# Evolution of Neuropeptide Precursors in Polyneoptera (Insecta)

Inaugural-Dissertation

zur

Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultät

der Universität zu Köln

vorgelegt von

**Marcel Bläser**

aus Bonn-Duisdorf

Köln, Juli 2020

Berichterstatter (Gutachter):

Prof. Dr. Reinhard Predel

PD Dr. Benjamin Altenhein

Tag der mündlichen Prüfung:

10.09.2020

# Table of Contents

# List of Abbreviations

## General

| | |
|---|---|
| AA | Amino acid |
| AED | Average evolutionary divergence |
| CNS | Central nervous system |
| ER | Endoplasmic reticulum |
| GPCR | G-protein coupled receptor |
| NCA-2 | *Nervus corporis allati*-2 |
| NCC-3 | *Nervus corporis cardiaci*-3 |
| PAM | Peptidyl-aminotransferase |
| PC | Prohormone convertases |
| SEG | subesophageal ganglion |
| VPMNs | ventral posterior median neurons |

## Neuropeptides

| | |
|---|---|
| ACP | Adipokinetic hormone/ corazonin-like peptide |
| AKH | Adipokinetic hormone |
| AST-C | Allatostatin-C |
| AST-CC | Allatostatin-CC |
| AST-CCC | Allatostatin-CCC |
| AT | Allatotropin |
| CCAP | Crustacean cardioactive peptide |
| CRF-DH | Corticotropin-releasing factor-like diuretic hormone |
| CT-DH/DH-31 | calcitonin-like diuretic hormone |
| NPF-1 | Neuropeptide F-1 |
| NPF-2 | Neuropeptide F-2 |
| MS | Myosuppressin |
| pk | Pyrokinin |
| pk-like | Pyrokinin-like |
| sNPF | Short Neuropeptide F |
| tryptoPK | Tryptopyrokinin |

## Amino acids

| | |
|---|---|
| Alanine | Ala, A |
| Arginine | Arg, R |
| Asparagine | Asn, N |
| Aspartate | Asp, D |
| Cysteine | Cy, C |
| Glutamate | Glu, E |
| Glutamine | Gln, Q |
| Glycine | Gly, G |
| Histidine | His, H |
| Isoleucine | Ile, I |
| Leucine | Leu, L |
| Lysine | Lys, K |
| Methionine | Met, M |
| Phenylalanine | Phe, F |
| Proline | Pro, P |
| Serine | Ser, S |
| Threonine | Thr, T |
| Tryptophan | Trp, W |
| Tyrosine | Tyr, Y |
| Valine | Val, V |

# Abstract

Neuropeptides are one of the major players in regulation of growth, reproduction and development as well as in information transfer in all multicellular animal organisms (Metazoa). Structurally, these short protein-like substances are among the most diverse signaling molecules, although they are under a stabilizing evolutionary pressure due to co-evolution with their respective receptor. Many of the peptidergic systems have a very ancient origin that can be traced back to the early evolution of Metazoa. In arthropods, especially insects, neuropeptide sequences and functions have been intensely studied in the past. In recent years, the increasing availability of genome and transcriptome data facilitated the investigation of complete neuropeptide precursor sequences and aided in the discovery of new neuropeptide genes. In my thesis, mainly transcriptome data were used to study the evolution of neuropeptide precursor sequences of polyneopteran insects, which resulted in three publications.

The first study represents an in-detail analysis of a novel neuropeptide gene, *tryptopyrokinin*. This gene was recently predicted from the genome of the migratory locust, *Locusta migratoria*, but no data on its expression were known from any organism. In this study, by means of transcriptomics and single cell mass spectrometry, products of two novel *tryptoPK* genes were identified in the gnathal ganglion of the locust. In addition, the evolutionary history of the *tryptoPK* genes was discussed using comparative neuroendocrinology and the sequences of their probable ancestral genes, *capa* and *pyrokinin*, were completed from newly generated transcriptomes.

In the second study, I used comprehensive transcriptome data from over 200 polyneopteran species, provided by the 1KITE initiative, to investigate the evolution of single-copy neuropeptide precursor sequences in Polyneoptera. I compiled a list of 21 orthologous single-copy neuropeptide precursors from all polyneopteran orders and discussed the degree of sequence conservation in the different neuropeptide families and between the different polyneopteran lineages. Sequence logo depictions based on the degree of sequence conservation were created. By mapping these sequence logos on the latest phylogeny of Polyneoptera, we were able to reveal lineage-specific sequence motifs (synapomorphies) across Polyneoptera. The data

further highlight that the sequences of the mature bioactive peptide show a higher degree of sequence conservation than the sequences of the remaining precursor. For most neuropeptide precursor families, the individual sequences of the mature neuropeptide that were present at the emergence of insects 400 mya could be postulated. We were able to show that while the degree of sequence conservation varies between the different neuropeptides and between orthologs of the same neuropeptide in different lineages, overall, the single-copy neuropeptide precursors of Polyneoptera show a relatively high degree of sequence conservation.

In the third study, we used neuropeptide precursor sequences to reconstruct phylogenetic relationships inside the polyneopteran order Blattodea. Amino acid sequences of 17 neuropeptide precursors from 40 blattodean species were successfully used to corroborate a recently proposed phylogeny of Blattodea that relied on 2,370 protein-coding nuclear single-copy genes. This illustrated the phylogenetic information encoded on neuropeptide precursor sequences and their power for phylogenetic analyses. In addition, the functionally different units of a neuropeptide precursor were analyzed separately for the first time. This revealed different degrees of phylogenetic information and hence sequence conservation in the different units. Nevertheless, the phylogenetic information of each unit was sufficient to resolve the phylogeny at the ordinal level. This study represents the first comprehensive use of complete neuropeptide precursor sequences for phylogenetic analyses and demonstrates their applicability as alternative dataset for the phylogenetic analyses of ambiguous sister-group relationships.

In my dissertation, I could generate valuable insight into the evolution of neuropeptide precursor sequences. Specifically, I was able to demonstrate that neuropeptide precursor sequence evolution does not follow a uniform rate. It is unique in every precursor. Albeit a rather high degree of sequence conservation in Polyneoptera, differences in evolutionary rate can be seen between orthologs of different lineages. Even on a single neuropeptide precursor, the degree of sequence conservation differs between the functional units. Furthermore, I was able to establish neuropeptide precursor sequences as a basis for phylogenetic analyses.

# Zusammenfassung

Neuropeptide sind einer der Hauptakteure bei der Steuerung von Wachstum, Reproduktion und Entwicklung sowie bei der Informationsübertragung in allen vielzelligen tierischen Organismen (Metazoen). Diese kurzen, proteinähnlichen Substanzen gehören zu den strukturell vielfältigsten Signalmolekülen, obwohl sie aufgrund ihrer Koevolution mit dem jeweiligen Rezeptor unter einem stabilisierenden evolutionären Druck stehen. Viele peptiderge Systeme lassen sich in ihrem Ursprung bis auf die frühe Evolution der Metazoen zurückführen. Bei Arthropoden, insbesondere Insekten, wurden Neuropeptidsequenzen und deren Funktionen in der Vergangenheit intensiv untersucht. In den letzten Jahren erleichterte die zunehmende Verfügbarkeit von Genom- und Transkriptomdaten die Untersuchung der Sequenzen kompletter Neuropeptidpräkursoren. In meiner Dissertation untersuchte ich die Evolution der Sequenzen von Neuropeptidpräkursoren in Polyneoptera mittels Transkriptomanalyse, was zur Veröffentlichung von drei Publikationen führte.

Die erste Studie meiner Arbeit stellt eine Detailanalyse eines neuartigen Neuropeptidgens dar, *tryptopyrokinin*. Dieses Gen wurde vor kurzem in dem Genom der Wanderheuschrecke *Locusta migratoria* gefunden, Daten über seine Expression waren jedoch bisher von keinem Organismus bekannt. In dieser Studie wurden mittels Transkriptomik und Einzelzellmassenspektrometrie Produkte von zwei neuartigen *tryptoPK*-Genen im Unterschlundganglion der Wanderheuschrecke identifiziert. Darüber hinaus wurde die Evolutionsgeschichte der *tryptoPK*-Gene mit Hilfe vergleichender Neuroendokrinologie diskutiert und die Sequenzen ihrer wahrscheinlichen Vorfahrgene, *capa* und *pyrokinin*, aus neu generierten Transkriptomen vervollständigt.

In der zweiten Studie verwendete ich umfassende Transkriptomdaten von über 200 Arten der Polyneoptera, um die Evolution der Sequenzen von Einzelkopie-Neuropeptid-Präkursoren in Polyneoptera zu untersuchen. Ich stellte eine Liste von 21 orthologen Einzelkopie-Neuropeptidpräkursoren aus allen Ordnungen der Polyneoptera zusammen und diskutierte den Grad der Sequenzerhaltung innerhalb der verschiedenen Neuropeptidfamilien und zwischen den verschiedenen Linien der Polyneoptera. Es wurden Sequenzlogo-Darstellungen des Grades der Sequenzkonservierung erstellt. Durch Kartierung dieser Sequenzlogos auf die neueste Phylogenie der Polyneoptera konnten wir linienspezifische Sequenzmotive (Synapomorphien) innerhalb der Polyneoptera aufdecken. Die Daten zeigen außerdem, dass die Sequenzen des reifen bioaktiven Peptids einen höheren Grad an Sequenzkonservierung aufweisen als die Sequenzen des verbleibenden

Präkursors. Für die meisten Neuropeptidpräkursorfamilien konnten die jeweiligen Sequenzen des reifen Neuropeptids, wie sie beim Entstehen der Insekten vor 400 Millionen Jahren vorhanden waren, postuliert werden. Wir konnten zeigen, dass die Einzelkopie-Neuropeptidpräkursoren der Polyneoptera insgesamt einen relativ hohen Grad der Sequenzkonservierung aufweisen, obwohl der Grad der Konservierung zwischen den verschiedenen Neuropeptiden und zwischen Orthologen desselben Neuropeptids in den verschiedenen Abstammungslinien variiert.

In der dritten Studie verwendeten wir die Sequenzen von Neuropeptidpräkursoren, um die phylogenetischen Beziehungen innerhalb der Polyneoptera Ordnung Blattodea zu rekonstruieren. Aminosäuresequenzen von 17 Neuropeptidpräkursoren aus 40 Arten der Blattodea wurden verwendet, um eine kürzlich veröffentlichte Phylogenie der Blattodea zu bestätigen, welche auf 2.370 protein-kodierende, nukleäre Einzelkopie-Gene beruhte. Dies veranschaulicht die phylogenetische Information, die auf den Sequenzen der Neuropeptidpräkursoren kodiert wird, und ihre Aussagekraft für phylogenetische Analysen. Darüber hinaus wurden die funktionell unterschiedlichen Einheiten eines Neuropeptidpräkursors zum ersten Mal getrennt analysiert. Hier konnten wir unterschiedliche Grade an phylogenetischer Information, und damit einhergehend, Sequenzerhaltung in den verschiedenen Einheiten nachweisen. Dennoch war die phylogenetische Information jeder einzelnen Einheit ausreichend, um die Phylogenie auf der ordinalen Ebene aufzuschlüsseln. Diese Studie stellt die erste umfassende Verwendung vollständiger Neuropeptidpräkursoren für phylogenetische Analysen dar und demonstriert ihre Anwendbarkeit als alternativer Datensatz für die phylogenetische Analyse unklarer Schwestergruppenbeziehungen.

In meiner Dissertation konnte ich wertvolle Einblicke in die Evolution der Sequenzen von Neuropeptidpräkursoren gewinnen. Ich konnte insbesondere zeigen, dass die Evolution der Neuropeptidpräkursoren keiner einheitlichen Geschwindigkeit folgt, sondern bei jedem Präkursor einzigartig ist. Trotz eines vergleichsweise hohen Grades an Sequenzkonservierung in Polyneoptera lassen sich Unterschiede in der Evolutionsrate zwischen Orthologen verschiedener Abstammungslinien feststellen. Selbst innerhalb eines einzelnen Neuropeptidpräkursors unterscheidet sich der Konservierungsgrad der Sequenz zwischen den funktionellen Einheiten. Darüber hinaus gelang es mir, Neuropeptidpräkursoren als Grundlage für phylogenetische Analysen zu etablieren.

# 1. Introduction

Before the discovery of neuropeptides, their functions were already known to humankind. In ancient times, extracts of certain organs were said to have stimulating or healing effects (Kleine and Rossmanith, 2016). Today, we know that these effects are based on the presence of neuropeptides in said organs. Neuropeptides represent the most diverse group of signalling molecules in all metazoan animals. These protein-like, short-sequence molecules (<50 amino acids (AA)) influence almost all bodily and behavioural functions, ranging from growth to mating to feeding. They are described as pleiotropic, i.e., acting on a variety of cell types and tissues (Nässel and Winther, 2010; Fricker, 2012; Schoofs et al., 2017; Nässel and Zandawala, 2019).

The term 'neuropeptide' was first coined by David de Wied to describe peptide hormones that are active on neural substrates (Wied, 1971; Wied, 2000). Neuropeptides derive through a specialized biosynthesis pathway from larger precursor proteins (Brownstein et al., 1980; Zupanc, 1996). This pathway is called the 'regulated secretory pathway' (Figure 1). A newly synthesized amino acid sequence, the neuropeptide preprohormone (also called precursor) is selected and sorted into the lumen of the endoplasmic reticulum (ER) and Golgi apparatus complex. Preprohormones are recognized by their signal peptide. This N-terminal sequence, usually 16-30 AA long, facilitates the entrance into the ER and is immediately cleaved off afterwards (Blobel and Dobberstein, 1975; Owji et al., 2018). Therefore, signal peptide sequences are not found in studies on the proteomic level. In the ER, the remaining protein (now called prohormone) is further proteolytically processed. Here, specialised endoproteases (furins, prohormone convertases [PCs]), cleave the protein into specific peptides at basic cleavage motifs (Hook et al., 2008). These motifs are mostly dibasic pairs of arginine (Arg, R) and lysine (Lys, K). Tribasic, and, under certain structural environments, also monobasic cleavage sites, are commonly found as well (Veenstra, 2000). The resulting peptides are determined by the respective PC. Since the PC content of a cell varies, cleavage of prohormones can be cell-specific, and a prohormone can lead to a different set of neuropeptides in different cells (e.g., Neupert et al., 2018). The cleavage of a prohormone does not only produce neuropeptides

that exert their biological activities through binding to a receptor, but it also leads to the release of so-called precursor peptides. These precursor peptides are not known to be receptor triggering and cause no known biological activity. It is hypothesized that they function as recognition sites for PCs (Wegener and Gorbashov, 2008; Pauls et al., 2014). Neuropeptides are, in the majority of cases, post-translationally modified. This usually includes C-terminal amidation. Here, peptidyl-aminotransferase (PAM) oxidizes a C-terminal glycine (Gly, G) into an amidation on the N-terminally preceding amino acid. The resulting amidated C-terminus leads to increased longevity of the neuropeptide and can be required for steric properties and recognition by the respective receptor (Merkler, 1994). The majority of all biologically active peptides are C-terminally amidated (Kim and Seong, 2001). In insects, more than 80% of neuropeptides possess a C-terminal amidation (Kim and Seong, 2001, Liessem et al., 2018). Further post-translational modifications include formation of di-sulphide bridges, oxidation, sulfation and cyclization of an N-terminal glutamine (Gln, Q) or, more rarely, glutamic acid (Glu, E), into a pyroglutamate residue (Garden et al., 1999; Predel et al., 1999). These modifications in contrast to e.g., the N-terminal amidation can occur spontaneously and must not depend on enzyme activity. After cleavage, the matured neuropeptides are stored into large, dense-core vesicles derived from the *trans*-Golgi network. Packed into these vesicles, neuropeptides can be released into the extracellular matrix by fusion of the vesicles with the cell membrane upon neuronal activity (Brownstein et al., 1980; Eckert et al., 2002; Pollák et al., 2005). The releasing stimulus is usually a sudden elevation of $Ca^{2+}$ levels in the compartment.

This specialized pathway is responsible for the unique release response of neuropeptides. The response time in the respective target cell after neuropeptide release is usually slow compared to the immediate response elicited by small neurotransmitters (e.g., amino acids, biogenic amines) that modulate ion concentrations directly by impacting ion channels (van den Pol, 2012). The advantage of the regulated response, however, is the simultaneous release of a significant number of stored vesicles in a major surge. The slow response of neuropeptide release is tied to the receptor type that most neuropeptides trigger: G-protein coupled receptors (GPCRs) (Vaudry and Seong, 2014). Effects triggered by GPCRs can be active for multiple seconds and therefore significantly longer than the results triggered, for example, by the faster responses of small neurotransmitters (van den Pol, 2012).

