10.1073/pnas.0914764107
- Kirkness, E. F., Haas, B. J., Sun, W., Braig, H. R., Perotti, M. A., Clark, J. M., et al. (2010). Genome sequences of the human body louse and its primary endosymbiont provide insights into the permanent parasitic lifestyle. *Proc. Natl. Acad. Sci. U.S.A.* 107, 12168–12173. doi: 10.1073/pnas.1003379107
- Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E. L. L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* 305, 567–580. doi: 10.1006/jmbi.2000.4315
- Kubli, E. (2003). Sex-peptides: seminal peptides of the *Drosophila* male. *Cell. Mol. Life Sci.* 60, 1689–1704. doi: 10.1007/s00018-003-3052
- Lagueux, M., Kromer, E., and Girardie, J. (1992). Cloning of a *Locusta* cDNA encoding neuropeptin A. *Insect Biochem. Mol. Biol.* 22, 511–516. doi: 10.1016/0965-1748(92)90027-C
- Lagueux, M., Lwoff, L., Meister, M., Goltzené, F., and Hoffmann, J. A. (1990). cDNAs from neurosecretory cells of brains of *Locusta migratoria* (Insecta, Orthoptera) encoding a novel member of the superfamily of insulins. *Eur. J. Biochem.* 187, 249–254. doi: 10.1111/j.1432-1033.1990.tb15302.x
- Lange, A. B., Peeff, N. M., and Orchard, I. (1994). Isolation, sequence, and bioactivity of FMRFamide-related peptides from the locust ventral nerve cord. *Peptides* 15, 1089–1094. doi: 10.1016/0196-9781(94)90075-2
- Lehmberg, E., Ota, R. B., Furuya, K., King, D. S., Applebaum, S. W., Ferenz, H. J., et al. (1991). Identification of a diuretic hormone of *Locusta migratoria*. *Biochem. Biophys. Res. Commun.* 179, 1036–1041. doi: 10.1016/0006-291X(91)91923-Z
- Liesch, J., Bellani, L. L., and Vossall, L. B. (2013). Functional and genetic characterization of neuropeptide Y-like receptors in *Aedes aegypti*. *PLoS Negl. Trop. Dis.* 7:e2486. doi: 10.1371/journal.pntd.0002486
- Simo, L., Koči, J., and Park, Y. (2013). Receptors for the neuropeptides, myoinhibitory peptide and SIFamide, in control of the salivary glands of the black-legged tick *Ixodes scapularis*. *Insect Biochem. Mol. Biol.* 43, 376–387. doi: 10.1016/j.ibmb.2013.01.002
- Ma, Z., Yu, J., and Kang, L. (2006). LocustDB: a relational database for the transcriptome and biology of the migratory locust (*Locusta migratoria*). *BMC Genomics* 7:11. doi: 10.1186/1471-2164-7-11
- Mayer, R. J., and Candy, D. J. (1969). Control of haemolymph lipid concentrations during locust flight: an adipokinetic hormone from the corpora cardiaca. *J. Insect Physiol.* 15, 611–620. doi: 10.1016/0022-1910(69)90259-5
- Meredith, J., Ring, M., Macins, A., Marschall, J., Cheng, N. N., Theilmann, D., et al. (1996). Locust ion transport peptide (ITP): primary structure, cDNA and expression in a baculovirus system. *J. Exp. Biol.* 199, 1053–1061.
