General and Comparative Endocrinology 229 (2016) 41–55

Contents lists available at ScienceDirect

General and Comparative Endocrinology

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

Neuropeptide evolution: Chelicerate neurohormone and neuropeptide genes may reflect one or more whole genome duplications

Jan A. Veenstra *

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

ARTICLE INFO

Article history:
Received 12 October 2015
Revised 20 November 2015
Accepted 29 November 2015
Available online 27 February 2016

Keywords:
Arthropods
CCRFamide
GPCR
Myriapod
House dust mite
Scorpion
Spider

ABSTRACT

Four genomes and two transcriptomes from six Chelicerate species were analyzed for the presence of neuropeptide and neurohormone precursors and their GPCRs. The genome from the spider *Stegodyphus mimosarum* yielded 87 neuropeptide precursors and 120 neuropeptide GPCRs. Many neuropeptide transcripts were also found in the transcriptomes of three other spiders, *Latrodectus hesperus*, *Parasteatoda tepidariorum* and *Acanthoscurria geniculata*. For the scorpion *Mesobuthus martensii* the numbers are 79 and 93 respectively. The very small genome of the house dust mite, *Dermatophagoides farinae*, on the other hand contains a much smaller number of such genes. A few new putative Arthropod neuropeptide genes were discovered. Thus, both spiders and the scorpion have an achatin gene and in spiders there are two different genes encoding myosuppressin-like peptides while spiders also have two genes encoding novel LGamides. Another finding is the presence of trissin in spiders and scorpions, while neuropeptide genes that seem to be orthologs of *Lottia* LFRYamide and *Platynereis* CCRFamide were also found. Such genes were also found in various insect species, but seem to be lacking from the Holometabola. The Chelicerate neuropeptide and neuropeptide GPCR genes often have paralogs. As the large majority of these are probably not due to local gene duplications, is plausible that they reflect the effects of one or more ancient whole genome duplications.

1. Introduction

Neuropeptides and neurohormones regulate and/or modulate many biological processes from basal physiological functions such as carbohydrate metabolism and water balance to cognitive functions. It is possible that they were among the first chemical messengers used by the nervous and endocrine systems to establish communication between different cells of a single organism. In the last two decades a large number of genomes have been sequenced which allow in theory to identify all the neuropeptide genes they contain. In practice, this is often more difficult as neuropeptide sequences are small and relatively variable and are usually only recognized when they either code for homologs of previously identified neuropeptides or a number of very similar peptide sequences separated by putative convertase cleavage sites. Nevertheless, it seems likely that most neuropeptide genes can be identified in a given sequenced Arthropod genome.

The large majority of neuropeptides act through G-protein coupled receptors (GPCRs) which are readily identified from genome sequences for the presence of their seven transmembrane regions that are well conserved. Thus, by analyzing simultaneously both neuropeptide and their putative receptors in its genome it is possible to get a fairly complete view of the neuropeptidome of a species. Using such methods it was previously shown that the neuropeptidomes of insects, annelids and mollusks are remarkably similar and share a large number of neuropeptide genes (Veenstra, 2010, 2011; Stewart et al., 2014).

As expected, neuropeptides and their GPCRs identified from the spider mite *Tetranychus urticae*, the first Chelicerate for which a complete genome was sequenced (Grbić et al., 2011), are most similar to those of insects (Veenstra et al., 2012). However, some of its neuropeptide genes were previously only known from mollusks. One of these, elevenin, has subsequently also been found in insects (Tanaka et al., 2014; Veenstra, 2014). Since then draft genomes have been published for four other Chelicerates, those of the African social velvet spider, *Stegodyphus mimosarum*, the Brazilian white-knee tarantula, *Acanthoscurria geniculata* (Sanggaard et al., 2014), the scorpion *Mesobuthus martensii* (Cao et al., 2013), and the house dust mite *Dermatophagoides farinae* (Chan et al., 2015). At the same time spider transcriptomes have become available for *Latrodectus hesperus* and *Parasteatoda tepidariorum* (Clarke et al., 2014; Posnien et al., 2014). It thus seemed interesting to take

* Address: INCIA UMR 5287 CNRS, Université de Bordeaux, allée Geoffroy St Hillaire, CS 50023, 33 615 Pessac Cedex, France.
E-mail address: jan.veenstra@u-bordeaux.fr

http://dx.doi.org/10.1016/j.ygcen.2015.11.019
0016-6480/

another look at Chelicerate neuropeptides to complete the picture of Arthropod neuropeptides and their evolution. Christie (2015) has already published a number of spider neuropeptide precursors based on the *Latrodectus* transcriptome, but, as shown here, the actual number of neuropeptide precursors in this transcriptome is considerably larger and analyzing whole genomes reveals additional neuropeptide genes that are not represented in the assembled transcriptomes. While this work was in progress the genome of the Myriapod *Strigamia maritima* was also published (Chipman et al., 2014). I included this genome in the analysis as the phylogenetic position of the Myriapods falls between insects and Chelicerates (Fig. 1). I felt it would be interesting to have Protostomian outgroups in the phylogenetic analysis of GPCRs, as it was expected that the more recently shared ancestry might lead to better resolved phylogenetic trees. I, therefore, also predicted the neuropeptide GPCRs from the genome of the mollusk *Lottia gigantea* (Simakov et al., 2013), a species from which a large number of neuropeptide genes has previously been described (Veenstra, 2010; Roch et al., 2011; Mirabeau and Joly, 2013). More recently a large number of GPCRs from the annelid *Platynereis dumerilii* were published, some of which were deorphanized (Bauknecht and Jékely, 2015). Given the interest of this data set together with the previously published list of putative neuropeptide precursors from this species (Conzelmann et al., 2013) those GPCR sequences were also included in the analysis. Apart from discovering a few novel putative Arthropod neuropeptide genes, the more interesting findings are that many spider and scorpion genes encoding neuropeptides and their receptors have been duplicated, perhaps the result of one or more whole genome duplications.