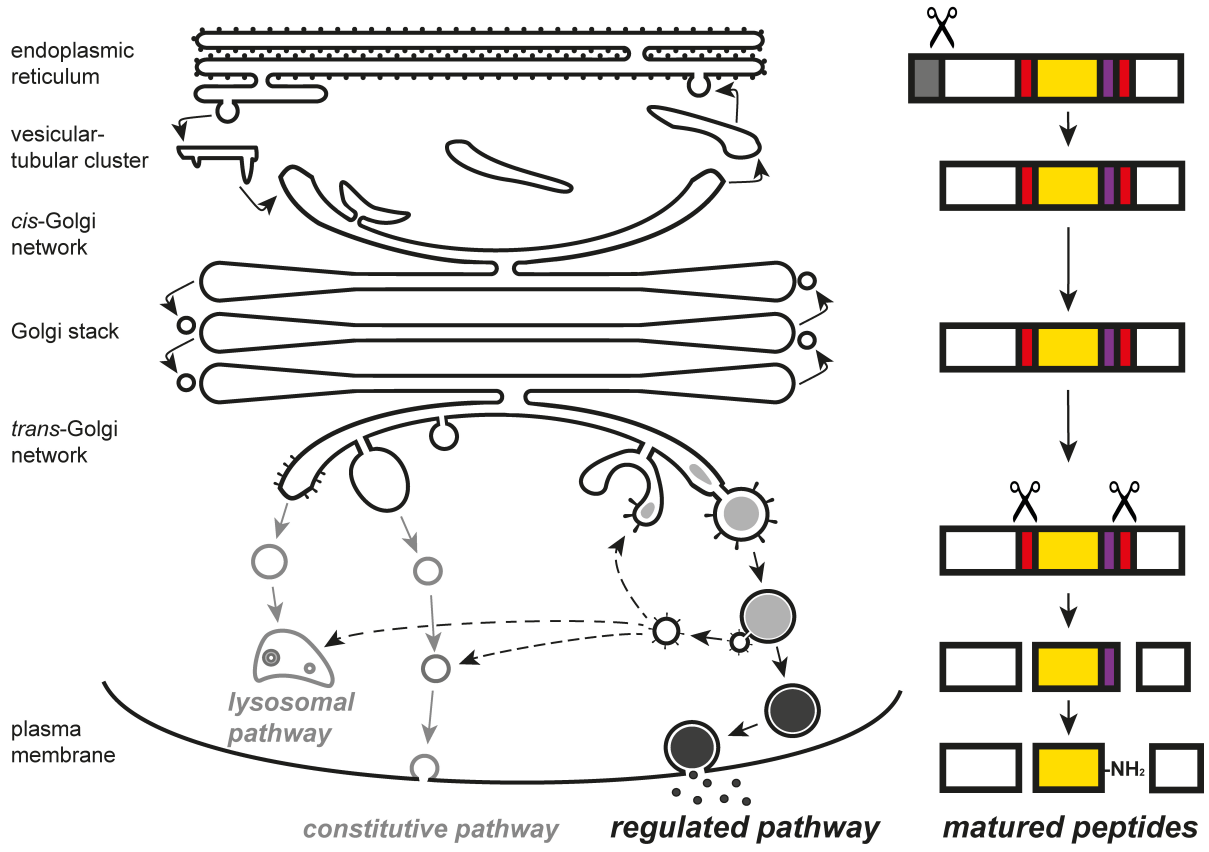



**Figure 1: Schematic overview of the biosynthesis and transport of neuropeptides through the regulated secretory pathway. The involved subcompartments are depicted along with the respective processing steps. The signal peptide (grey bar) is cleaved off of the preprohormone by entry into the secretion routes, only the prohormone remains. In the *trans*-Golgi network, neuropeptides are packed into vesicles, in which further maturing occurs. Proteolytic enzymes cleave the prohormone at specific motives (red bars) into neuropeptides and precursor peptides. Posttranslational modifications, e.g., amidation (-$NH_2$), occur further on. The purple bar represents an N-terminal Gln. Mature, dense-core vesicles are stored and released upon stimulus. For comparison, the constitutive pathway and the lysosomal pathway (depicted in grey) are shown. Adapted from Murphy and Gainer, 2016.**

Many GPCRs are only activated by a single type of ligand. This leads to a distinct co-evolution between ligand and receptor. To retain the functionality of the receptor-ligand relationship, mutations of the amino acid sequence of neuropeptides must be mirrored by mutations on the sequence of the receptor that retain the binding properties. Otherwise, the mutations of the neuropeptides must not alter their steric properties or represent silent mutations. This co-evolution results in a strong stabilizing selection on the sequences of neuropeptides (Hauser et al., 2006; Wegener and Gorbashov, 2008; Grimmelikhuijzen and Hauser, 2012). The C-terminus, which is usually the major site for interactions between neuropeptides and their respective receptor, is especially affected by this stabilizing selection (e.g., Sullivan et al., 1987; Simonds et al., 1989; Conklin et al., 1996). Orthologous neuropeptides of different species can

often be identified by their conserved C-terminal sequences (e.g., the C-terminal FXRFamide in extended FMRFamide-like peptides, Taussig and Scheller, 1986; Schneider and Taghert, 1988; Nagata, 2016). Similar C-terminal sequences of different neuropeptides can also be found in a single species. In fact, some receptors are triggered by multiple neuropeptides, even from different genes (Li and Kim, 2008). These neuropeptides often belong to the same neuropeptide family. Neuropeptide families are mostly defined by their shared C-terminal sequence motifs and primary amino-acid structure (Coast and Schooley, 2011; Merighi, 2011). These sequences are often paralogs of a common ancestral gene. Paralog genes can undergo separate mutations and gain new functions over time (e.g., Hansen et al., 2010).

Duplications do not always occur on the level of complete genes. When the DNA segment that encodes the neuropeptide on the gene is duplicated internally, this leads to the presence of multiple paracopies of a neuropeptide-coding sequence on the gene (Wegener and Gorbashow, 2008). There are cases of neuropeptide precursors with over 20 paracopies, like the extended FMRFamides in *Aplysia californica* (Taussig and Scheller, 1986). Paracopies are defined as copies of a neuropeptide encoded on the same gene in a single species (Wegener and Gorbashov, 2008; Larhammar, 2009). These internal duplications lead to a differentiation of neuropeptide precursors into single-copy and multiple-copy neuropeptide precursors. Products of multiple-copy neuropeptides are often structurally similar and target the same receptors. This leads to redundancy in the signal transmission, which lowers the stabilizing selection pressure on the individual neuropeptide sequence. Single-copy neuropeptides are therefore usually better conserved than their multiple-copy counterparts (Wegener and Gorbashov, 2008).

Ever since their discovery in humans and other vertebrates, neuropeptides have been also extensively studied in non-vertebrate taxa. Due to their shorter reproduction rate, simpler nervous systems and smaller genomes, insects have emerged as model organisms for studies on neuropeptides (Starratt and Brown, 1975; Loof and Schoofs, 1990; Nässel, 1993; Li et al., 2008; Ons et al., 2009; Hansen et al., 2010; Huybrechts et al., 2010; Predel et al., 2010; Schoofs et al., 2017; Pandit et al., 2018; Predel et al., 2018; Nässel and Zandawala, 2019) and GPCRs (Li et al., 1999; reviewed in Caers et al., 2012). In fact, multiple neuropeptides have been first described in insects and their homologs were later identified in vertebrates (Loof and Schoofs, 1990). Aided by findings in insects, among others, recent studies come to the conclusion that peptidergic

systems are conserved throughout Bilateria and presumably have a pre-bilaterian origin (Grimmelikhuijzen and Hauser, 2012; Jékely, 2013; Mirabeau and Joly, 2013; Mendoza et al., 2014). Many of these recent studies are enabled by the rise of genomic and transcriptomic data throughout the last decade (Elphick et al., 2018). Before the rise of molecular data, neuropeptide studies relied on peptidomic approaches, such as mass spectrometry or immunohistochemistry (e.g., van Veelen et al., 1994; Li et al., 1999; Predel, 2001). One major source of transcriptomes in insects is the 1KITE initiative (www.1kite.org), which has brought together leading experts of multiple fields of insect biology and aims to investigate the transcriptomes of more than 1,000 insect species, covering all recognized insect orders. The 1KITE consortium also included neuropeptide investigations, namely the analysis of neuropeptide evolution in non-pterygote hexapods (Derst, et al., 2016). This study showed that the majority of known insect neuropeptide precursors were already present in the last common ancestor of hexapods and the frequent absence of complete precursors that can be observed in some, mostly holometabolous, insect lineages (e.g., Chang et al., 2018; Nässel and Zandawala, 2019; Veenstra, 2019a) is an evolutionary trend that likely arose after the divergence of pterygote insects. Nevertheless, neuropeptide precursor sequences show evolutionary trends and variation shared within orders (Derst et al., 2016). The investigations carried out in my thesis are also mainly based on transcriptomes acquired from the 1KITE consortium. However, since these data were collected several years ago and were not done specifically to yield the best neuropeptide coverage possible, we also generated transcriptomic data with state of the art methods for species of special interest.

The main goal of my thesis was to investigate the evolution of neuropeptide precursor sequences on three levels: (i) the individual evolution of a selected neuropeptide family, (ii) a broad comparison of the evolution of multiple neuropeptide precursors and (iii) the evolution of the functionally different units of neuropeptide precursors and the phylogenetic information contained on these units and on complete neuropeptide precursors. The aim was to build on previous studies and deepen the understanding of neuropeptides and their evolution in insects, especially in the insect group Polyneoptera, using mostly transcriptomic data.

Polyneoptera belong to the Pterygota, i.e., winged insects, and represents the sister group to the remaining Neoptera (Beutel et al., 2014; Wipfler et al., 2019). Due to this phylogenetic position, Polyneoptera represent a promising model group for the study of the evolution of insect

neuropeptides and their precursors. Polyneoptera are composed of 10 orders of insects (Figure 2): Blattodea (cockroaches and termites), Dermaptera (earwigs), Embioptera (web-spinners), Grylloblattodea (ice crawlers), Mantodea (praying mantises), Mantophasmatodea (heel walkers), Orthoptera (crickets and relatives), Phasmatodea (stick insects and relatives), Plecoptera (stoneflies) and Zoraptera (angel insects). The internal phylogeny of these lineages has recently been revealed by a study of the Polyneoptera subgroup of the 1KITE consortium (Wipfler et al., 2019). A reliable phylogeny can therefore be used as a basis for evolutionary studies in this group (Figure 2).

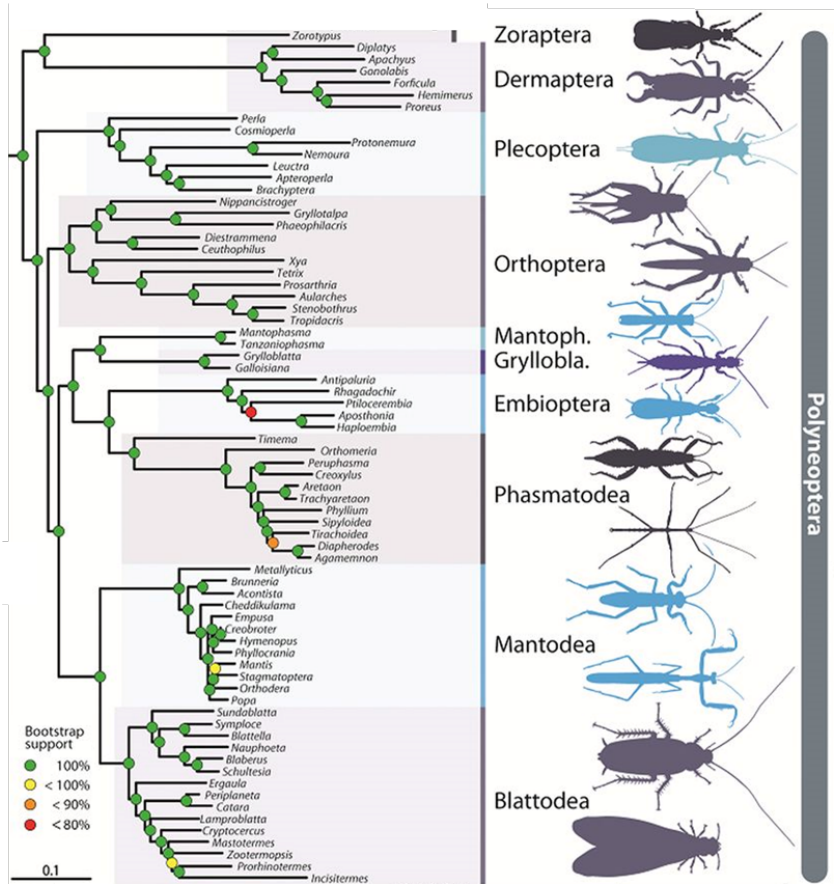



**Figure 2: Schematic overview of Polyneoptera and their internal phylogenetic relationships, modified after Wipfler et al (2019).**

Polyneopteran species have been used as model organisms in neuropeptide investigations for decades. In fact, the first neuropeptide sequence described from Insecta, proctolin, was found in the American cockroach, *Periplaneta americana* (Starratt and Brown, 1975). Furthermore, leucosulfakinin, found in 1986 in the cockroach *Leucophaea maderae* (Nachman et al., 1986),

was one of the first neuropeptides that was linked to a homolog peptidergic system in vertebrates, namely the vertebrate hormones human gastrin II and cholecystokinin. In the current age of transcriptomes and genomes, there is still a focus on polyneopteran species, resulting, among other findings, in the descriptions of novel neuropeptides and corresponding genes, e.g., HanSolin and RFLa in *Carausius morosus* (Phasmatodea) (Liessem et al., 2018), or the *tryptopyrokinin* genes in *Locusta migratoria* (Orthoptera) and *Zootermopsis nevadensis* (Blattodea) (Veenstra, 2014).

The newly described *tryptopyrokinin* genes (Veenstra, 2014) belong to the neuropeptide family of FXPRLamides/Pyrokinins and CAPA. Tryptopyrokinin neuropeptides share a WFGPRXa C-terminus and were hitherto only known to be products of *capa* and *pyrokinin* genes in insects (reviewed in Derst et al, 2016). No less than four genes coding for tryptopyrokinins were predicted from the genome of the well-studied desert locust *L. migratoria*, but so far, none of the products of these genes were experimentally verified. In addition to this, the closely related genes *pyrokinin* and *capa* were only partially known in *L. migratoria*. We created our own transcriptomes of the CNS of *L. migratoria* and used a combination of transcriptomic and peptidomic approaches to study the tryptoPKs in *L. migratoria*. With this combined approach, I was able to recover three complete tryptopyrokinin precursors and we identified products of two of these genes with single cell mass spectrometry (Redeker, Bläser et al., 2017). Furthermore, with the use of immunohistochemistry, we resolved the expression patterns of these two genes. With the new transcriptomes, I also completed the sequences of the *capa* and *pyrokinin* gene of *L. migratoria*, and described a further, novel neuropeptide gene, *pyrokinin-like*, whose products were also confirmed via mass spectrometry.

My first study combined peptidomics and transcriptomics to investigate a certain neuropeptide family of a single species in great detail and used both approaches to confirm each other. The huge amount of available transcriptomic data from sources like 1KITE and other projects (e.g., Liessem et al., 2018, Predel and Roth, unpublished) enabled us to increase the number of neuropeptides and species we investigated. This increase in quantity, however, meant that peptidomic analysis for all species was no longer a viable option. On the other hand it enabled us to investigate fundamental evolutionary trends such as absence or duplication of genes across a large dataset. Moreover, parts of the neuropeptide precursor that are not processed, such

as the signal peptide or cleavage sites, were included into the analyses. In my second publication, I analyzed 21 single-copy neuropeptide precursors of over 220 species, covering all polyneopteran orders and investigated the degree of sequence conservation between and within the different lineages (Bläser and Predel, 2020). In many cases, these data also enabled us to postulate the individual neuropeptide sequences that were present at the time of the insect emergence more than 400 million years ago. We were able to show that the degree of sequence conservation is relatively high throughout Polyneoptera, but synapomorphies for multiple individual lineages and orders could be identified. Furthermore, we showed that the reported frequent absence of complete precursors in some, mostly holometabolous insect lineages (e.g., Chang et al., 2018; Nässel and Zandawala, 2019), does not occur in Polyneoptera. This evolutionary trend therefore likely occurred after the divergence of Condylognatha and Holometabola.