- Mirabeau, O., and Joly, J. S. (2013). Molecular evolution of peptidergic signaling systems in bilaterians. *Proc. Natl. Acad. Sci. U.S.A.* 110, E2028–E2037. doi: 10.1073/pnas.1219956110
- Mitsumasa, K., Tanaka, Y., Niimi, T., Yamashita, O., and Yaginuma, T. (2009). Novel gene encoding precursor protein consisting of possible several neuropeptides expressed in brain and frontal ganglion of the silkworm, *Bombyx mori*. *Peptides* 30, 1233–1240. doi: 10.1016/j.peptides.2009.03.023
- Mizoguchi, A., and Okamoto, N. (2013). Insulin-like and IGF-like peptides in the silkworm *Bombyx mori*: discovery, structure, secretion, and function. *Front. Physiol.* 4:217. doi: 10.3389/fphys.2013.00217
- Muren, J. E., and Nässel, D. R. (1997). Seven tachykinin-related peptides isolated from the brain of the Madeira cockroach: evidence for tissue-specific expression of isoforms. *Peptides* 18, 7–15. doi: 10.1016/S0196-9781(96)00243-4
- Naccarati, C., Audsley, N., Keen, J. N., Kim, J. H., Howell, G. J., Kim, Y. J., et al. (2012). The host-seeking inhibitory peptide, Aea-HP-1, is made in the male accessory gland and transferred to the female during copulation. *Peptides* 34, 150–157. doi: 10.1016/j.peptides.2011.10.027
- Nässel, D. R., Persson, M. G., and Muren, J. E. (2000). Baratin, a nonamidated neurostimulating neuropeptide, isolated from cockroach brain: distribution and actions in the cockroach and locust nervous systems. *J. Comp. Neurol.* 422, 267–286. doi: 10.1002/(SICI)1096-9861(20000626)422:2<267::AID-CNE8>3.0.CO;2-J
- Nuss, A. B., Forschler, B. T., Crim, J. W., TeBrugge, V., Pohl, J., and Brown, M. R. (2010). Molecular characterization of neuropeptide F from the eastern subterranean termite *Reticulitermes flavipes* (Kollar) (Isoptera: Rhinotermitidae). *Peptides* 31, 419–428. doi: 10.1016/j.peptides.2009.09.001
- O'Brien, L. E., Soliman, S. S., Li, X., and Bilder, D. (2011). Altered modes of stem cell division drive adaptive intestinal growth. *Cell*. 147, 603–614. doi: 10.1016/j.cell.2011.08.048
- Ohler, U. (2006). Identification of core promoter modules in *Drosophila* and their application in accurate transcription start site prediction. *Nucleic Acids Res.* 34, 5943–5950. doi: 10.1093/nar/gkl608
- Okamoto, N., Yamanaka, N., Satake, H., Saegusa, H., Kataoka, H., and Mizoguchi, A. (2009a). An ecdysteroid-inducible insulin-like growth factor-like peptide regulates adult development of the silkworm *Bombyx mori*. *FEBS J.* 276, 1221–1232. doi: 10.1111/j.1742-4658.2008.06859.x
- Okamoto, N., Yamanaka, N., Yagi, Y., Nishida, Y., Kataoka, H., O'Connor, M. B., et al. (2009b). Fat body derived IGF-like peptide regulates post-feeding growth in *Drosophila*. *Dev. Cell* 17, 885–891. doi: 10.1016/j.devcel.2009.10.008
- Oudejans, R. C. H. M., Kooiman, F. P., Heerma, W., Versluis, C., Slotboom, A. J., and Beenackers, A. M. T. (1991). Isolation and structure elucidation of a novel adipokinetic hormone (Lom-AKH-III) from the glandular lobes of the corpus cardiacum of the migratory locust, *Locusta migratoria*. *Eur. J. Biochem.* 195, 351–359. doi: 10.1111/j.1432-1033.1991.tb15713.x
- Overend, G., Cabrero, P., Guo, A. X., Sebastian, S., Cundall, M., Armstrong, H., et al. (2011). The receptor guanylate cyclase Gyc76C and a peptide ligand, NPLP1-VQQ, modulate the innate immune IMD pathway in response to salt stress. *Peptides* 34, 209–218. doi: 10.1016/j.peptides.2011.08.019
- Paemen, L., Schoofs, L., Proost, P., Decock, B., and de Loof, A. (1991b). Isolation, identification and synthesis of Lom-AG-Myotropin-II, a novel peptide in the male accessory reproductive glands of *Locusta migratoria*. *Insect Biochem.* 21, 243–248.