2. Materials and methods

Local BLAST (Altschul et al., 1997; Camacho et al., 2009) was used to analyze the published transcriptome of *Latrodectus* as well as the *Stegodyphus*, *Mesobuthus*, *Strigamia* and *Dermatophagoides* genomes (all obtained from NCBI) as well as the *Parasteatoda* transcriptome (downloaded from http://asgard.rc.fas.harvard.edu/download.html). The *Acanthoscurria* genome was analyzed directly at NCBI using the web interface and contigs that might contain neuropeptide genes or parts thereof were downloaded for further analysis. The peptide sequences of known arthropod and molluscan neuropeptides and their G-protein coupled receptors were used as queries in the BLAST searches. The neuropeptide and GPCR sequences were from the following species, *T. urticae* (Veenstra et al., 2012), *Locusta migratoria*, *Zootermopsis nevadensis* (Veenstra, 2014) and *L. gigantea* (Veenstra, 2010). It should be noted that such homology searches are often limited in determining the correct C- and N-terminal parts of protein sequences when there are no or very incomplete transcriptome data available. Gene models were constructed using Artemis (Rutherford et al., 2000) as described previously (Veenstra, 2014). The assembled *Mesobuthus* transcriptome was downloaded from http://lifecenter.sgst.cn/main/en/scorpion.jsp. Unfortunately, the original RNAseq data from *Mesobuthus* are not in the data bases and could thus not be used. Short sequence read archives (SRAs) containing RNAseq data from *Stegodyphus*, *Strigamia* and *Dermatophagoides* were decompressed with the SRA toolkit (http://www.ncbi.nlm.nih.gov/Traces/sra/?view=software) and used to make BLAST databases. The latter were then searched for sequences coding for the predicted RNA sequences that were subsequently used as input for the Trinity program (Haas et al., 2013) in order to check, correct and/or improve the various gene models. This procedure was very effective for *Strigamia* not only for improving the predicted neuropeptide GPCR gene models, but also for the completion and correction of several previously predicted neuropeptide precursors (Chipman et al., 2014). This method furthermore identified the *Strigamia* allatostatin C receptor, for which the gene is lacking from the genome assembly (Chipman et al., 2014). In *Stegodyphus* and *Dermatophagoides* on the other hand it led to just a few corrections as the RNAseq data is less extensive. However, the various transcriptomes from three *Stegodyphus* species (Mattila et al., 2012) were helpful in completing predicted neuropeptide precursors from the *Stegodyphus* genome. Differences in the quality of the genome assemblies analyzed here are explained by variables such as the absence or presence of homozygocity, sufficient coverage, existence of good transcriptome data and length of sequence reads (Richards and Murali, 2015).

When novel neuropeptides genes were found, their presence in other species was studied by using the BLAST program on the various protostomian genomes, transcriptome assemblies and ESTs collections in the NCBI data bases.

In the four Chelicerate genome assemblies scaffolds are either absent or relatively small and thus the positions of the various genes relative to one another are known in only a few cases. This makes it also impossible to predict complete protein sequences when different exons of the same gene are located in different contigs. This is particularly a problem for the prediction of GPCRs where introns of more than 100,000 base pairs are not rare. In those cases where different exons were sufficiently similar to orthologs from other Arthropods and when the genome appeared to encode a single copy of a particular receptor, I have taken the liberty to join those different contigs or scaffolds into a single pseudo-scaffold for reconstructing the GPCR. The same was done when there were two copies of a gene with one present in a single continuous DNA sequence and the other in various pieces. However, numerous Chelicerate GPCRs exist as three or four copies

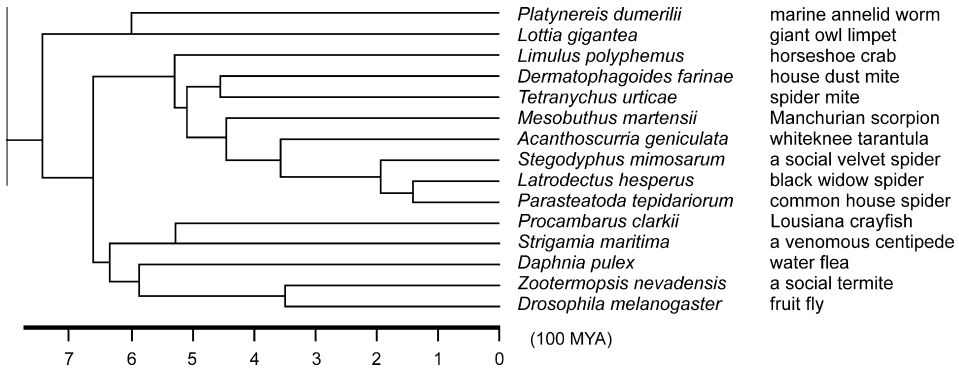

Fig. 1. Simplified phylogenetic tree of the species analyzed in this paper as well as a few others. Note that the Chelicerate phylogeny has not been unambiguously resolved (Sharma et al., 2014). MYA, million years ago.