Because we were able to recover lineage specific traits on the sequences of neuropeptides, these results also further indicated that neuropeptide datasets like these can be used for phylogenetic analyses. Neuropeptides underlie a strong stabilizing pressure as explained above. Thus, analyzing sequences of neuropeptides provides comprehensive data to elucidate phylogenetic relationships. Neuropeptide sequences can therefore be used as an alternative approach to resolve existing ambiguities in phylogenies. Taxon-specific sequences of expressed neuropeptides have already been used in the past to reconstruct phylogenetic relationships of closely related polyneopteran lineages (e.g., Blattodea: Roth et al., 2009, Mantophasmatodea: Predel et al. 2012). These previous phylogenetic studies based on neuropeptide sequences exclusively used the sequences of potential receptor-binding and thus bioactive neuropeptides. Most of these studies relied on extensive mass spectrometric results and therefore incorporated rather small datasets. Furthermore, the parts of the precursor that are not receptor-binding, like the sequences of the signal peptide and precursor peptides, were thus so far excluded from the analyses. Transcriptomic datasets not only allow the investigation of a larger taxon sampling and the inclusion of the complete precursor in the analyses but can also be used to gain novel insight into the sequences of previously excluded parts of neuropeptide precursors. In my third publication, we used the comprehensive dataset of neuropeptide precursors I had amassed to elucidate the phylogenetic relationships within Blattodea (Bläser et al., 2020). Blattodea were chosen due to the above average coverage of neuropeptide precursors in our dataset and the

presence of a recent phylogenetic analysis from the 1KITE initiative (Evangelista et al., 2019). The study by the 1KITE Blattodea subgroup is based mainly on the identical transcriptomes used in my study and a traditional nuclear phylogenomic approach with 2,370 protein-coding single-copy genes. This allowed a direct comparison between the two approaches. In total we combined 17 neuropeptide precursors from 40 blattodean species and were able to confirm the internal relationships proposed by Evangelista and colleagues (2019) with maximal node support. The third publication of my thesis therefore acts as a case study to demonstrate the power of neuropeptide precursor sequences for phylogenetic analyses. We also tested each functional unit of the precursor separately for their phylogenetic information. All units contained phylogenetic signal, although the degree of phylogenetic information varied. This study represents the first extensive utilization of complete neuropeptide precursor sequences for phylogenetic analyses. The results suggest that this alternative approach can be used as an alternative dataset to analyze ambiguous phylogenetic relationships in other groups.

The overall aim of my study was to investigate the evolution of neuropeptide sequences in insects, especially in Polyneoptera. This included the in-depth analysis of a selected novel neuropeptide precursor, tryptopyrokinin, and the evolution of its neuropeptide family in one species (*L. migratoria*) (Redeker, Bläser et al., 2017), as well as a broad evolutionary analysis of 21 single-copy neuropeptide precursors in over 200 species of Polyneoptera (Bläser and Predel, 2020). The latter study showcased lineage-specific evolutionary trends and enabled us to postulate the ancestral states of neuropeptide sequences that were present at the time of the insect emergence more than 400 million years ago. Furthermore, my study was designed to establish the usage of neuropeptide precursor sequences obtained from transcriptomes as a basis for phylogenetic analysis (Bläser et al., 2020). This marks the first time that a comprehensive set of complete precursors was used for phylogenetic analysis. Through a comparison of the functionally different parts of the neuropeptide precursor, insight on the phylogenetic information, and therefore on the evolutionary history of these parts, was gained.

# 2. Published Studies

## 2.1 Identification and Distribution of Products from Novel *tryptopyrokinin* Genes in the Locust, *Locusta migratoria*

Jana Redeker[1], Marcel Bläser[1], Susanne Neupert, Reinhard Predel

Published in Biochemical and Biophysical Research Communications 486 (2017) pp 70-75. 

Supplementary Material is provided as Supplementary Material SM 1.

Author contributions

MB, SN and RP conceived and designed the experiments; JR prepared material for transcriptome sequencing; MB analyzed the transcriptome data; JR performed single cell mass spectrometry experiments; JR, MB, SN and RP analyzed the data; JR, MB and SN prepared figures and tables; JR, MB and RP wrote the manuscript.

[1] Authors contributed equally.

Biochemical and Biophysical Research Communications 486 (2017) 70–75

Contents lists available at ScienceDirect

Biochemical and Biophysical Research Communications

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

# Identification and distribution of products from novel *tryptopyrokinin* genes in the locust, *Locusta migratoria*

Jana Redeker [1], Marcel Bläser [1], Susanne Neupert, Reinhard Predel*

*Institute for Zoology, Functional Peptidomics Group, University of Cologne, D-50674 Cologne, Zuelpicher Str. 47b, Germany*

ARTICLE INFO

*Article history:*
Received 15 February 2017
Accepted 27 February 2017
Available online 28 February 2017

*Keywords:*
Insect
Locust
Neuropeptide
Pyrokinin
CAPA
Peptidomics

ABSTRACT

A recent analysis of the genome of *Locusta migratoria* indicated the presence of four novel insect neuropeptide genes encoding for multiple tryptopyrokinin peptides (tryptoPKs); hitherto only known from *pyrokinin* or *capa* genes. In our study, mature products of *tryptoPK* genes 1 and 2 were identified by mass spectrometry; precursor sequences assigned to the *tryptoPK* genes 3 and 4 are likely partial sequences of a single precursor. The expression of *tryptoPK* genes 1 and 2 is restricted to two cells in the subesophageal ganglion, exhibiting not only a unique neuropeptidome but also a very distinctive axonal projection. Comparative neuroendocrinology revealed that homologous cells in other insects also produce tryptoPKs but use other genes to generate this pattern. Since *capa* and *pyrokinin* genes are discussed as ancestors of the *tryptoPK* genes, we completed the hitherto only partially known precursor sequences of these genes by means of transcriptome analyses. The distribution of mature products of CAPA and pyrokinin precursors in the CNS is compared with that of tryptoPKs. In addition, a novel pyrokinin-like precursor is described.



## 1. Introduction

Most of the neuropeptide genes in insects and crustaceans have largely similar ancestors in arthropods [e.g. 1, 2]. There is sufficient evidence to suggest that orthologous neuropeptide genes are even known from insects and mammals which supports a long evolutionary history and conservation of neuropeptide genes [3]. On the other hand, many lineage-specific developments of neuropeptide genes are known in metazoa and these developments can usually be traced back to gene duplications. Representatives for the development of lineage-specific neuropeptide genes in insects are *pyrokinin* (*pk*) and *capa* genes. Current knowledge about these genes and their mature products suggests an ancient condition in arthropods of a single gene encoding periviscerokinins (PVKs) and pyrokinins (PKs) [4]. At the origin of hexapods, the ancestor of *pk* and *capa* genes rather suddenly encodes a third peptide species, differentiated from conventional PKs by the presence of Trp preceding the C-terminal core motif of PKs; namely FXPRLamide [5]. This is important since it is known from *Drosophila melanogaster*, that PVKs, PKs and tryptoPK each activate specific receptors [6–8]. It can only be speculated that the presence of three different receptor ligands in a single precursor posed a challenge for a ligand-specific release. It is also not known yet, how long in the evolution of hexapods the trypto-PK receptor evolved as functionally separated unit. Clearly however, the appearance of tryptoPKs was followed in Hexapoda by a gene duplication which led to *pk* (*pban* of Lepidoptera, *hugin* of *D. melanogaster*) and *capa* genes [see 4]. Following gene duplication, both genes evolved differentially. In today's insects, the *pk* gene is expressed in neurosecretory cells of the subesophageal ganglion (SEG) and the *capa* gene is mainly expressed in neurosecretory cells of abdominal ganglia. Although there are several exemptions known from insects, most *pk* genes encode for PKs and a single tryptoPK and the *capa* genes encode for PVKs and also a single tryptoPK. A separate release of CAPA-tryptoPK is known from many insects and likely the result of differential processing of the CAPA precursor [9–11]. On the other hand, no data exist regarding a differential processing of PK precursors in insects which likely results in a consequent co-release of PKs and tryptoPKs from the *pk* expressing cells. A recent analysis of the genome of the migratory locust *Locusta migratoria* indicated an

*Abbreviations:* NCA-2, nervus corporis allati-2; NCC-3, nervus corporis cardiaci-3; PSO, perisympathetic organ; PK, pyrokinin; PVK, periviscerokinin; RC, retrocerebral complex; tryptoPK, tryptopyrokinin; VPMN, ventral posterior median neuron.

* Corresponding author.
*E-mail address:* rpredel@uni-koeln.de (R. Predel).
[1] Authors contributed equally.

http://dx.doi.org/10.1016/j.bbrc.2017.02.135
0006-291X/

alternative option for a separate release of tryptoPKs that was not known from insects so far. In fact, not less than 4 genes coding for multiple tryptoPKs were described [12]. However, not a single tryptoPK which can be deduced from these genes is known from the well-studied locusts [13].

Here, we describe the identification of mature products from two *tryptoPK* genes in *L. migratoria* as well as their cellular origin and axonal projection. In addition, only partially known precursor sequences of *pk* and *capa* genes have been completed and their distribution in the CNS analyzed. Finally and for the first time in insects, a second PK precursor and the neuropeptides processed from this precursor have been identified.

## 2. Materials and methods

### 2.1. Transcriptome sequencing, library construction and BLAST search

Total RNA was extracted from the CNS of single adult locusts (*L. migratoria*) kept at 4 °C during the preparation. CNS samples were preserved in RNAlater solution (QIAGEN GmbH, Hilden, Germany) and sequenced using the Illumina Truseq RNA Sample Preparation Kit (Illumina, San Diego, USA) at Beijing Genomics Institute (China). Isolation of mRNA, generation of cDNA and PCR amplification of suitable fragments were performed as described in Ref. [4]. Raw data were initially filtered by removing adapters, reads with more than 5% of unknown basses and reads with low quality sequences (reads having more than 20% bases with quality value lower or equal to 10). Transcripts were then *de novo* assembled using Trinity [14] with default options and used for tBLASTn on a local computer.

### 2.2. Tissue preparation and cell dissection for MALDI-TOF mass spectrometry

*Corpora cardiaca*, *corpora allata*, perisympathetic organs and nerves were dissected as described for cockroach samples by Predel [15] and directly transferred into a drop of water on the sample plate for MALDI-TOF mass spectrometry. Immediately after the transfer, the water was removed and the tissue samples were dried prior to deposition of matrix. Preparations of neuroendocrine cells or cell cluster from isolated SEGs were performed as described in Predel et al. [16].

### 2.3. MALDI TOF mass spectrometry

10 mg/ml 2,5−dihydroxybenzoic acid (Sigma-Aldrich, Steinheim, Germany) dissolved in 20% acetonitrile/1% formic acid or alternatively 10 mg/ml α-cyano-4-hydroxycinnamic acid dissolved in 60% ethanol, 36% acetonitrile, 4% water (Sigma-Aldrich, Steinheim, Germany) were used as matrix salts and loaded onto the dried samples using a glass capillary (10–20 nl for single cell preparations) or 0.1–10 μl pipettes (about 0.3 μl for direct tissue profiling). Mass fingerprint spectra were acquired in positive ion mode with an UltrafleXtreme TOF/TOF mass spectrometer (Bruker Daltonik, Bremen, Germany). MSMS fragment spectra were acquired in gas-off mode using an ABI 4800 proteomics analyzer (Applied Biosystems, Framingham, USA). Due to the limited nature of all samples, acquisitions were generally taken in manual mode.

### 2.4. Immunocytochemistry

An antiserum raised against the CAPA peptide *Periplaneta* PVK-2 [17], which recognizes PKs, tryptoPKs and PVKs, was used for whole-mount immunostainings. Tissue samples were fixed with Histofix (Sigma-Aldrich) at 4 °C for 12 h, then washed with phosphate-buffered saline (ph 7.2) containing 1% TritonX-100 (1% Tx) for 24 h and incubated for 5 days at 4 °C in a polyclonal anti-PVK-2 serum at a concentration of 1:4000 diluted in PBS-1% Tx containing 2,5 mg/ml bovine serum and 10% normal goat serum. Subsequently, the samples were washed in PBS 1% Tx for 24 h at 4 °C followed by incubation with Cy3-coupled goat anti-rabbit antiserum at a concentration of 1:3000 (Jackson ImmunoResearch) for 5 days at 4 °C. After washing in PBS-1% Tx for 24 h at 4 °C, the samples were embedded in Mowiol (Merck KGaA, Darmstadt, Germany). A confocal laser scanning microscope (Zeiss LSM 510 Meta system; Jena, Germany), equipped with a C-Apochromat 10×/0.45W objective, a Plan-Apochromat 20x/0.75 objective and a Helium-Neon1 laser was used to scan serial optical sections at optical section thicknesses from 0.3 to 0.8 μm. Images were exported and processed to adjust brightness and contrast using Adobe Photoshop 7.0 software (San Jose, CA).

## 3. Results

As a first step, transcriptome analysis of the CNS was performed to complete sequences of PK and CAPA precursors, which were only partially known from genome data [12], and to confirm the sequences of the four tryptoPK precursors. The resulting sequences of tryptoPK precursors 1 and 2 (Supplementary Material 1) show a surprisingly high number of amino acid substitutions when compared with the respective precursors predicted from the locust genome (e. g. 25 in the tryptoPK 2 precursor). Although different populations of *L. migratoria* were used in both studies, such variability is without precedent even for neuropeptide precursors of closely related insect species [18]. Transcriptome data indicated that the published tryptoPK precursors 3 and 4 [12] contain in fact partial sequences of a single precursor; designated as tryptoPK 3 precursor here. It can also be concluded from the transcriptome data, that most of the already known PKs from *L. migratoria* [see 13] are processed from the PK precursor (Supplementary Material 1), except for pQSVPTFTPRL-$NH_2$ [Lom-PK-2; 19]. A blast search in the transcriptome assembly of the CNS, using the sequence of this peptide, revealed a second PK precursor containing two PKs only. Thus, *L. migratoria* not only has *tryptoPK* genes, which were hitherto unknown from insects, but also an unusual second *pk* gene which shows low similarity with the other *pk* gene, except for the C-terminus of the PKs. To avoid confusion in the assignment of mature peptides, we name the novel gene *pk-like* gene (*pkl*). Sequences of putative neuropeptides which can be predicted from PK, PKL, CAPA and tryptoPK precursors are listed in Table 1 insofar as mature products from these precursors were found by mass spectrometry. With the complete precursor sequences in hand, these peptides were renamed according to their position in the different precursors. The resulting nomenclature is used throughout this manuscript. Theoretical masses ($[M+H]^+$) of peptides predicted from the different precursors provided the basis to search for tryptoPKs and related PKs, PKLs, and PVKs in the CNS and neurohemal organs by means of mass spectrometric direct tissue profiling.

It was already known from cockroaches that, except for a few *capa*-interneurons in the brain/SEG, all *capa* and *pk* expressing cells in the CNS store their products in neurohemal organs [16]. Therefore, we searched for neuropeptides deduced from the three tryptoPK precursors first in major neurohemal organs of *L. migratoria*. As expected [19,20], mass spectra of abdominal perisympathetic organs (PSOs) verified the presence of CAPA peptides including hitherto unknown PVKs and CAPA-tryptoPKs (Table 1) but products of the *tryptoPK* genes were not detected. Mass spectra of thoracic PSOs did not contain any ion signals typical of PKs, PKLs, PVKs or

**Table 1**
List of mature peptides, which have been confirmed by mass spectrometry, from *tryptoPK*, *capa* and *pk/pkl* genes. Peptides additionally verified by MSMS experiments are marked with an asterisk; amino acid substitutions due to allelic differences are underlined. Designations of already known peptides are changed according to their position in the respective precursors. Note, that mass spectra suggest an N-terminus of PK-6 and a signal peptide of the PK precursor different from those predicted (see Supplementary Material 1) and also revealed N-terminally different PK-4 sequences with similar abundances. AG, abdominal ganglion; aPSO, abdominal perisympathetic organ; LCN, lateral cardiac nerve; MdC, neurosecretory cells of the mandibular neuromere in the subesophageal ganglion; MxC, neurosecretory cells of the maxillary neuromere in the SEG; RC, retrocerebral complex; VPMN, ventral posterior median neuron.

| Peptide name | Historic designation | Peptide sequence | *m/z*, $[M+H]^+$ | MS | Source |
|---|---|---|---|---|---|
| *TryptoPK precursor 1* | | | | | |
| tPK1-1 | – | GSTLSENTGVWFGPRYa | 1769.86 | + | VPMN, RC, LCN |
| tPK1-2 | – | AAKQPALWFGPRVa | 1439.83 | +* | ” |
| tPK1-3 | – | SAQPPGLWFGPRVa | 1410.76 | +* | ” |
| tPK1-4 | – | SDAQVDDMLWFGPRPa | 1732.81 | +* | ” |
| tPK1-5 | – | GAKHPGLWFGPRFa | 1468.80 | +* | ” |
| *TryptoPK precursor 2* | | | | | |
| tPK2-1 | – | SIPEPGTLFGPRIa | 1382.78 | + | VPMN, RC, LCN |
| tPK2-2 | – | SHTEAGVWFGPRYa | 1505.73 | +* | ” |
| tPK2-3 | – | SYPEPGMWFGPRVa | 1521.73 | +* | ” |
| tPK2-4 | – | SQAEPGVWFGARIa | 1416.74 | + | ” |
| tPK2-5 | – | NQPEPGTWFGARIa | 1471.74 | +* | ” |
| tPK2-6 | – | SHPEPGMWFGPRVa | 1495.73 | +* | ” |
| tPK2-7 | – | SHPEPGTWFGPRIa | 1479.75 | +* | ” |
| *PK precursor* | | | | | |
| PK-1 | PK-4 | EGDFTPRLa | 933.50 | +* | MxC, MdC, RC |
| ext. PK-1 | – | HGGGGSWVSRREGDFTPRLa | 2070.04 | + | ” |
| PK-2 | (PK-9,10) | ESAEQGGGVSAWQGGEPQQEEQVLAGPFVPRLa | 3336.62 | +* | ” |
| PK-3 | PK-3 | GAVPAAQFSPRLa | 1212.68 | +* | ” |
| PK-4 | PK-5 | RQQPFVPRLa | 1139.68 | +* | ” |
| PK-$4^{2-9}$ (pQ) | PK-8 | pQQPFVPRLa | 966.55 | +* | ” |
| PK-5 | PK-1 | DSGDEWPQQPFVPRLa | 1769.86 | +* | ” |
| PK-6 | PK-6 | RLHQNGMPFSPRLa | 1551.83 | +* | ” |
| PK-$6^{2-13}$ | PK-7 | LHQNGMPFSPRLa | 1395.73 | +* | ” |
| *PK-like precursor* | | | | | |
| PKL-1 | PK-2 | pQSVPTFTPRLa | 1127.62 | +* | MxC, MdC, RC |
| PKL-2 | – | DSAGDELAEEDAVDDGGDGLPQPQLAPPFWPRPa | 3460.59 | + | ” |
| *CAPA precursor* | | | | | |
| PVK-1 (pQ) | – | pQDGDKGISKLKKTSSLFPHPRIa | 2434.36 | +* | AG cells, aPSO |
| PVK-$1^{12-21}$ | PKL-1 | TSSLFPHPRIa | 1153.65 | +* | ” |
| PVK-2 | PVK-1 | AAGLFQFPRVa | 1104.63 | +* | ” |
| PVK-3 | – | KGLVASARVa | 899.58 | +* | ” |
| PVK-4 | PVK-2 | GLLAFPRVa | 871.55 | +* | ” |
| CAPA tPK-1a | PVK-3 | DGGEPAAPLWFGPRVa | 1567.80 | +* | ” |
| CAPA tPK-1b | – | DGGQPAAPLWFGPRVa | 1566.82 | +* | ” |
| CAPA tPK-2a | – | SGSDGDRDRDRQGLWFGPRVa | 2275.11 | + | ” |
| CAPA tPK-2b | – | SGSDGDRDRDRQGLWFGRRVa | 2334.16 | + | ” |

tryptoPKs. Mass spectra from the retrocerebral complex (RC with *corpora cardiaca* and *corpora allata*), however, revealed ion signals mass identical with theoretical masses of peptides from tryptoPK precursors 1 and 2. These ion signals appeared with much lower ion intensity than ion signals typical of products from the *pk* gene (Fig. 1).