- Paemen, L., Tips, A., Schoofs, L., Proost, P., Van Damme, J., and de Loof, A. (1991a). Lom-AG-myotropin: a novel myotropic peptide from the male accessory glands of *Locusta migratoria*. *Peptides* 12, 7–10. doi: 10.1016/0196-9781(91)90158-L
- Paluzzi, J. P., and 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. doi: 10.1016/j.mce.2012.07.007
- Paluzzi, J. P., Park, Y., Nachman, R. J., and 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. doi: 10.1073/pnas.1003666107
- Panhuis, T. M., Clark, N. J., and Swanson, W. J. (2006). Rapid evolution of reproductive proteins in abalone and *Drosophila*. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 361, 261–268. doi: 10.1098/rstb.2005.1793
- Park, Y., Kim, Y. J., and Adams, M. E. (2002). Identification of G protein-coupled receptors for *Drosophila* PRXamide peptides, CCAP, corazonin, and AKH supports a theory of ligand-receptor coevolution. *Proc. Natl. Acad. Sci. U.S.A.* 99, 11423–11428. doi: 10.1073/pnas.162276199
- Pascual, N., Castresana, J., Valero, M. L., Andreu, D., and Belles, X. (2004). Orocokinin in insects and other invertebrates. *Insect Biochem. Mol. Biol.* 34, 1141–1146. doi: 10.1016/j.ibmb.2004.07.005
- Patel, H., Orchard, I., Veenstra, J. A., and Lange, A. B. (2014). The distribution and physiological effects of three evolutionarily and sequence-related neuropeptides in *Rhodnius prolixus*: adipokinetic hormone, corazonin and adipokinetic hormone/corazonin-related peptide. *Gen. Comp. Endocrinol.* 195, 1–8. doi: 10.1016/j.ygcen.2013.10.012
- Peeff, N. M., Orchard, I., and Lange, A. B. (1994). Isolation, sequence, and bioactivity of PDVDHVFRLFamide and ADVGHVFRLFamide peptides from the locust central nervous system. *Peptides* 15, 387–392. doi: 10.1016/0196-9781(94)90193-7
- Petersen, T. N., Brunak, S., von Heijne, G., and Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods* 8, 785–786. doi: 10.1038/nmeth.1701
- Poels, J., Van Loy, T., Vandersmissen, H. P., Van Hiel, B., Van Soest, S., Nachman, R. J., et al. (2010). Myoinhibiting peptides are the ancestral ligands for the promiscuous *Drosophila* sex peptide receptor. *Cell. Mol. Life Sci.* 67, 3511–3522. doi: 10.1007/s00018-010-0393-8
- Predel, R., and Gäde, G. (2002). Identification of the abundant neuropeptide from abdominal perisymphatic organs of locusts. *Peptides* 23, 621–627. doi: 10.1016/S0196-9781(01)00669-6
- Price, M. N., Dehal, P. S., and Arkin, A. P. (2010). FastTree 2 – approximately maximumlikelihood trees for large alignments. *PLoS ONE* 5:e9490. doi: 10.1371/journal.pone.0009490
- Proux, J. P., Miller, C. A., Li, J. P., Carney, R. L., Girardie, A., Delaage, M., et al. (1987). Identification of an arginine vasopressin-like diuretic hormone from *Locusta migratoria*. *Biochem. Biophys. Res. Commun.* 149, 180–186. doi: 10.1016/0006-291X(87)91621-4
- Rémy, C., and 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. doi: 10.1016/0016-6480(80)90092-1
- Rholam, M., Brakch, N., Germain, D., Thomas, D. Y., Fahy, C., Boussetta, H., et al. (1995). Role of amino acid sequences flanking dibasic cleavage sites in precursor proteolytic processing. The importance of the first residue C-terminal of the cleavage site. *Eur. J. Biochem.* 227, 707–714. doi: 10.1111/j.1432-1033.1995.0707p.x
- Richard, O., Tamarelle, M., Geoffre, S., and Girardie, J. (1994). Restricted occurrence of *Locusta migratoria* ovary maturing parsin in the brain-corpora cardiaca complex of various insect species. *Histochemistry* 102, 233–239. doi: 10.1007/BF00268900
- Richards, S., Gibbs, R. A., Gerardo, N. M., Moran, N., Nakabachi, A., David Stern, D., et al. (2010). Genome sequence of the pea aphid *Acyrtosiphon pisum*. *PLoS Biol.* 8:e1000313. doi: 10.1371/journal.pbio.1000313
- Robb, S., Packman, L. C., and Evans, P. D. (1989). Isolation, primary structure and bioactivity of schistoFLRF-amide, a FMRF-amide-like neuropeptide from the locust, *Schistocerca gregaria*. *Biochem. Biophys. Res. Commun.* 160, 850–856. doi: 10.1016/0006-291X(89)92512-6
- Roller, L., Yamanaka, N., Watanabe, K., Daubnerová, I., Žitňan, D., Kataoka, H., et al. (2008). The unique evolution of neuropeptide genes in the silkworm *Bombyx mori*. *Insect Biochem. Mol. Biol.* 38, 1147–1157. doi: 10.1016/j.ibmb.2008.04.009
- Rosenkilde, C., Cazzamali, G., Williamson, M., Hauser, F., Søndergaard, L., DeLotto, R., et al. (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. doi: 10.1016/j.bbrc.2003.08.022
- Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M. A., et al. (2000). Artemis: sequence visualization and annotation. *Bioinformatics* 16, 944–945. doi: 10.1093/bioinformatics/16.10.944
- Schoofs, L., Holman, G. M., Hayes, T. K., and de Loof, A. (1990d). "Isolation and identification of a sulfakinin-like peptide with sequence homology to vertebrate gastrin and cholecystokinin, from the brain of *Locusta migratoria*," in *Chromatography and Isolation of Insect Hormones and Pheromones*, eds A. McCaffery and I. Wilson (New York, NY: Plenum Press), 231–241.