Ventral median neurosecretory cells in the SEG are a major source of PKs stored in the RC of insects. As shown by immunostainings, this is also the case in *L. migratoria* (Fig. 2). Most of these cells (mandibular and maxillary cell cluster) have projections into the *nervi corporis allati-2* (NCA-2) which link the SEG with the RC [21]. Indeed, ion signals of putative peptides from tryptoPK precursors 1 and 2 were also found in mass spectra from distal portions of the NCA-2 (not shown). However, these ion signals disappeared in mass spectra from NCA-2 preparations in the vicinity of the SEG. In accordance with that, ion signals of products from the *pk* gene and, with lower ion intensity, ion signals of products from the *pkl* gene were detected in mass spectra of the ventral median cell cluster of the mandibular and maxillary neuromers of the SEG (Fig. 3A). Products of the *tryptoPK* genes could not be observed in these mass spectra. Instead, potential products of tryptoPK precursors 1 and 2 were found in preparations of the *nervi corporis cardiaci-3* (NCC-3; not shown); these nerves intermingle with the NCA-2 near the RC. Immunostainings suggest ((Fig. 2) and [22]) that the cellular origin of tryptoPKs, which were detected in the NCC-3, are PVK-immunoreactive ventral posterior median neurons (VPMNs) in the labial neuromere of the SEG. In fact, mass spectra of VPMNs verified the presence of all predicted tryptoPKs from the *tryptoPK* genes 1 and 2 (Fig. 3B). The abundance of tryptoPKs in these cells allowed MSMS analyses to confirm the identity of most of the tryptoPKs. Although our transcriptome data obtained from preparations of SEG of *L. migratoria* suggested the presence of three tryptoPK precursors, not a single ion signal typical of products from tryptoPK precursor 3 was observed. Possibly, the absence in mass spectra of peptides from the *tryptoPK* 3 gene resulted from a much lower quantity of mature peptides from this gene. Products of *pk* and *pkl* genes were not detectable in mass spectra from VPMNs.

Projections of VPMNs of *L. migratoria* have been thoroughly investigated by Bräunig [22] who described, in addition to axons running anteriorly into the NCC-3, also projections running posteriad into the lateral cardiac nerves via abdominal ganglia and segmental nerves. The presence of products of the tryptoPK

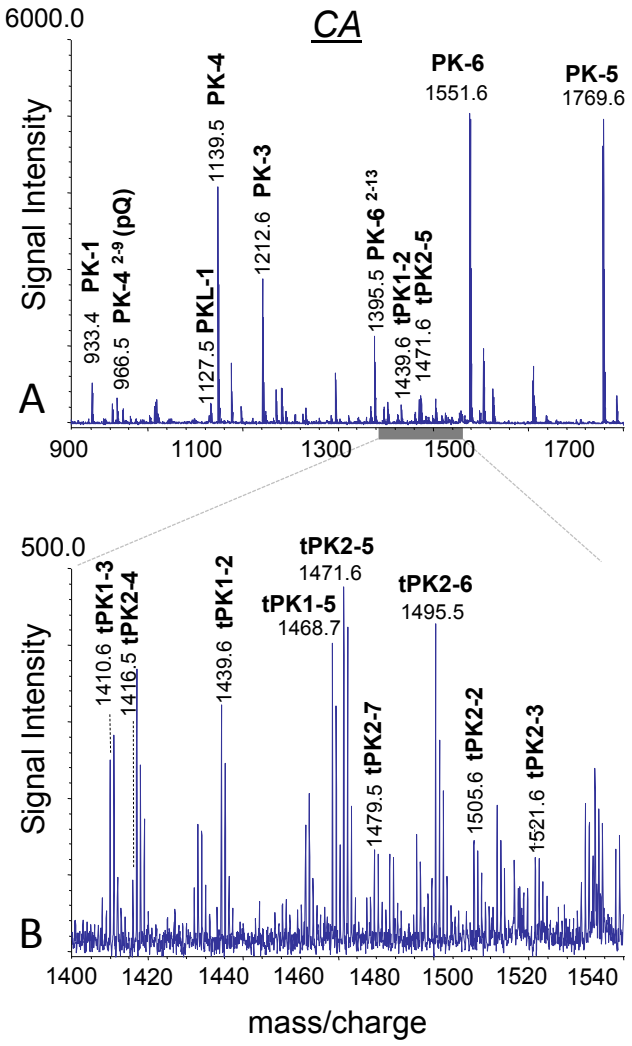



**Fig. 1.** Direct tissue profiling of a single corpus allatum (CA) indicating the presence of products from PK, PKL and tryptoPK precursors 1 and 2. **A)** Mass range of *m/z* 900–1800; only PKs and tryptoPKs are labelled. Peptides from the PK precursor show higher signal intensity than those of the PKL and tryptoPK precursors. **B)** Mass range of *m/z* 1400–1550 with a large number of ion signals typical of tryptoPKs (tPKs).

precursors 1 and 2 in the lateral cardiac nerves was confirmed with mass spectrometric direct tissue profiling (Fig. 4). Hence, the two VPMNs in the SEG have not only a unique peptidome but also a very distinct axonal projection. A prominent ion signal at 1389.8 appearing in mass spectra of VPMNs as well as lateral cardiac nerves indicates a co-processing with tryptoPKs of a still unknown peptide in VPMNs rather than a contamination during the dissection of VPMNs. Altogether, no hints for the occurrence of tryptoPKs in tissues other than those crossed by VPMNs were found in our analyses which also included nerves and different parts of the CNS (not shown).

## 4. Discussion

We could demonstrate for the first time the processing of products of two *tryptoPK* genes in *L. migratoria*. Since our transcriptome data obtained from CNS samples suggest the expression of three *tryptoPK* genes in the SEG, the failure to detect mature products from *tryptoPK* gene 3 possibly reflects a much lower abundance of these peptides. This assumption is corroborated by the fact that the identity of nearly all predicted tryptoPKs of *tryptoPK* genes 1 and 2 could be unambiguously confirmed by fragment analyses. We also succeeded in the detection of the expression site of *tryptoPK* genes 1 and 2 in the SEG. Due to the sequence similarity of products of *tryptoPK* genes with other PKs and also the CAPA-tryptoPKs, available antisera raised against PKs of different insects cannot distinguish between these peptides.

The current data suggest that locusts have a unique set of PKs in the broader sense, encoded in three *tryptoPK* genes, two *pk* (*pk* and *pkl*) genes and the *capa* gene. Given the fact that these genes likely evolved from a single ancestor gene in arthropods, this is an exceptional example of neuropeptide evolution. There are some hints, however, that *tryptoPK* genes are more common in insects, at least in Polyneoptera. Veenstra [12] described a single putative *tryptoPK* gene from the termite *Zootermopsis nevadensis* which precursor sequence has a moderate similarity with sequences of *tryptoPK* genes from *L. migratoria*. In addition, a comparison of the data obtained here with mass spectra obtained from cockroach neurons that are homologous to VPMNs in locusts [16] indicates that *Periplaneta* PK-6 and 7 could be processed from a specific *tryptoPK* gene as well. When we change the point of view and focus only on the VPMNs and homologous SEG cells in other insect lineages, we see an interesting development. In insect as different as cockroaches and fruit flies, these two cells seem to have an expression pattern and axonal projection which is typical of these cells only [see also 10]. Whether *pk* or *capa* genes are expressed in these cells, the mature products are usually restricted to tryptoPKs; possibly a result of differential processing. Only in the moth *Manduca sexta*, these cells express *capa* and *pk* genes and process all peptides which can be predicted from both precursors. This pattern likely represents the ancient condition in insects. In locusts these cells do not express *pk* or *capa* anymore; the *pk* genes also miss the tryptoPKs typical of *pk* genes in many other insects. Instead, the novel *tryptoPK* genes are expressed in these cells. As a consequence, differential processing to release solely tryptoPKs is not necessary

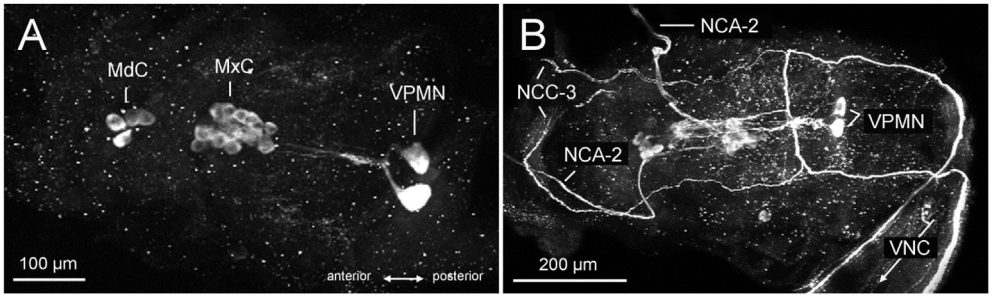



**Fig. 2.** Anti-Pea PVK-2 like immunofluorescence staining (whole mount) of the SEG of *L. migratoria*, with immunostained neurosecretory cell cluster in the mandibular (MdC), maxillary (MxC) and labial (VPMN) neuromeres of the SEG. These cells project into NCA-2, NCC-3 and posteriad into the ventral nerve cord (VNC).

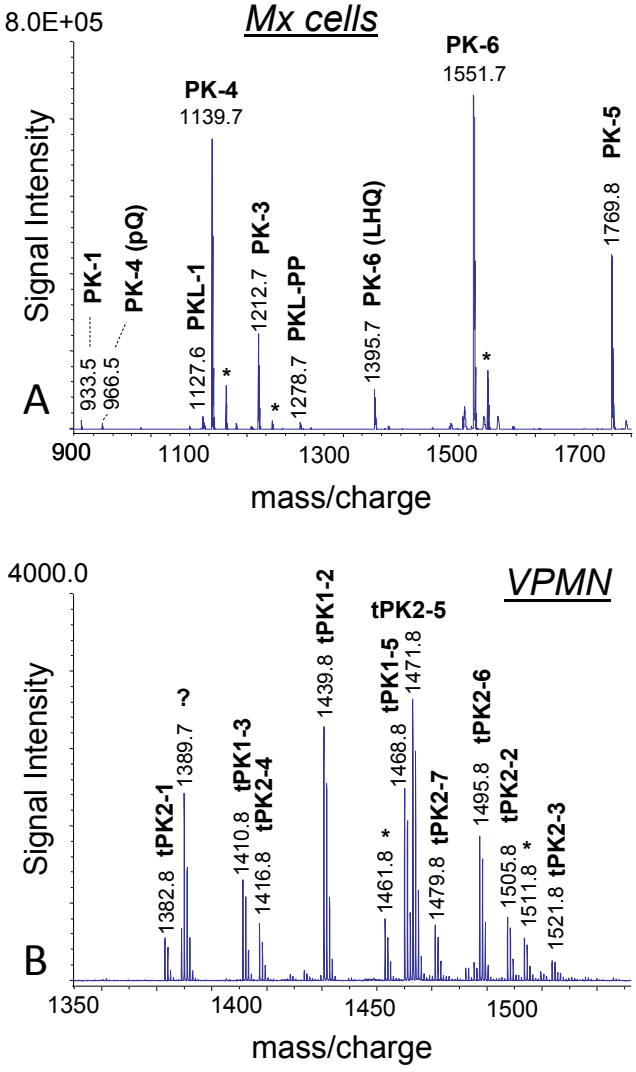



**Fig. 3.** Direct tissue profiling of **A)** cells of the *pk* expressing cell cluster in the maxillary neuromere of the SEG (MxC), indicating the presence of PKs from *pk* and *pkl* genes (*m/z* 900–1800) and **B)** a VPMN from the labial neuromere in the SEG (*m/z*1350-1550). The mass spectrum indicates, by mass matches, the presence of all predicted tryptoPKs from the tryptoPK precursors 1 and 2. Note the completely different set of peptides in both spectra. These cells could not be distinguished by immunocytochemistry (see Fig. 2). The identity of most tryptoPKs was confirmed by subsequent MSMS analyses (see Table 1). Asterisks mark sodium adduct ions or Met-oxidized peptides. tPK 1-1, tryptopyrokinin-1 of tryptoPK precursor 1 etc.; PP, precursor peptide.

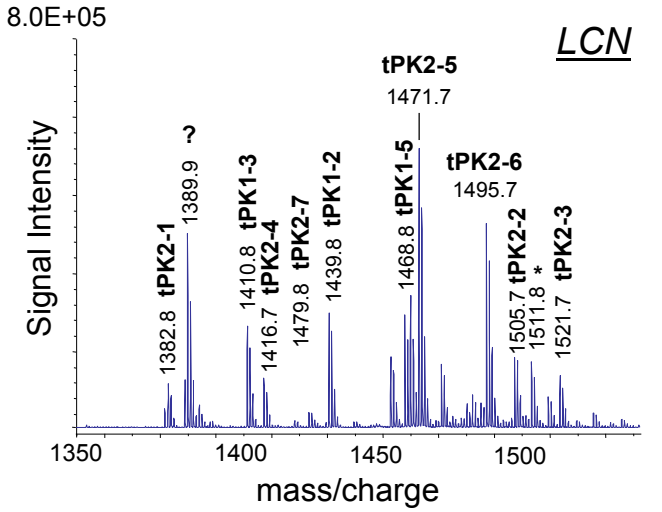



**Fig. 4.** Direct tissue profiling of lateral cardiac nerve (LCN) indicating the presence of products from tryptoPK precursors 1 and 2 (*m/z* 1350–1550). Asterisks mark sodium adduct ions or Met-oxidized peptides. tPK 1-1, tryptopyrokinin-1 of tryptoPK precursor 1.

anymore.

As a general conclusion, the ancestor of *pk/capa* genes underwent fast evolutionary changes with the emergence of hexapods, including multiple gene duplications, receptor ligand sorting and differential expression of duplicated genes. Some of these developments, notably the release of CAPA-PVKs from abdominal PSOs, are typical of insects and related with newly acquired functions such as the regulation of Malpighian tubule activity. Other developments are typical of specific insect lineages only; the occurrence of *tryptoPK* genes in *L. migratoria* seems to be a feature which exists exclusively in Polyneoptera. Only two cells with a distinct projection pattern in the CNS were found to process mature peptides of *tryptoPK* genes 1 and 2. In other insects, these neurons in the labial neuromere of the SEG either express products of the *pk* gene, products of the *capa* gene or products of both of these genes. The processing of the respective genes is usually reduced to *tryptoPKs* in these cells and yields a characteristic peptidome. In locusts, the novel *tryptoPK* genes replace the expression of *pk/capa* genes. As discussed here for insects, sequence-related neuropeptides are always produced in these cells and it is obvious that the respective neurons recruit different genes in different insects to conserve the same function. A general function in insects, however, is still unknown for tryptoPKs.

## Acknowledgements

This project was financially supported by the Deutsche Forschungsgemeinschaft (PR766/11-1). The authors would like to thank Dr. Lapo Ragionieri (Cologne, Germany) for assembling transcriptome raw data and supporting transcriptome analysis and Prof. P. Bräunig (Aachen, Germany) for providing adults of *L. migratoria*.

## Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.bbrc.2017.02.135.

## Transparency document

Transparency document related to this article can be found online at http://dx.doi.org/10.1016/j.bbrc.2017.02.135.

## References

[1] J.A. Veenstra, S. Rombauts, M. Grbić, In silico cloning of genes encoding neuropeptides, neurohormones and their putative G-protein coupled receptors in a spider mite, Insect Biochem. Molec. Biol. 42 (2012) 277–295.
[2] G. Mayer, L. Hering, J.M. Stosch, et al., Evolution of pigment-dispersing factor neuropeptides in panarthropoda: insights from onychophora (velvet worms) and tardigrada (water bears), J. Comp. Neurol. 523 (2015) 1865–1885.
[3] O. Mirabeau, J. Jean-Stéphane, Molecular evolution of peptidergic signaling systems in bilaterians, Proc. Natl. Acad. Sci. U. S. A. 110 (2013) 2028–2037.
[4] C. Derst, H. Dircksen, K. Meusemann, et al., Evolution of neuropeptides in nonpterygote hexapods, BMC Evol. Biol. 16 (2016) 51.
[5] R. Predel, R.J. Nachman, The FXPRLamide (pyrokinin/PBAN) peptide family, in: A. Kastin (Ed.), Handbook of Biologically Active Peptides, Academic Press, Elsevier, Louisiana, 2006, pp. 207–212.
[6] G. Cazzamali, M. Torp, F. Hauser, et al., The Drosophila gene CG9918 codes for a pyrokinin-1 receptor, Biochem. Biophys. Res. Commun. 335 (2005) 14–19.

[7] Y. Park, Y.J. Kim, M.E. Adams, 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 (2002) 11423–11428.
[8] A. Iversen, G. Cazzamali, M. Williamson, et al., Molecular cloning and functional expression of a Drosophila receptor for the neuropeptides capa-1 and-2, Biochem. Biophys. Res. Commun. 299 (2002) 628–633.
[9] R. Predel, C. Wegener, Biology of the CAPA peptides in insects, Cell. Mol. Life Sci. S63 (2006) 2477–2490.
[10] S. Neupert, W. Huetteroth, J. Schachtner, R. Predel, Conservation of the function counts: homologous neurons express sequence-related neuropeptides that originate from different genes, J. Neurochem. 111 (2009) 757–765.
[11] S. Neupert, W.K. Russell, D.H. Russell, R.Predel, Two capa-genes are expressed in the neuroendocrine system of *Rhodnius prolixus*, Peptides 31 (2010) 408–411.
[12] J.A. Veenstra, The contribution of the genomes of a termite and a locust to our understanding of insect neuropeptides and neurohormones, Front. Physiol. 5 (2014) 454.
[13] E. Clynen, L.Schoofs, Peptidomic survey of the locust neuroendocrine system, Insect Biochem. Molec. Biol. 39 (2009) 491–507.
[14] M.G. Grabherr, B.J. Haas, M. Yassour, et al., Full-length transcriptome assembly from RNA-seq data without a reference genome, Nat. Biotechnol. 29 (2011) 644–652.
[15] R. Predel, Peptidergic neurohemal system of an insect: mass spectrometric morphology, J. Comp. Neurol. 436 (2001) 363–375.
[16] R. Predel, M. Eckert, E. Pollák, et al., Peptidomics of identified neurons demonstrates a highly differentiated expression pattern of FXPRLamides in the neuroendocrine system of an insect, J. Comp. Neurol. 500 (2007) 498–512.
[17] R. Predel, J. Rapus, M. Eckert, et al., Isolation of periviscerokinin-2 from the abdominal perisympathetic organs of the American cockroach, Peripl. Am. Peptides 19 (1998) 801–809.
[18] R. Predel, S. Neupert, W. Huetteroth, et al., Peptidomics-based phylogeny and biogeography of mantophasmatodea (Hexapoda), Syst. Biol. 61 (2012) 609–629.
[19] E. Clynen, J. Huybrechts, A. De Loof, L. Schoofs, Mass spectrometric analysis of the perisympathetic organs in locusts: identification of novel periviscerokinins, Biochem. Biophys. Res. Commun. 300 (2003) 422–428.
[20] R. Predel, G. Gäde, Identification of the abundant neuropeptide from abdominal perisympathetic organs of locusts, Peptides 23 (2002) 621–627.
[21] P. Bräunig, C. Böhme, B. Staufer, Morphology of locust neurosecretory cells projecting into the nervus corporis allati II of the suboesophageal ganglion, Microsc. Res. Tech. 35 (1996) 230–241.
[22] P. Bräunig, A suboesophageal ganglion cell innervates heart and retrocerebral glandular complex in the locust, J. Exp. Biol. 156 (1991) 567–582.

## 2.2 Evolution of Neuropeptide Precursors in Polyneoptera (Insecta)

Marcel Bläser, Reinhard Predel

Published in Frontiers in Endocrinology 11:197 (2020). 

Supplementary Material is provided as Supplementary Material SM 2.

Author contributions

MB and RP conceived and designed the experiments; MB mined the transcriptome dataset for neuropeptide precursor sequences and performed statistical analysis; MB and RP analyzed the data; MB prepared figures and tables and wrote the first draft of the manuscript; MB and RP wrote the paper.

frontiers in Endocrinology

ORIGINAL RESEARCH
published: 15 April 2020
doi: 10.3389/fendo.2020.00197

# Evolution of Neuropeptide Precursors in Polyneoptera (Insecta)

*Marcel Bläser and Reinhard Predel**

*Department of Biology, Institute for Zoology, University of Cologne, Cologne, Germany*

Neuropeptides are among the structurally most diverse signaling molecules and participate in intercellular information transfer from neurotransmission to intrinsic or extrinsic neuromodulation. Many of the peptidergic systems have a very ancient origin that can be traced back to the early evolution of the Metazoa. In recent years, new insights into the evolution of these peptidergic systems resulted from the increasing availability of genome and transcriptome data which facilitated the investigation of the complete neuropeptide precursor sequences. Here we used a comprehensive transcriptome dataset of about 200 species from the 1KITE initiative to study the evolution of single-copy neuropeptide precursors in Polyneoptera. This group comprises well-known orders such as cockroaches, termites, locusts, and stick insects. Due to their phylogenetic position within the insects and the large number of old lineages, these insects are ideal candidates for studying the evolution of insect neuropeptides and their precursors. Our analyses include the orthologs of 21 single-copy neuropeptide precursors, namely ACP, allatotropin, AST-CC, AST-CCC, CCAP, CCHamide-1 and 2, CNMamide, corazonin, CRF-DH, CT-DH, elevenin, HanSolin, NPF-1 and 2, MS, proctolin, RFLamide, SIFamide, sNPF, and trissin. Based on the sequences obtained, the degree of sequence conservation between and within the different polyneopteran lineages is discussed. Furthermore, the data are used to postulate the individual neuropeptide sequences that were present at the time of the insect emergence more than 400 million years ago. The data confirm that the extent of sequence conservation across Polyneoptera is remarkably different between the different neuropeptides. Furthermore, the average evolutionary distance for the single-copy neuropeptides differs significantly between the polyneopteran orders. Nonetheless, the single-copy neuropeptide precursors of the Polyneoptera show a relatively high degree of sequence conservation. Basic features of these precursors in this very heterogeneous insect group are explained here in detail for the first time.

**Keywords: neuropeptides, transcriptome, Polyneoptera, insect evolution, Blattodea, Orthoptera, Phasmatodea, Dermaptera**

OPEN ACCESS

***Edited by:***
*Elizabeth Amy Williams, University of Exeter, United Kingdom*

***Reviewed by:***
*Sheila Ons, National University of La Plata, Argentina*
*Sven Bradler, University of Göttingen, Germany*

****Correspondence:***
*Reinhard Predel*
*rpredel@uni-koeln.de*

***Specialty section:***
*This article was submitted to Neuroendocrine Science, a section of the journal Frontiers in Endocrinology*

***Received:*** *30 January 2020*
***Accepted:*** *19 March 2020*
***Published:*** *15 April 2020*

***Citation:***
*Bläser M and Predel R (2020) Evolution of Neuropeptide Precursors in Polyneoptera (Insecta). Front. Endocrinol. 11:197. doi: 10.3389/fendo.2020.00197*

## INTRODUCTION

Neuropeptides are among the structurally most diverse signaling molecules in multi-cellular animal organisms. As such, they participate in intercellular information transfer from neurotransmission to intrinsic or extrinsic neuromodulation and regulate physiological processes including growth, reproduction, development, and behavior. For a single insect species, up to 50

neuropeptide genes can be expected coding for single or multiple copies of neuropeptides (1, 2). The sequences of single-copy neuropeptides, which are the focus of our study, are on average better conserved than those of multiple-copy peptides because amino acid (AA) substitutions in the single ligand of a particular neuropeptide receptor are potentially more likely to lead to a general loss of function than substitutions involving only one of several related neuropeptides produced from the same precursor. Thus, mutations that alter the sequences of single-copy neuropeptides must either be accompanied by parallel mutations in receptor genes that maintain the binding properties of the respective receptors or should not alter the steric properties of the peptides to maintain functionality (3). Most neuropeptides activate peptide-specific G-protein coupled receptors and many of these peptidergic systems have a very ancient origin that can be traced back to the early evolution of the Metazoa [e.g., (4)]. In fact, in many cases orthologies between neuropeptide and/or corresponding receptor genes of distantly related lineages can be identified (5–7). The identification of cockroach sulfakinins (8) with an already then suspected relationship to the cholecystokinins of vertebrates was a first strong indication of the conservation of peptidergic systems across protostomes and deuterostomes, these taxa diverged more than 700 million years ago (9, 10). Mirabeau and Joly (6) later described eight conserved peptidergic systems (vasopressin, neuropeptide Y/F, tachykinin, gonadotropin-releasing hormone/adipokinetic hormone, cholecystokinin/sulfakinin, neuromedin U/pyrokinin, corticotropin-releasing factor, calcitonin) which probably already occurred in the last common ancestor of Bilateria. More recently, Elphick et al. (4) have described as much as 30 neuropeptide signaling systems with orthologs in protostomes and deuterostomes.

Insects have always been in the focus of neuropeptide research (11). Today, the fruitfly *Drosophila melanogaster* is the model organism also for the study of neuropeptide functions (12). However, due to the large number of harmful insects that have an impact on human health or reduce yields in agriculture and forestry, other groups of insects, in particular true bugs, beetles, lepidopterans, and various groups of flies have also been intensively examined [e.g., (13–18)]. Beneficial insects such as honey bees and predators or parasites of pest insects [e.g., (19–21)] were also investigated in detail. Nevertheless, many neuropeptides in insects were first described from Polyneoptera such as locusts, cockroaches and stick insects; including proctolin, the first ever identified insect neuropeptide (22). In recent years, most of the insights into the evolution of peptidergic systems then resulted from the increasing availability of genome and transcriptome data which facilitated the investigation of the complete neuropeptide precursor sequences. Such data have been used in several comprehensive studies, mainly to compile neuropeptide precursor sequences within higher taxa (23–25). In a recent study a large dataset of neuropeptide precursors from Blattodea was used to demonstrate the considerable phylogenetic information contained in these sequences (3). A particular focus on the general evolution of neuropeptide precursors has been placed in a study on the precursors of 12 *Drosophila* species (26). In this study the authors also discussed the potentially different evolution of single-copy and multiple-copy precursors, with mutations in the neuropeptide sequences of single-copy precursors being exposed to stronger stabilizing selection.

In our study, we used a comprehensive transcriptome dataset of about 200 species from the 1KITE initiative (http://www.1kite.org/) to study the evolution of single-copy neuropeptide precursors in Polyneoptera. This group comprises well-known orders such as cockroaches and termites (Blattodea), locusts (Orthoptera) and stick insects (Phasmatodea), but also rather unknown orders with few species such as ice crawlers (Grylloblattodea), heel walkers (Mantophasmatodea), and angel insects (Zoraptera). The internal relationships of Polyneoptera were recently resolved as part of the 1KITE project (27). Due to their phylogenetic position within the Insecta (**Figure 1**) and the large number of old lineages, these insects are ideal candidates for studying the evolution of insect neuropeptides and their precursors. Furthermore, the taxon sampling of the 1KITE initiative includes multiple species from all higher lineages of Polyneoptera (with the exception of Zoraptera with only one species). This enabled us to discuss changes in single-copy neuropeptide precursor sequences with respect to insect evolution.

Our analyses include the orthologs of 21 single-copy neuropeptide precursors. The neuropeptides of several of these precursors were first described from Polyneoptera (cockroaches: proctolin, corazonin, myosuppressin; stick insects: HanSolin, RFLamide), but were later also found in other insects. Based on the sequences obtained, the degree of sequence conservation between and within the different polyneopteran lineages is discussed. Furthermore, the data are used to postulate the individual neuropeptide sequences that were present at the time of the insect emergence more than 400 million years ago, as well as to detect taxon-specific losses of peptidergic systems. The dataset from the 1KITE initiative has become "historic" and contains a number of transcriptomes with partly insufficient coverage of neuropeptide precursors. The few new transcriptomes we have prepared specifically for this study (Mantophasmatodea) show much better data coverage. Nevertheless, the comprehensive taxon sampling of the 1KITE initiative allows a sufficient validation of almost all statements we provide about the evolution of neuropeptide precursors in Polyneoptera.

## MATERIALS AND METHODS

### Orthology Assessment and Alignment of Neuropeptide Precursor Sequences

Orthology assessment and alignments were performed as described in Bläser et al. (3). Briefly, we mined transcriptome sequences, provided by the 1KITE initiative (GenBank Umbrella BioProject ID PRJNA183205), for single-copy neuropeptide precursors in the datasets from each order of Polyneoptera; starting with neuropeptide precursor sequences of *Carausius morosus* (2), *Locusta migratoria* (1) and Blattodea (3). Once a full set of single-copy neuropeptide precursors was obtained, we used this information to search for precursors in the remaining

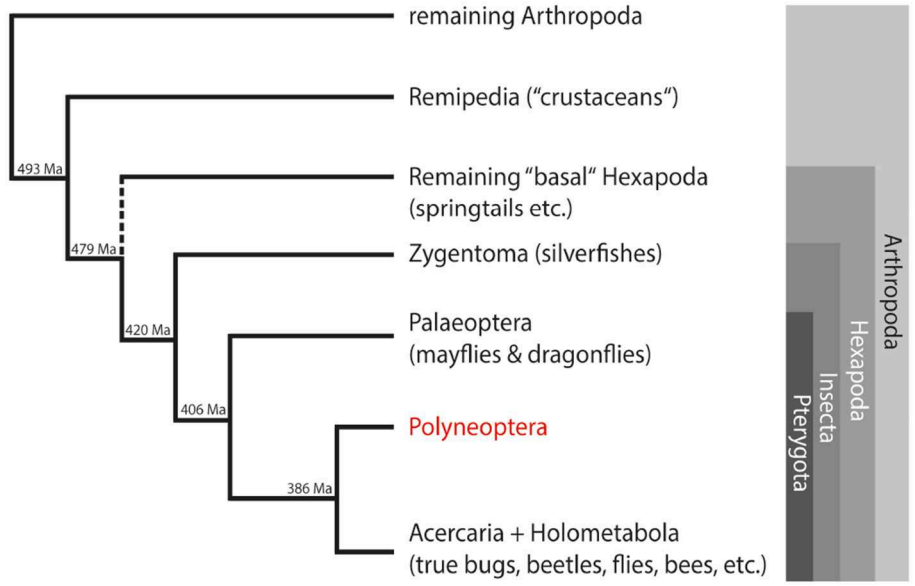



**FIGURE 1 |** Simplified overview of the phylogenetic relationships among insects and their closest relatives within the arthropods. Estimated divergence times of the different lineages are added at the nodes (28). Ma, million years ago.

species of this order or species from related orders. Assembled transcripts were analyzed with the tblastn algorithms provided by NCBI (https://blast.ncbi.nlm.nih.gov/Blast.cgi). Identified candidate nucleotide precursor gene sequences were translated into AA sequences using the ExPASy Translate tool (29) with the standard genetic code. Orthologous neuropeptide precursors were aligned using the MAFFT-L-INS-i algorithm (30) (dvtditr (aa) Version 7.299b alg=A, model=BLOSUM62, 1.53, −0.00, −0.00, noshift, amax=0.0). Alignments generated with the MAFFT-L-INS-i algorithm were then manually checked for misaligned sequences using N-termini of signal peptides and conserved AA residues (cleavage signals, Cys as target for disulphide bridges) as anchor points. Incompletely translated transcripts of neuropeptide precursors and transcripts of questionable quality were either combined to generate complete precursors when possible, or labeled with question marks at the respective AA positions.

## Assessment of Precursor Characteristics

Individual AA alignments of each group of orthologous neuropeptide precursors from each order were merged in BioEdit 7.2.5 (31). The coverage of single-copy neuropeptide precursors in our dataset (**Additional File 1**), minimal and maximal length of precursors as well as number of identified transcripts and the position of the conserved neuropeptide sequences in the precursor were manually determined for each neuropeptide precursor in each order of Polyneoptera, respectively. Additionally, the predicted neuropeptide sequences in the individual AA alignments were determined and further analyzed using BioEdit 7.2.5. The lengths of these sequences as well as N-terminal and C-terminal cleavage sites were manually determined.