- Schoofs, L., Holman, G. M., Hayes, T. K., Kochansky, J. P., Nachman, R. J., and de Loof, A. (1990b). Locustatachykinin III and IV: two additional insect neuropeptides with homology to peptides of the vertebrate tachykinin family. *Regul. Pept.* 31, 199–212. doi: 10.1016/0167-0115(90)90006-I
- Schoofs, L., Holman, G. M., Hayes, T. K., Nachman, R. J., and de Loof, A. (1990a). Locustatachykinin I and II, two novel insect neuropeptides with homology to peptides of the vertebrate tachykinin family. *FEBS Lett.* 261, 397–401. doi: 10.1016/0014-5793(90)80601-E
- Schoofs, L., Holman, G. M., Hayes, T. K., Nachman, R. J., and de Loof, A. (1990e). Isolation, identification and synthesis of locustamyotropin II, an additional neuropeptide of *Locusta migratoria*. Member of the cephalomyotropic peptide family. *Insect Biochem.* 20, 479–484
- Schoofs, L., Holman, G. M., Hayes, T. K., Nachman, R. J., and de Loof, A. (1991a). Isolation, primary structure, and synthesis of locustapyrokinin: a myotropic peptide of *Locusta migratoria*. *Gen. Comp. Endocrinol.* 81, 97–104. doi: 10.1016/0016-6480(91)90129-T
- Schoofs, L., Holman, G. M., Hayes, T. K., Nachman, R. J., and de Loof, A. (1991b). Isolation, identification and synthesis of locustamyoinhibiting peptide (LOM-MIP), a novel biologically active neuropeptide from *Locusta migratoria*. *Regul. Pept.* 36, 111–119. doi: 10.1016/0167-0115(91)90199-Q
- Schoofs, L., Holman, G. M., Hayes, T. K., Nachman, R. J., Kochansky, J. P., and de Loof, A. (1992b). Isolation, identification and synthesis of locustamyotropin III and IV, two additional neuropeptides of *Locusta migratoria*: members of the locustamyotropin peptide family. *Insect Biochem. Mol. Biol.* 22, 447–452. doi: 10.1016/0965-1748(92)90140-A
- Schoofs, L., Holman, G. M., Hayes, T. K., Tips, A., Nachman, R. J., Vandesande, E., et al. (1990c). Isolation, identification and synthesis of locustamyotropin (Lom-MT), a novel biologically active insect peptide. *Peptides* 11, 427–433. doi: 10.1016/0196-9781(90)90038-7
- Schoofs, L., Holman, G. M., Nachman, R., Proost, P., Van Damme, J., and de Loof, A. (1993b). Isolation, identification and synthesis of locustapyrokinin II from *Locusta migratoria*, another member of the FXPR-L-amide peptide family. *Comp. Biochem. Physiol. C* 106, 103–109.
- Schoofs, L., Holman, G. M., Paemen, L., Veelaert, D., Amelinckx, M., and de Loof, A. (1993a). Isolation, identification, and synthesis of PDVDHVFRLFamide (SchistoFLRFamide) in *Locusta migratoria* and its association with the male accessory glands, the salivary glands, the heart, and the oviduct. *Peptides* 14, 409–421. doi: 10.1016/0196-9781(93)90126-2
- Schoofs, L., Holman, G. M., Proost, P., Van Damme, J., Hayes, T. K., and de Loof, A. (1992a). Locustakinin, a novel myotropic peptide from *Locusta migratoria*, isolation, primary structure and synthesis. *Regul. Pept.* 37, 49–57. doi: 10.1016/0167-0115(92)90063-Z
- Schoofs, L., Veelaert, D., Holman, G. M., Hayes, T. K., and de Loof, A. (1994). Partial identification, synthesis and immunolocalization of locustamyoinhibin, the third myoinhibiting neuropeptide isolated from *Locusta migratoria*. *Regul. Pept.* 52, 139–156. doi: 10.1016/0167-0115(94)90046-9
- Siebert, K. J. (1999). Locust corpora cardiaca contain an inactive adipokinetic hormone. *FEBS Lett.* 447, 237–240. doi: 10.1016/S0014-5793(99)00299-9
- Siebert, K., Morgan, P., and Mordue, W. (1985). Primary structures of locust adipokinetic hormones II. *Biol. Chem. Hoppe-Seyler.* 366, 723–727. doi: 10.1515/bchm3.1985.366.2.723
- Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* 7, 539. doi: 10.1038/msb.2011.75

- Šimo, L., and Park, Y. (2014). Neuropeptidergic control of the hindgut in the black-legged tick *Ixodes scapularis*. *Int. J. Parasitol.* 44, 819–826. doi: 10.1016/j.ijpara.2014.06.007
- Šimo, L., Žitňan, D., and Park, Y. (2009). Two novel neuropeptides in innervation of the salivary glands of the black-legged tick, *Ixodes scapularis*: myoinhibitory peptide and SIFamide. *J. Comp. Neurol.* 517, 551–563. doi: 10.1002/cne.22182
- Stangier, J., Hilbich, C., Burdzik, S., and Keller, R. (1992). Orcokinin: a novel myotropic peptide from the nervous system of the crayfish, *Orconectes limosus*. *Peptides* 13, 859–864. doi: 10.1016/0196-9781(92)90041-Z
- Sterkel, M., Oliveira, P. L., Urlaub, H., Hernandez-Martinez, S., Rivera-Pomar, R., and Ons, S. (2012). OKB, a novel family of brain-gut neuropeptides from insects. *Insect Biochem. Mol. Biol.* 42, 466–473. doi: 10.1016/j.ibmb.2012.03.003
- Stone, J. V., Mordue, W., Batley, K. E., and Morris, H. R. (1976). Structure of locust adipokinetic hormone, a neurohormone that regulates lipid utilisation during flight. *Nature* 263, 207–211. doi: 10.1038/263207a0
- Stracker, T. H., Thompson, S., Grossman, G. L., Riehle, M. A., and Brown, M. R. (2002). Characterization of the AeaHP gene and its expression in the mosquito *Aedes aegypti* (Diptera: Culicidae). *J. Med. Entomol.* 39, 331–342. doi: 10.1603/0022-2585-39.2.331
- Tamarelle, M., Coast, G. M., and Veenstra, J. A. (2000). Ovary maturing parsin and diuretic hormone are produced by the same neuroendocrine cells in the migratory locust, *Locusta migratoria*. *Peptides* 21, 737–739. doi: 10.1016/S0196-9781(00)00187-X
- Tanaka, Y., Suetsugu, Y., Yamamoto, K., Noda, H., and 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. doi: 10.1016/j.peptides.2013.07.027
- Taussig, R., Kaldany, R. R., and Scheller, R. H. (1984). A cDNA clone encoding neuropeptides isolated from *Aplysia* neuron L11. *Proc. Natl. Acad. Sci. U.S.A.* 81, 4988–4992.
- Tawfik, A. I., Tanaka, S., de Loof, A., Schoofs, L., Baggerman, G., Waelkens, E., et al. (1999). Identification of the gregarization-associated dark-pigmentotropin in locusts through an albino mutant. *Proc. Natl. Acad. Sci. U.S.A.* 96, 7083–7087.
- Terhzaz, S., O'Connell, F. C., Pollock, V. P., Kean, L., Davies, S., Veenstra, J. A., et al. (1999). Isolation and characterization of a leucokinin-like peptide of *Drosophila melanogaster*. *J. Exp. Biol.* 202, 3667–3676.