Alignments of single-copy neuropeptide precursors of the individual polyneopteran orders as well as combined alignments of all polyneopteran lineages were used to estimate the average evolutionary divergence (AED) over all sequence pairs in Mega X (32). Standard error estimates were obtained by implementing 500 bootstrap replicates using the Poisson correction model (33). We used the pairwise deletion option to remove all ambiguous sites. The results of these analyses are shown in **Additional File 3**. The median AED of each single-copy neuropeptide precursor was calculated with Microsoft Excel and compared to the overall AED value of the respective single-copy neuropeptide precursor. To calculate the overall AED value, complete sequences of each single-copy neuropeptide from each polyneopteran order (see **Additional File 1**) were merged into a single file and aligned again. Furthermore, the alignments of the predicted neuropeptide sequences were used to calculate the overall AED value for the conserved neuropeptide sequences for each neuropeptide. Finally the overall median AED of all single-copy neuropeptides for each order was calculated using Microsoft Excel.

These analyses enabled us to compare internal sequence variation for each neuropeptide in all polyneopteran orders (AED for each order), as well as between orders (overall AED). The median AED also allowed an assessment of relative levels of sequence conservation. The overall AED of the predicted neuropeptide sequences enables a comparison of sequence conservation between neuropeptide sequences and the complete precursor sequence.

## Sequence Logo Generation and Topology Mapping

Sequence logos of the aligned neuropeptide precursor orthologs were generated using the tool WebLogo version 2.8.2 (34). Each stack represents one position in the multiple sequence alignment. The overall height of a stack indicates the sequence conservation at this AA position; the height of letters within

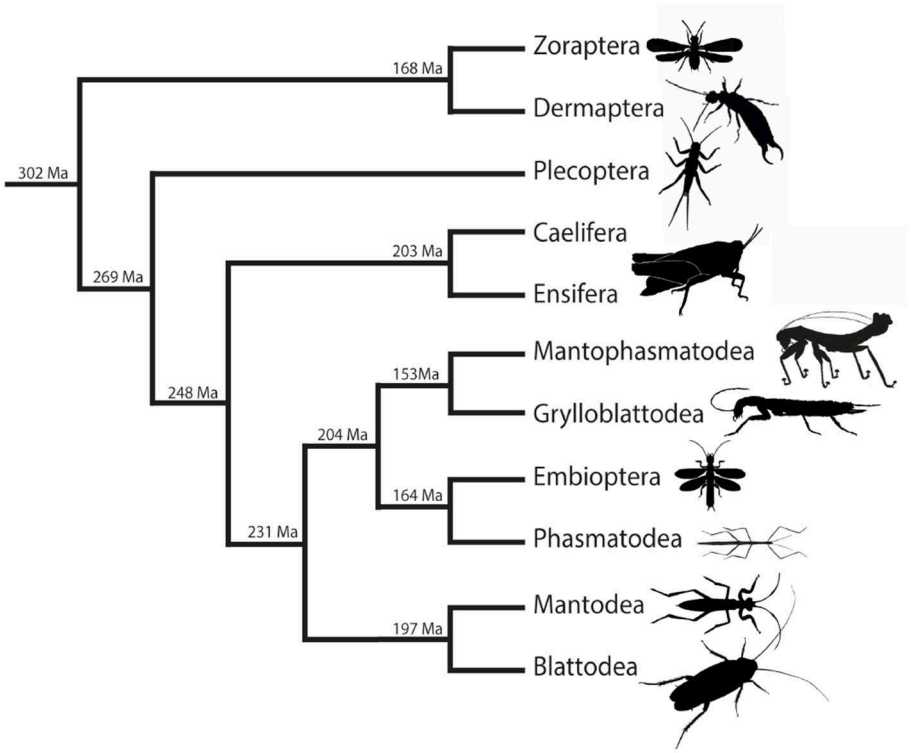



**FIGURE 2 |** Simplified overview of the phylogenetic relationships among Polyneoptera [modified from Wipfler et al. (27)]. Estimated divergence times of the different lineages are added at the nodes [see Misof et al. (28)].

the stack indicates the relative frequency of each AA at that position. For the color scheme of AA residues, the default settings were selected. Resulting sequence logos were manually mapped (Adobe Illustrator CS6; version 16.0.0.) on a simplified tree showing the phylogeny of Polyneoptera (27).

## RESULTS AND DISCUSSION

The BLAST searches in the Polyneoptera transcriptome assemblies of the 1KITE initiative were performed with single-copy neuropeptide precursor sequences of *C. morosus* (2), *L. migratoria* (1), and Blattodea (3). Due to the varying quality of the transcriptome data and the generally low quantity of several neuropeptide-coding RNA sequences in whole body transcriptomes (see **Additional File 1**), the number of identified precursor sequences is significantly lower than the number of species analyzed. In addition, the yield of neuropeptide precursors is different for the different precursors; for example, much fewer CCHamide-1 precursors could be identified across the different lineages than precursors for other neuropeptides such as proctolin and NPF-1 (see **Additional File 1**). Nevertheless, the extensive material of the 1KITE initiative guaranteed sufficient information for almost all orders of Polyneoptera. The only exception was Zoraptera, where only a single transcriptome is available. In total, we have included 21 different single-copy precursors in our analysis. The precursors for adipokinetic hormones (AKHs) were not included in our study because the number of AKH genes varies considerably between and within the polyneopteran orders and the orthologies could not be resolved. All neuropeptide precursor sequences identified in this study are listed in **Additional File 2**, sorted by the different polyneopteran lineages. The phylogenetic relationships in Polyneoptera are illustrated in **Figure 2** which also shows the estimated divergence times of the different orders. The estimated divergence times between orders vary between 300 and 150 Ma, which corresponds to the time scale of the parallel (independent) evolution of the respective precursors. Due to the long separate history of the two orthopteran lineages Caelifera and Ensifera (**Figure 2**), we have treated them separately in our analyses. The information about the neuropeptide sequences for each lineage is used in the following to determine the respective ancestral neuropeptide sequences for the Polyneoptera. A comparison with orthocopies of Zygentoma and Remipedia (24) allows in many cases also a statement about the possible ancestral neuropeptide sequences of the insects or even of the hexapods.

### ACP (Additional File 3A)

A single adipokinetic hormone/ corazonin-like peptide (ACP) precursor with a length of 85–109 AA is present in almost all polyneopteran orders. The only exception was found in Dermaptera, where the ACP precursor is absent. It is noteworthy that of the 37 phasmatodean species analyzed, an ACP precursor was found only in 5 species of Oriophasmata (35). However, the transcriptomes of Phasmatodea in general show a rather high percentage of missing data (see **Additional File 1**) and therefore the low number of ACP precursors in Phasmatodea might also

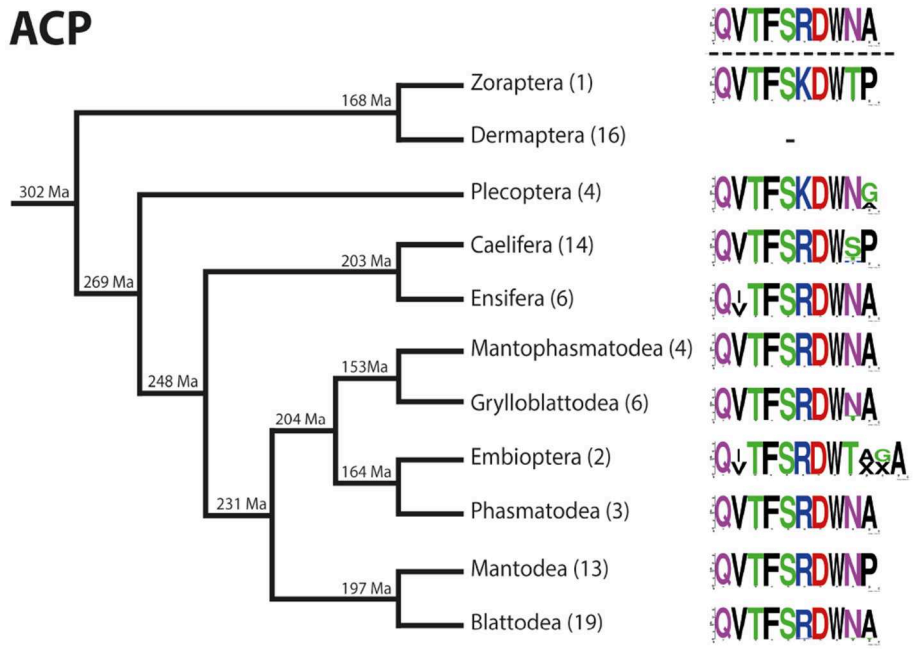



**FIGURE 3 |** Sequence logo representation showing the degree of amino acid sequence conservation of the ACP neuropeptides for each order; mapped on a phylogenetic tree of Polyneoptera [modified from Wipfler et al. (27)]. Only the completely obtained precursor sequences were considered, the respective number is given in parentheses for each taxon. An "X" in the sequence represents a gap. The hypothetical ancestral state of the ACP sequence in Polyneoptera is listed at the top.

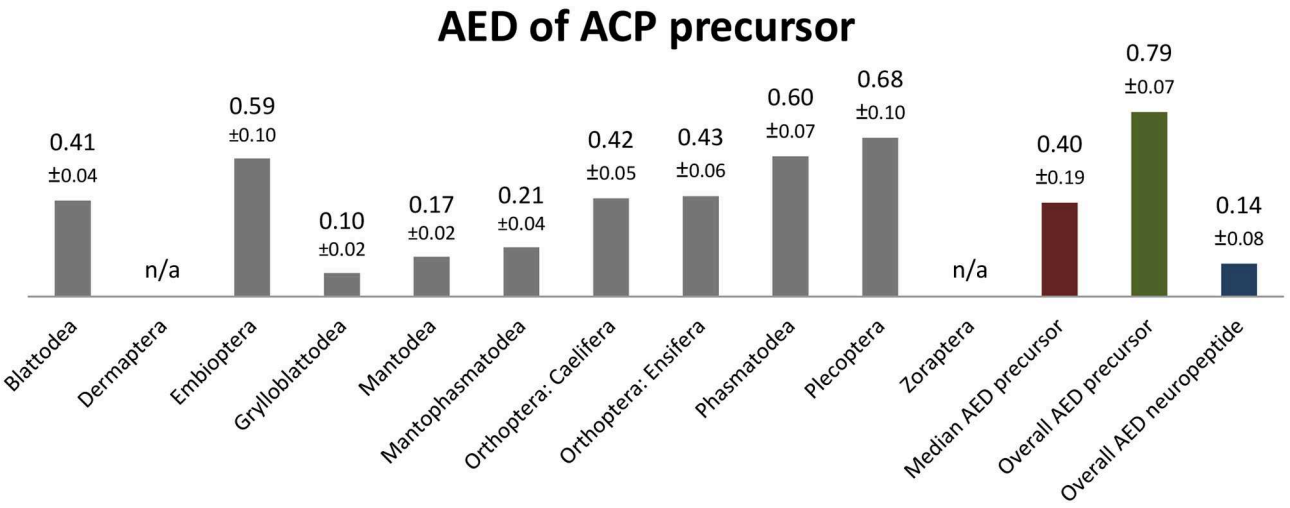



**FIGURE 4 |** Bar chart depicting the average evolutionary divergence (AED) of the ACP precursors for the different polyneopteran lineages with standard error estimates (S.E.). In addition, the median AED and the AED values for the available set of complete ACP precursors and predicted ACP sequences of Polyneoptera are given.

be a result of this incomplete dataset. Each precursor contains a usually very well conserved ACP motif with an amidation site. The ACP sequence immediately follows the signal peptide and terminates upstream of a dibasic KR cleavage site.

The ACP sequences are mostly decapeptides; only in one species of Embioptera (*Rhagadochir virgo*) ACP is a duodecapeptide. The sequence (p)QVTFSRDWNA-$NH_2$, which also occurs in the remipedian *X. tulumensis*, was found in various orders of Polyneoptera (**Figure 3**). This sequence is likely ancestral for all Hexapoda. Amino acid substitutions in ACPs of Polyneoptera are largely limited to substitutions from $Val^2$ to $Ile^2$ (few Ensifera and Embioptera), $Arg^6$ to $Lys^6$ (Zoraptera, all Plecoptera), and several substitutions of the two C-terminal AA (**Figure 3**). The median average evolutionary divergence (AED) for the ACP precursor is 0.40 (**Figure 4**). Grylloblattodea show the lowest ACP precursor variation, while Plecoptera

possess the most variable ACP precursor sequences. The overall AED for the ACP precursors is 0.79, the actual neuropeptide sequence is significantly better conserved (overall AED: 0.14; **Figure 4**).

## AST-CC (Additional File 3B)

A single allatostatin-CC (AST-CC) precursor with a length of 108-172 AA is present in all polyneopteran orders. The precursors contain a well conserved AST-CC motif without a C-terminal amidation site. Only in *Apteroperla tikumana* (Plecoptera) a second partial precursor with an alternative AST-CC motif was found. The AST-CC sequence is always located C-terminal in the precursor, N-terminally flanked by a dibasic RR cleavage site [monobasic Arg in *Tryonicus parvus* (Blattodea)] and terminates upstream of variable C-terminal cleavage sites or the AST-CC precursor sequence ends directly with the AST-CC sequence.

Most AST-CC sequences of Polyneoptera are non-adecapeptides, only some derived AST-CC sequences of Orthoptera have 18 (few Caelifera) or 20 AA [few Ensifera and *T. parvus* (Blattodea)]. The sequence GQQKGRVYWRCYFNAVTCF-OH which also occurs in the silverfish *T. domestica* was found in most orders of Polyneoptera (not in Dermaptera, Caelifera, Embioptera). This sequence might therefore be regarded as ancestral for all Pterygota. While the C-terminal motif YWRCYFNAVTCF-OH is highly conserved in all Polyneoptera, the N-terminus shows a number of lineage-specific AA substitutions, particularly at position 2 and 3. The median AED for the AST-CC precursor is 0.24. Grylloblattodea and Mantophasmatodea show the lowest AST-CC precursor variation, while Embioptera possess the most variable AST-CC precursors. The overall AED for the AST-CC precursor is 0.49, the actual neuropeptide sequence is significantly better conserved (overall AED: 0.08).

## AST-CCC (Additional File 3C)

A single allatostatin-CCC (AST-CCC) precursor with a length of 93-121 AA is present in all polyneopteran orders. The precursors contain a highly conserved AST-CCC motif with an amidation site. The AST-CCC sequence is always located C-terminal in the precursor, N-terminally flanked by a dibasic KR cleavage site and terminates upstream of a monobasic K cleavage site (KK in few species of Caelifera).

All AST-CCC sequences of Polyneoptera are tetradecapeptides. The sequence SYWKQCAFNAVSCF-$NH_2$ which also occurs in the remipedian *X. tulumensis* was found in all orders of Polyneoptera. This sequence might therefore be regarded as ancestral for all Hexapoda. Amino acid substitutions in AST-CCC are limited to a substitution from $Lys^4$ to $Arg^4$ in few species of Dermaptera and Embioptera and all species of Mantodea. The median AED for the AST-CCC precursor is 0.23. Mantophasmatodea show the lowest AST-CCC precursor variation, while Ensifera possess the most variable AST-CCC precursors. The overall AED for the AST-CCC precursor is 0.36, the actual neuropeptide sequence is significantly better conserved (overall AED: 0.03).

In our study we could not find an AST-C precursor in any transcriptome. The presence of an AST-C precursor was previously suggested for *L. migratoria* (36, 37), but the corresponding AST-C neuropeptide has not been confirmed biochemically (e.g., fragment analysis).

## AT (Additional File 3D)

A single allatotropin (AT) precursor with a length of 102–142 AA is present in all polyneopteran orders. The precursor contains a usually well-conserved AT motif with a C-terminal amidation site. The AT sequence follows a precursor peptide (17-21 AA) inserted between the signal peptide and the AT sequence. Only in the Dermaptera the AT sequence is directly C-terminal of the signal peptide. In all other taxa, the AT sequence is N-terminally flanked by a monobasic Arg cleavage site, while all sequences terminate upstream of a dibasic KR cleavage site. Specific features of AT precursors were found in Embioptera, where the species *Haploembia palaui* has a second AT precursor and another species (*Ptilocerembia catherinae*) has a second, longer AT motif immediately C-terminal of the first AT sequence.