- Terhzaz, S., Rosay, P., Goodwin, S. F., and Veenstra, J. A. (2007). The neuropeptide SIFamide modulates sexual behavior in *Drosophila*. *Biochem. Biophys. Res. Commun.* 352, 305–310. doi: 10.1016/j.bbrc.2006.11.030
- Terrapon, N., Li, C., Robertson, H. M., Ji, L., Meng, X., Booth, W., et al. (2014). Molecular traces of alternative social organization in a termite genome. *Nat. Commun.* 5, 3636. doi: 10.1038/ncomms4636
- Tyrer, N. M., Davis, N. T., Arbas, E. A., Thompson, K. S., and Bacon, J. P. (1993). Morphology of the vasopressin-like immunoreactive (VPLI) neurons in many species of grasshopper. *J. Comp. Neurol.* 329, 385–401. doi: 10.1002/cne.903290309
- Vanhems, E., Delbos, M., and Girardie, J. (1990). Insulin and neuroparsin promote neurite outgrowth in cultured locust CNS. *Eur. J. Neurosci.* 2, 776–782. doi: 10.1111/j.1460-9568.1990.tb00468.x
- Van Hiel, M. B., Vandersmissen, H. P., Proost, P., and Vanden Broeck, J. (2014). Cloning, constitutive activity and expression profiling of two receptors related to relaxin receptors in *Drosophila melanogaster*. *Peptides*. doi: 10.1016/j.peptides.2014.07.014. [Epub ahead of print].
- Van Kesteren, R. E., Smit, A. B., De Lange, R. P., Kits, K. S., Van Golen, F. A., Van Der Schors, R. C., et al. (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.
- Van Wielendaele, P., Dillen, S., Marchal, E., Badisco, L., and Vanden Broeck, J. (2012). CRF-like diuretic hormone negatively affects both feeding and reproduction in the desert locust, *Schistocerca gregaria*. *PLoS ONE*. 7:e31425. doi: 10.1371/journal.pone.0031425
- Van Wielendaele, P., Dillen, S., Zels, S., Badisco, L., and Vanden Broeck, J. (2013a). Regulation of feeding by Neuropeptide F in the desert locust, *Schistocerca gregaria*. *Insect Biochem. Mol. Biol.* 43, 102–114. doi: 10.1016/j.ibmb.2012.10.002
- Van Wielendaele, P., Wynant, N., Dillen, S., Badisco, L., Marchal, E., and Vanden Broeck, J. (2013c). *In vivo* effect of Neuropeptide F on ecdysteroidogenesis in adult female desert locusts (*Schistocerca gregaria*). *J. Insect Physiol.* 59, 624–630. doi: 10.1016/j.jinsphys.2013.03.005
- Van Wielendaele, P., Wynant, N., Dillen, S., Zels, S., Badisco, L., and Vanden Broeck, J. (2013b). Neuropeptide F regulates male reproductive processes in the desert locust, *Schistocerca gregaria*. *Insect Biochem. Mol. Biol.* 43, 252–259. doi: 10.1016/j.ibmb.2012.12.004
- Veelaert, D., Schoofs, L., Proost, P., Van Damme, J., Devreese, B., Van Beeumen, J., et al. (1995). Isolation and identification of Lom-SG-SASP, a salivation stimulating peptide from the salivary glands of *Locusta migratoria*. *Regul. Pept.* 57, 221–226. doi: 10.1016/0167-0115(95)00035-A
- Veenstra, J. A. (2000). Mono- and dibasic proteolytic cleavage sites in insect neuroendocrine peptide precursors. *Arch. Insect Biochem. Physiol.* 43, 49–63. doi: 10.1002/(SICI)1520-6327(200002)43:2<49::AID-ARCH1>3.0.CO;2-M
- Veenstra, J. A. (2009). Allatostatin C and its paralog allatostatin double C: the arthropod somatostatins. *Insect Biochem. Mol. Biol.* 39, 161–170. doi: 10.1016/j.ibmb.2008.10.014
- Veenstra, J. A. (2010a). Neurohormones and neuropeptides encoded by the genome of *Lotia gigantea*, with reference to other mollusks and insects. *Gen. Comp. Endocrinol.* 167, 86–103. doi: 10.1016/j.ygcen.2010.02.010
- Veenstra, J. A. (2010b). What the loss of the hormone neuroparsin in the *melanogaster* subgroup of *Drosophila* can tell us about its function. *Insect Biochem. Mol. Biol.* 40, 354–361. doi: 10.1016/j.ibmb.2010.03.001
- Veenstra, J. A., Agricola, H.-J., and Sellami, A. (2008). Regulatory peptides in fruit fly midgut. *Cell Tissue Res.* 334, 499–516. doi: 10.1007/s00441-008-0708-3