With a single exception (Ensifera: *Comicus calcaris*; 12 AA), the AT sequences (AT-1 of *P. catherinae*) are tridecapeptides. The sequence GFKNVALSTARGF-$NH_2$ which also occurs in the silverfish *T. domestica* was found in many orders of Polyneoptera (not in Zoraptera, Dermaptera, Mantophasmatodea, Grylloblattodea, Embioptera). This sequence might therefore be regarded as ancestral for all Pterygota. Common AA substitutions of AT sequences affect the positions 5 and/or 6 from the N-terminus, resulting in lineage-specific AA at these positions. Highly derived sequences of AT are typical of all Dermaptera and most Embioptera; in Dermaptera these substitutions even affect the N- and C-terminal AA, which are conserved in all other polyneopteran orders. The median AED for the ACP precursor is 0.3. Grylloblattodea and Mantophasmatodea show the lowest ACP precursor variation in Polyneoptera, while Embioptera possess the most variable ACP precursors. The overall AED for the ACP precursors is 0.48, the actual neuropeptide sequence is significantly better conserved (overall AED: 0.10).

## CCAP (Additional File 3E)

A single crustacean cardioactive peptide (CCAP) precursor with a length of 143-174 AA is present in all polyneopteran orders. The precursor contains a fully conserved CCAP motif with a C-terminal amidation site. The CCAP sequence follows a precursor peptide (24–27 AA) inserted between the signal peptide and the CCAP sequence. The CCAP sequences are N-terminally flanked by dibasic KR cleavage sites and terminate upstream of KKR or RKR (few Orthoptera and Embioptera) cleavage sites. The sequence PFCNAFTGC-$NH_2$ which also occurs in the remipedian *X. tulumensis* was found in all orders of Polyneoptera. This sequence might therefore be regarded as ancestral for all Hexapoda. A single AA substitution from $Phe^6$ to $Leu^6$ was found in *Creoxylus spinosus* (Phasmatodea). The median AED for the CCAP precursor is 0.22. Grylloblattodea and Mantophasmatodea show the lowest CCAP precursor variation in Polyneoptera, while Plecoptera possess the most variable CCAP precursors.

The overall AED for the ACP precursors is 0.45, the actual neuropeptide sequence is significantly better conserved (overall AED: 0.00).

## CCHamide-1 (Additional File 3F)

A single CCHamide-1 precursor with a length of 115-241 AA is present in almost all polyneopteran orders. The only exception was found in Ensifera, where the CCHamide-1 precursor is absent. The precursor contains a usually well-conserved CCHamide-1 motif with a C-terminal amidation site. The CCHamide-1 sequence in the precursor immediately follows the signal peptide and terminates upstream of a dibasic KR cleavage site.

With very few exceptions (N-terminally extended in Embioptera and possibly also in Caelifera), the predicted CCHamide-1 sequences are always tetradecapeptides. The sequence GSCLSYGHSCWGAH-$NH_2$ which also occurs in the silverfish *T. domestica* was found in most orders of Polyneoptera (not in Zoraptera). This sequence might therefore be regarded as ancestral for all Pterygota. Significant intraordinal variation is only present in Dermaptera, Plecoptera, Blattodea, and Mantodea. The most common AA substitution across different orders was that of $Ala^{13}$ to $Gly^{13}$ (Zoraptera and few Mantophasmatodea, Mantodea and Blattodea). The median AED for the CCHamide-1 precursor is 0.52. Mantophasmatodea show the lowest CCHamide-1 precursor variation, while Plecoptera have the most variable CCHamide-1 precursors. The overall AED for the CCHamide-1 precursors is 0.81, the actual neuropeptide sequence is significantly better conserved (overall AED: 0.07).

## CCHamide-2 (Additional File 3G)

A single CCHamide-2 precursor with a length of 97-174 AA is present in all polyneopteran orders. The precursors contain a well-conserved CCHamide-2 motif with a C-terminal amidation site. The CCHamide-2 sequence in the precursor follows immediately the signal peptide and terminates upstream of a dibasic KR cleavage site. An analysis of the *C. morosus* peptidome (2) has shown that the N-terminal KR of CCHamide-2 is not recognized as a cleavage site and we hypothetically assume that the same N-terminus occurs in all polyneopteran CCHamide-2. Two precursors with different CCHamide-2 sequences were found in the species *H. palaui* (Embioptera). Additionally, in several species of Caelifera a second transcript of the CCHamide-2 precursor was found. In these species, 16 AA (P/SYGVRR/TPGD/AIQI/TRRAG) are inserted following the N-terminal KR in the respective CCHamide-2 sequences.

With few exceptions in Blattodea (*Nocticola* sp.: 16 AA; *Catara rugosicollis*, *Coptotermes* sp.: 14 AA), the predicted CCHamide-2 sequences are always pentadecapeptides. The sequence KRGCSAFGHSCFGGH-$NH_2$ which also occurs in several silverfish species, but not *T. domestica*, was found in most orders of Polyneoptera (not in Plecoptera, Dermaptera). This sequence might therefore be regarded as ancestral for all Pterygota. Common AA substitutions in the CCHamide-2 sequences are $Ala^4$ to $Ser^4$ (Phasmatodea, several Blattodea, few Plecoptera) and $Phe^{10}$ to $Tyr^{10}$ (few Embioptera, Plecoptera, and Dermaptera). The AA at position 3 of the N-terminus show lineage-specific AA substitutions in the majority of Dermaptera (Ser to Gln), Plecoptera (Ser to Thr) and Caelifera (Ser to Met). The median AED for the CCHamide-2 precursor is 0.37. Mantophasmatodea show the lowest CCHamide-2 precursor variation in Polyneoptera, while Dermaptera possess the most variable CCHamide-2 precursors. The overall AED for the CCHamide-2 precursors is 0.72, the actual neuropeptide sequence is significantly better conserved (overall AED: 0.10).

## CNMamide (Additional File 3H)

CNMamide precursors with a length of 127–179 AA are present in all polyneopteran orders. Two transcripts with different CNMamide sequences were found in several Blattodea. The less commonly found longer transcripts (7 species of Blattodea) are orthologs of the single CNMamide precursor of Mantodea (sister group of Blattodea). In Caelifera, we identified two precursors in the species *Haplotropis brunneriana* and *Pielomastax soochowensis*. One of these precursors is very similar to the single CNMamide precursors of other Polyneoptera, while the second precursor, found in the majority of caeliferan species (27), has a rather variable and N-terminally extended sequence. In Ensifera, we have also identified two very different precursor sequences, but these sequences were always found in different species. It therefore remains unclear whether these sequences represent different transcripts or result from the rapid sequence diversification of CNMamide precursors. All precursors contain the CNMamide motif with a C-terminal amidation site. The CNMamide sequence in the precursor is located C-terminal in the precursor, N-terminally flanked by a dibasic KR cleavage site (KK in the ensiferan *Acheta domesticus* and *Phaeophila crisbredoides*), and terminates upstream of variable C-terminal cleavage sites (mostly RKR).

Most CNMamides are tetradecapeptides, but the full variation is from 13 to 18 AA. The sequence GSYMSLCHFKICNM-$NH_2$ which also occurs in the silverfish *T. domestica* was found in many orders of Polyneoptera (not in Dermaptera, Caelifera, Mantophasmatodea, Embioptera, Mantodea). This sequence might therefore be regarded as ancestral for all Pterygota. Common AA substitutions in the CNMamide sequences are $Gly^1$ to $Asn^1$ (all Mantophasmatodea), $Gly^1$ to $Thr^1$ (all Embioptera), and $Ser^5$-$Leu^6$ to $Thr^5$-$Met^6$ (all Mantodea). Particularly the CNMamide sequences of Caelifera, Blattodea, and Ensifera in which two different precursors are present are quite variable at the N-terminus. A similar sequence variation was also found in Dermaptera and Plecoptera. The median AED for the CCHamide-2 precursor is 0.44. Mantophasmatodea show the lowest CCHamide-2 precursor variation in Polyneoptera, while Plecoptera possess the most variable CCHamide-2 precursors. The overall AED for the CCHamide-2 precursors is 0.95, the actual neuropeptide sequence is significantly better conserved (overall AED: 0.26).

## Corazonin (Additional File 3I)

A corazonin precursor with a length of 85-140 AA was found in almost all polyneopteran orders. The only exception was obtained in Zoraptera, where the corazonin precursor is absent. For two species, *Nippancistroger testaceus* (Ensifera) and *Medauroidea*

*extradentata* (Phasmatodea), we identified a second precursor with different corazonin sequences. Otherwise, the corazonin precursors contain highly conserved corazonin motifs with C-terminal amidation sites. The corazonin sequence in the precursor follows immediately the signal peptide and terminates upstream of a RKR cleavage site.

Corazonin sequences are almost exclusively undecapeptides. Only in one species of Dermaptera (*Nesogaster amoenus*), corazonin has 9 AA and the C-terminal dipeptide is missing. The sequence (p)QTFQYSRGWTN-$NH_2$, which also occurs in Malacostraca and even Myriapoda, was found in most orders of Polyneoptera (not in Phasmatodea and Dermaptera). This sequence might therefore be regarded as ancestral for all Hexapoda. Peptidomics confirmed for different polyneopteran taxa that the N-terminal Gln of corazonin is almost completely converted to pGlu (38). Amino acid substitutions in corazonin of Polyneoptera are largely limited to substitutions from $Arg^7$ to $His^7$ (many Phasmatodea, Caelifera and Dermaptera). Considering the phylogenetic position of the respective insect taxa (**Figure 1**), the $His^7$-corazonins probably evolved several times independently of each other. Significant intraordinal variation is only present in Dermaptera. In Mantophasmatodea, all species of a single lineage (Austrophasmatidae) have a unique corazonin sequence with two AA substitutions ($Gln^4$ to $His^4$; $Arg^7$ to $Gln^7$), while in all other Mantophasmatodea the original sequence is retained (39). The median AED for the corazonin precursor is 0.37. Mantophasmatodea show the lowest corazonin precursor variation, while Dermaptera possess the most variable corazonin precursor sequences. The overall AED for the corazonin precursors is 0.82, the actual neuropeptide sequence is significantly better conserved (overall AED: 0.07).

## CRF-DH (Additional File 3J)

A precursor for the corticotropin-releasing factor-like diuretic hormone (CRF-DH) was found with a length of 138-284 AA in almost all polyneopteran orders. A partial sequence (N-terminal) of a possible CRF-DH was identified for the single species of Zoraptera. This sequence is not further considered here. The CRF-DH precursors contain a mostly well-conserved CRF-DH motif with C-terminal amidation site. The CRF-DH sequence is located in the middle of the precursor and is flanked by dibasic KR cleavage sites in most polyneopteran taxa. Notable exceptions are the precursors of Dermaptera which have a C-terminal RKR cleavage site (not in *Parapsalis infernalis*) and lack the N-terminal KR cleavage motif. Therefore, the sequences of the mature CRF-DHs of Dermaptera cannot be predicted with certainty and require biochemical confirmation first.

Most CRF-DHs of Polyneoptera consist of 46 AA, shorter sequences are indicated for several Plecoptera (42-46 AA), Ensifera (45-46 AA), Caelifera (44-46 AA), Mantophasmatodea (44 AA), and a single species of Blattodea (45 AA in *Nocticola*). Dermaptera have N-terminal extended CRF-DHs (see above). Due to considerable sequence variations, particularly in Dermaptera, an ancestral sequence of CRF-DH for Polyneoptera cannot be determined with certainty. All species contain a consensus sequence of PSLSIVNxxDVLRQRxxLExxRxRMR within the CRF-DH. The variable AA (x) within this sequence decrease significantly if the CFR-DHs of Dermaptera are not considered (PSLSIVNxxDVLRQRLLLExARRRMR). The median AED for the CFR-DH precursor is 0.30. Mantophasmatodea show the lowest CFR-DH precursor variation, while Ensifera possess the most variable CFR-DH precursor sequences. The overall AED for the CFR-DH precursors is 0.62, the actual neuropeptide sequence is significantly better conserved (overall AED: 0.31).

## CT-DH (Additional File 3K)

A precursor for the calcitonin-like diuretic hormone (CT-DH) was found with a length of 107-178 AA in all polyneopteran orders. The CT-DH precursors contain a highly conserved CT-DH motif with C-terminal amidation site. The CT-DH sequence is located in the middle of the precursor, N-terminally flanked by a dibasic KR cleavage site, and terminates upstream of an RRRR cleavage site (RKRR in Plecoptera).

All CT-DHs of Polyneoptera consist of 31 AA. The sequence GLDLGLSRGFSGSQAAKHLMGLAAANYAGGP-$NH_2$ which also occurs in the silverfish *T. domestica* was found in most orders of Polyneoptera (not in Zoraptera, Dermaptera, Caelifera). This sequence might therefore be regarded as ancestral for all Pterygota. The remarkable sequence conservation of CT-DH is unique for such a long neuropeptide. The few AA substitutions in CT-DHs of Polyneoptera are often lineage-specific and cover all species within the corresponding insect orders: $Phe^{10}$ to $Tyr^{10}$ and $Tyr^{27}$ to $Phe^{27}$ (Zoraptera), $Leu^6$ to $Met^6$ (Dermaptera), $Ser^7$ to $Asn^7$, and $Ser^{13}$ to $Ala/Thr^{13}$ (Caelifera). Significant intraordinal variation is only present in Caelifera. The median AED for the CFR-DH precursor is as low as 0.15. Mantophasmatodea and Grylloblattodea show the lowest CT-DH precursor variation, while Dermaptera possess the most variable CT-DH precursor sequences. The overall AED for the CT-DH precursors is 0.44, the actual neuropeptide sequence is significantly better conserved (overall AED: 0.05).

## Elevenin (Additional File 3L)

An elevenin precursor with a length of 99-170 AA was found in all polyneopteran orders. For a single species, *Paratemnopteryx couloniana* (Blattodea), we have identified a second and sequence-related precursor. The elevenin precursors contain quite variable elevenin motifs without C-terminal amidation site. The elevenin sequence in the precursor follows immediately the signal peptide and terminates upstream of a dibasic KR cleavage site (RKR or KKR in some Caelifera). In Caelifera, the predicted elevenin sequence contains a $Lys^2Arg^3$ motif that might be used as cleavage site. However, as noted above for CCHamide-2, KR motifs that immediately follow the signal peptide sequence do not necessarily function as cleavage signals for prohormone convertases.

The elevenins of Polyneoptera are variable in length and consist of 17 (Embioptera) up to 22 AA (Dermaptera, multiple species of Blattodea). Due to considerable sequence variations (**Figure 5**), an ancestral sequence of elevenin for Polyneoptera cannot be determined. Most polyneopteran taxa have species with a consensus C-terminus of CRGVAA-OH (CRGASA-OH in Mantophasmatodea) and a conserved position of the two

Cys residues; specific features also found in *T. domestica*. Significant intraordinal variation is present in Dermaptera, Caelifera, Embioptera, Phasmatodea, and Blattodea. The median AED for the elevenin precursor is 0.42. Grylloblattodea and Mantophasmatodea show the lowest elevenin precursor variation, while Caelifera possess the most variable elevenin precursor sequences. The overall AED for the elevenin precursors is 0.87, the actual neuropeptide sequence is better conserved (overall AED: 0.45).

## HanSolin (Additional File 3M)

HanSolin was recently described from *C. morosus* (2). A subsequent search for HanSolin in Coleoptera (25) also revealed orthologous precursors in these holometabolous insects. Here we found a single HanSolin precursor with a length of 88-139 AA in all polyneopteran orders. The HanSolin precursors contain quite variable HanSolin motifs with a conserved C-terminus; including a C-terminal amidation site. HanSolin is always located C-terminal in the precursor, N-terminally mostly flanked by a monobasic Arg cleavage site (RR in Mantophasmatodea, multiple Caelifera and Ensifera), and terminates upstream of a dibasic C-terminal RR cleavage site (occasionally KR or monobasic R for different orders).

The HanSolins of Polyneoptera seem to be very variable in length and consist of 8 (Caelifera) up to 16 AA (Ensifera). In many cases, the N-terminal sequences of the predicted mature peptides require biochemical confirmation. Due to considerable sequence variations, an ancestral sequence of HanSolin for Polyneoptera cannot be determined. Most polyneopteran taxa have species with a consensus C-terminal hexapepide of GQPLRW-$NH_2$ (GMPLRF-$NH_2$ in Zoraptera, GLPLRW-$NH_2$ in Mantophasmatodea); this C-terminus is also found in *T. domestica* (Bläser and Predel, unpublished). Significant intraordinal variation is present in Dermaptera, Plecoptera, Caelifera, Ensifera, Embioptera, and Phasmatodea. The median AED for the HanSolin precursor is 0.39 (**Figure 6**). Mantophasmatodea show the lowest HanSolin precursor variation, while Ensifera and Plecoptera possess the most variable HanSolin precursor sequences. The overall AED for the HanSolin precursors is 0.83, the actual neuropeptide sequence is significantly better conserved (overall AED: 0.33).

## MS (Additional File 3N)

A myosuppressin (MS) precursor with a length of 143-174 AA is present in all polyneopteran orders. All of these precursors contain a highly conserved MS motif with a C-terminal amidation site. For three species, *L. migratoria, Prosarthria teretrirostris* (both Caelifera) and *Hemimerus* sp. (Dermaptera), we identified a second precursor with related MS sequences. We also found a second transcript with identical MS sequences in five species of Mantodea, but with an insertion of 39 AA in the middle part of the precursor. The MS sequence is always located C-terminal in the precursor, N-terminally flanked by a dibasic KR cleavage site and terminates upstream of an RRR cleavage motif.