- Veenstra, J. A., and Ida, T. (2014). More *Drosophila* enteroendocrine peptides: orcokinin B and the CCHamides 1 and 2. *Cell Tissue Res.* 357, 607–621. doi: 10.1007/s00441-014-1880-2
- Veenstra, J. A., and Lambrou, G. (1995). Isolation of a novel RFamide peptide from the midgut of the American cockroach, *Periplaneta americana*. *Biochem. Biophys. Res. Commun.* 213, 519–524. doi: 10.1006/bbrc.1995.2162
- Veenstra, J. A., Patillo, J. M., and Petzel, D. H. (1997). A single cDNA encodes all three *Aedes* leucokinin, which stimulate both fluid secretion and hindgut contractions. *J. Biol. Chem.* 272, 10402–10407. doi: 10.1074/jbc.272.16.10402
- Veenstra, J. A., Rombauts, S., and 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. doi: 10.1016/j.ibmb.2011.12.009
- Vogel, K. J., Brown, M. R., and Strand, M. R. (2013). Phylogenetic investigation of peptide hormone and growth factor receptors in five dipteran genomes. *Front. Endocrinol. (Lausanne)*. 4:193. doi: 10.3389/fendo.2013.00193
- Wang, X., Fang, X., Yang, P., Jiang, X., Jiang, F., Zhao, D., et al. (2014). The locust genome provides insight into swarm formation and long-distance flight. *Nat. Commun.* 5, 2957. doi: 10.1038/ncomms3957
- Wegener, C., Reintl, T., Jänsch, L., and 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. doi: 10.1111/j.1471-4159.2005.03634.x
- Weinbörmair, G., Pohlhammer, K., and Dürnberger, H. (1975). Das Axonsystem des ventro-medianen neurosekretorischen Zellpaares im Suboesophagealganglion bei *Teleogryllus commodus* Walker. Darstellung im Totalpräparat mit Hilfe von Resorcinfuchsin. *Mikroskopie* 31, 147–154.
- Wolfner, M. F. (2002). The gifts that keep on giving: physiological functions and evolutionary dynamics of male seminal proteins in *Drosophila*. *Heredity (Edinb)*. 88, 85–93. doi: 10.1038/sj.hdy.6800017
- Yamada, K., Hirotsu, T., Matsuki, M., Butcher, R. A., Tomioka, M., Ishihara, T., et al. (2010). Olfactory plasticity is regulated by pheromonal signaling in *Caenorhabditis elegans*. *Science* 329, 1647–1650. doi: 10.1126/science.1192020
- Yamanaka, N., Hua, Y.-J., Roller, L., Spalovská-Valachová, I., Mizoguchi, A., Kataoka, H., et al. (2010). *Bombxy* prothoracicostatic peptides activate the sex peptide receptor to regulate ecdysteroid biosynthesis. *Proc. Natl. Acad. Sci. U.S.A.* 107, 2060–2065. doi: 10.1073/pnas.0907471107
- Yamanaka, N., Yamamoto, S., Žitňan, D., Watanabe, K., Kawada, T., Satake, H., et al. (2008). Neuropeptide receptor transcriptome reveals unidentified neuroendocrine pathways. *PLoS ONE* 3:e3048. doi: 10.1371/journal.pone.0003048
- Zandawala, M., Li, S., Hauser, F., Grimmekhuijzen, C. J. P., and Orchard, I. (2013). Isolation and functional characterization of calcitonin-like diuretic hormone

receptors in *Rhodnius prolixus*. *PLoS ONE* 8:e82466. doi: 10.1371/journal.pone.0082466

Zerbino, D. R., and Birney, E. (2008). Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome Res.* 18, 821–829. doi: 10.1101/gr.074492.107

Conflict of Interest Statement: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 14 September 2014; paper pending published: 12 October 2014; accepted: 03 November 2014; published online: 19 November 2014.

Citation: Veenstra JA (2014) The contribution of the genomes of a termite and a locust to our understanding of insect neuropeptides and neurohormones. *Front. Physiol.* 5:454. doi: 10.3389/fphys.2014.00454

This article was submitted to *Invertebrate Physiology*, a section of the journal *Frontiers in Physiology*.

Copyright © 2014 Veenstra. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.