MS sequences are almost exclusively decapeptides. Only in one species of Blattodea (*Reticulitermes santonensis*) MS

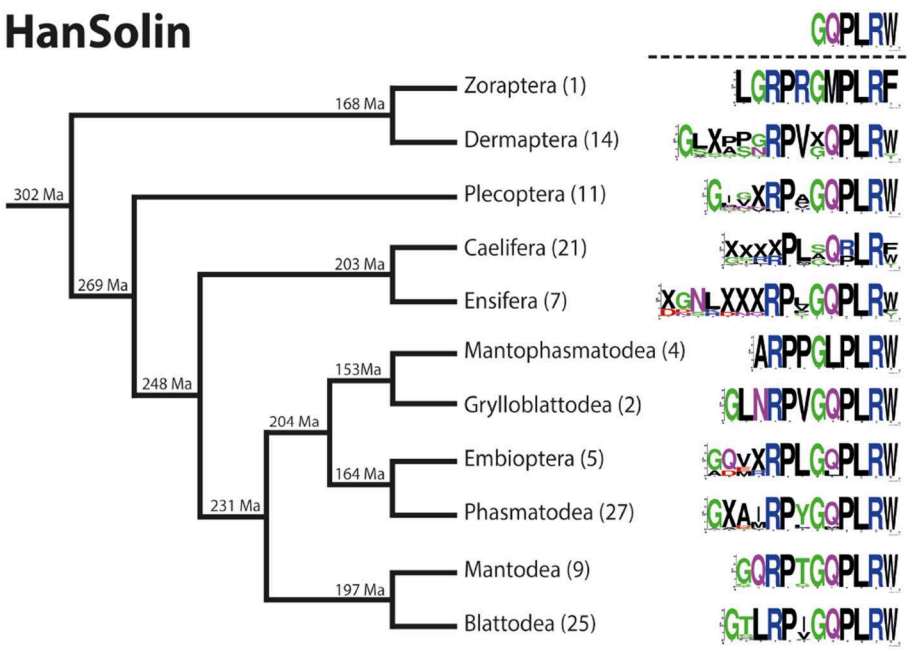



**FIGURE 5 |** Sequence logo representation showing the degree of amino acid sequence conservation of the HanSolin neuropeptides for each order; mapped on a phylogenetic tree of Polyneoptera [modified from Wipfler et al. (27)]. Only the completely obtained precursor sequences were considered, the respective number is given in parentheses for each taxon. An "X" in the sequence represents a gap. The hypothetical ancestral C-terminus of the HanSolin sequence in Polyneoptera is listed at the top.

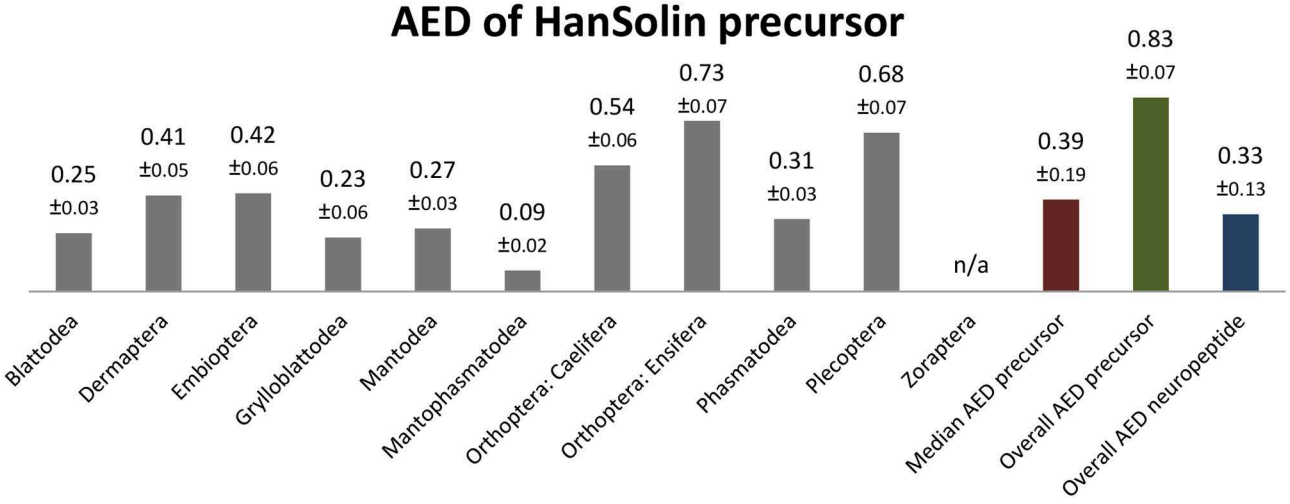



**FIGURE 6 |** Bar chart depicting the average evolutionary divergence (AED) of the HanSolin precursors for the different polyneopteran lineages with standard error estimates (S.E.). In addition, the median AED and the AED values for the available set of complete HanSolin precursors and predicted HanSolin sequences of Polyneoptera are given.

is a highly derived undecapeptide (KEDSQHMFLRF-$NH_2$). The sequence (p)QDVDHVFLRF-$NH_2$, which also occurs in Remipedia, was found in most orders of Polyneoptera (not in Caelifera and Dermaptera). This sequence might therefore be regarded as ancestral for all Hexapoda. Peptidomics confirmed for different polyneopteran taxa that the N-terminal Gln of MS is only partially converted to pGlu [e.g., (2, 40)]. With the exception of the MS of *R. santonensis* (see above) AA substitutions are restricted to the N-terminal AA (P/T in Caelifera, H in several Ensifera) and the position 6 ($Val^6$ to $Ile^6$ in Dermaptera). The second precursor of few species contains additional AA substitutions (EDVGHVFLRF-$NH_2$ *in L. migratoria*; KDIEHVFLRF-$NH_2$ in *P. teretrirostris*, QDVHHNFLRF-$NH_2$ in *Hemimerus* sp.). The median AED for the MS precursor is 0.26. Grylloblattodea and Mantophasmatodea show the lowest MS precursor variation in Polyneoptera, while Ensifera possess the most variable MS precursors. The overall AED for the MS precursors is 0.57, the actual neuropeptide sequence is significantly better conserved (overall AED: 0.05).

## NPF-1 (Additional File 3O)

The insect neuropeptide F-1 (NPF-1) precursor with a length of 81–97 AA was found in all polyneopteran orders. In most taxa we have also identified a second and longer transcript showing an insertion of about 40 AA in the middle of the NPF-1 neuropeptide (= NPF-$1_b$). Such transcripts are also known from *T. domestica* and are therefore a basic feature of Pterygota. The remaining sequences of NPF-1 are identical in both transcripts. The NPF-1 precursors contain a well-conserved NPF-1 motif with C-terminal amidation site. To a slightly lesser extent, this also applies to the insertion in the long transcript. The NPF-1 sequence in the precursor follows immediately the signal peptide and terminates upstream of a dibasic KR cleavage site.

Most NPF-$1_a$ neuropeptides of Polyneoptera consist of 33 AA, but the full range is 30–36 AA. The C-terminus LQELDRYYSQVARPRF-$NH_2$ is fully conserved in the majority of Polyneoptera (E to M/R in Embioptera and Mantophasmatodea; V to N/K in Plecoptera). Particularly significant intraordinal variation of the N-terminus is present in Dermaptera, Plecoptera, Caelifera, Grylloblattodea, and Embioptera. The median AED for the CFR-DH precursor is 0.27. Mantophasmatodea and Mantodea show the lowest CT-DH precursor variation, while Ensifera and Plecoptera possess the most variable CT-DH precursor sequences. The overall AED for the CT-DH precursors is 0.45, the actual neuropeptide sequence is significantly better conserved (overall AED: 0.19).

## NPF-2 (Additional File 3P)

An insect neuropeptide F-2 (NPF-2) precursor with a length of 85–134 AA was found in all polyneopteran orders. The NPF-2 precursors contain a quite variable NPF-2 motif with a well-conserved C-terminus and C-terminal amidation site. The NPF-2 sequence in the precursor follows immediately the signal peptide and terminates upstream of a dibasic KR cleavage site.

The NPF-2 neuropeptides of Polyneoptera are variable in length and consist of 43–47 AA. Only the C-terminus PRF-$NH_2$ is fully conserved in all analyzed species while a C-terminal RPRF-$NH_2$ was at least found in members of all order of Polyneoptera. Thus, the information about the C-terminal AA is not sufficient to distinguish between the neuropeptides NPF-1 and NPF-2. Particularly significant intraordinal variation of the N-terminus is present in Dermaptera, Plecoptera, Ensifera, and Embioptera. The median AED for the NPF-2 precursor is 0.29. Mantophasmatodea and Mantodea show the lowest NPF-2 precursor variation, while Plecoptera possess the most variable NPF-2 precursor sequences. The overall AED for the NPF-2

precursors is 0.65, the actual neuropeptide sequence is slightly better conserved (overall AED: 0.41).

## Proctolin (Additional File 3Q)

A proctolin precursor with a length of 74–104 AA is present in almost all polyneopteran orders. The only exception was found in Dermaptera, where the proctolin precursor is absent. For the single species of Zoraptera (*Zorotypus caudelli*) we identified a second precursor. All these precursors contain a highly conserved proctolin motif without a C-terminal amidation site. The proctolin sequence immediately follows the signal peptide and terminates upstream of a monobasic Arg cleavage site.

Proctolin sequences are exclusively pentapeptides. The sequence RYLPT-OH, which also occurs in Remipedia and even Myriapoda, was found in all orders of Polyneoptera. This sequence might therefore be regarded as ancestral for Hexapoda. With the exception of the proctolin of *Systella rafflesii* ($Pro^4$ to $His^4$) and *Zubovskia* sp. ($Thr^5$ to $Val^5$; both Caelifera), all species possess the original sequence RYPLT-OH. The median AED for the proctolin precursor is 0.31. Grylloblattodea, Mantophasmatodea, and Mantodea show the lowest proctolin precursor variation in Polyneoptera, while Ensifera and Caelifera possess the most variable proctolin precursors. The overall AED for the ACP precursors is 0.52, the actual neuropeptide sequence is significantly better conserved (overall AED: 0.00).

## RFLamide (Additional File 3R)

RFLamides were only recently described from *C. morosus* (2). An RFLamide precursor with a length of 122–211 AA is present in all polyneopteran orders. Most precursors contain a well-conserved RFLamide with a C-terminal amidation site. The RFLamide sequence is always located C-terminal in the precursor, N-terminally flanked by an RKR or RRR cleavage site (dibasic RR in *Protonemura ausonia*, Plecoptera) and terminates upstream of quite variable cleavage motifs (monobasic Arg up to 5 basic AA which terminate the precursor sequence).

RFLamides are mostly duodecapeptides. Only in Mantophasmatodea and Grylloblattodea the RFLamides are 14 mers with an extended C-terminus. In these taxa the first Arg of the original cleavage motif is replaced by a Met, resulting in RFLamides with two additional AA without a C-terminal amidation site. The unique C-terminus of these insects is a remarkable synapomorphy of Mantophasmatodea and Grylloblattodea. The sequence PASAIFTNIRFL-$NH_2$ was found in most orders of Polyneoptera (not in Zoraptera, Mantophasmatodea, Grylloblattodea). This sequence might therefore be regarded as ancestral for all Polyneoptera (**Figure 7**). Amino acid substitutions in RFLamides of Polyneoptera are largely limited to substitutions of $Ser^3$, $Ala^4$, and $Ile^5$. Mantophasmatodea and Grylloblattodea have the most derived sequences, each with several lineage-specific features; in addition to the distinct C-terminus, which is identical in both taxa. Significant intraordinal variation is present in Plecoptera, Ensifera, and Embioptera. The median AED for the RFLamide precursor is 0.28 (**Figure 8**). Grylloblattodea (only 2 species) show the lowest RFLamide precursor variation in Polyneoptera, while Plecoptera possess the most variable RFLamide precursors. The overall AED for the RFLamide precursors is 0.62, the actual neuropeptide sequence is significantly better conserved (overall AED: 0.14).

## SIFamide (Additional File 3S)

A SIFamide precursor with a length of 71–103 AA is present in all polyneopteran orders. In two blattodean species (*Prorhinotermes* and *Schultesia*) we found a second SIFamide precursor with identical SIFamide sequences; probably slightly different alleles. Most precursors contain a very well-conserved SIFamide with a C-terminal amidation site. An exception was found in *R. virgo* (Embioptera), which has three consecutive copies of SIFamide in the precursor and therefore cannot not be treated as a single-copy peptide. The SIFamides (actually SIYamides) of this species do indeed have more derived sequences and are not considered in our analyses. Whether the transition to multiple copies is a specific feature of Embioptera cannot yet be determined, since we could not find SIFamide precursors in the other Embioptera species. The SIFamide sequence in the precursors of all other species follow immediately the signal peptide and terminates upstream of a dibasic KR cleavage site.

The SIFamides are mostly duodecapeptides or longer. Only in the SIFamides of Dermaptera the N-terminal AA is missing (= undecapeptides). Since the length of the signal peptide of SIFamide precursors cannot always be predicted with certainty (41), the N-terminus of SIFamides should be confirmed by peptidomics in taxa not yet examined. The sequence TYRKPPFNGSIF-$NH_2$ was found in most orders of Polyneoptera (not in Dermaptera). This sequence might therefore be regarded as ancestral for all Polyneoptera. In the SIFamide of *T. domestica* (Zygentoma) only the N-terminal AA is different ($Thr^1$-$Gly^1$). Amino acid substitutions in SIFamides of Polyneoptera are largely limited to substitutions of $Thr^1$ and $Tyr^2$. The median AED for the SIFamide precursor is 0.22. Grylloblattodea show the lowest SIFamide precursor variation in Polyneoptera, while Caelifera possess the most variable SIFamide precursors. The overall AED for the SIFamide precursors is 0.5, the actual neuropeptide sequence is significantly better conserved (overall AED: 0.09).

A second gene coding for a related neuropeptide, SMYamide, was described for *L. migratoria* (Caelifera) and *Zootermopsis nevadensis* [Blattodea; (1)]. We found SMYamide precursors in Caelifera, Ensifera, Embioptera, Phasmatodea, Mantodea, and Blattodea. The phylogenetic position of these taxa suggests that the SMYamide gene has evolved within the Polyneoptera. This is corroborated by the fact that so far no orthologs of SMY genes have been reported from any other insect.

## sNPF (Additional File 3T)

A short neuropeptide F (sNPF) precursor with a length of 86–134 AA is present in all polyneopteran orders. All of these precursors contain a highly conserved sNPF motif with a C-terminal amidation site. A specific feature of most Caelifera is the presence of a second and longer sNPF neuropeptide immediately after the first sNPF sequence in the precursor. While the two species of *Xya* (Caelifera) still show the original pattern with a single sNPF sequence, *P. teretrirostris* (Caelifera) even has three

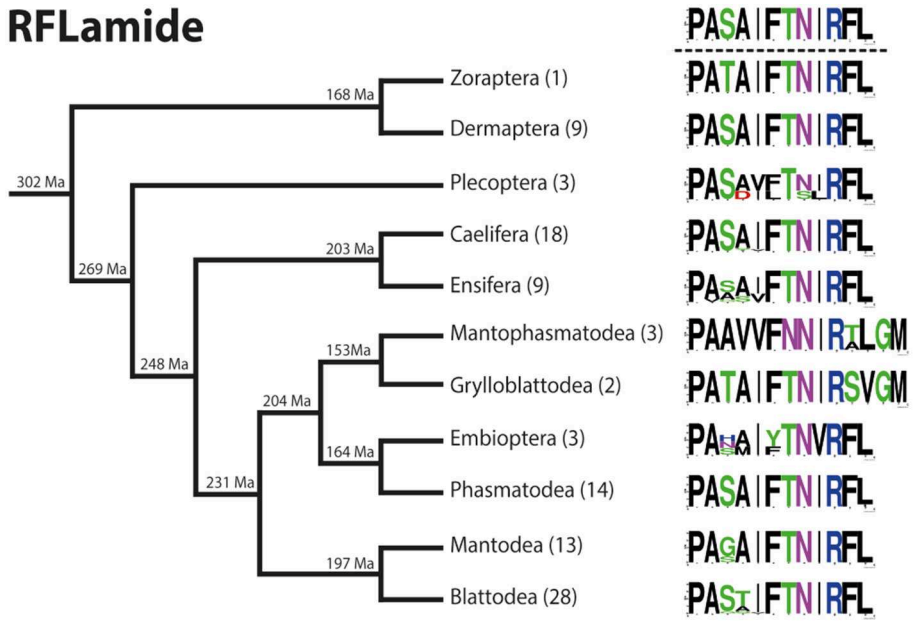



**FIGURE 7 |** Sequence logo representation showing the degree of amino acid sequence conservation of the RFLamide neuropeptides for each order; mapped on a phylogenetic tree of Polyneoptera [modified from Wipfler et al. (27)]. Only the completely obtained precursor sequences were considered, the respective number is given in parentheses for each taxon. An "X" in the sequence represents a gap. The hypothetical ancestral state of the RFLamide sequence in Polyneoptera is listed at the top.

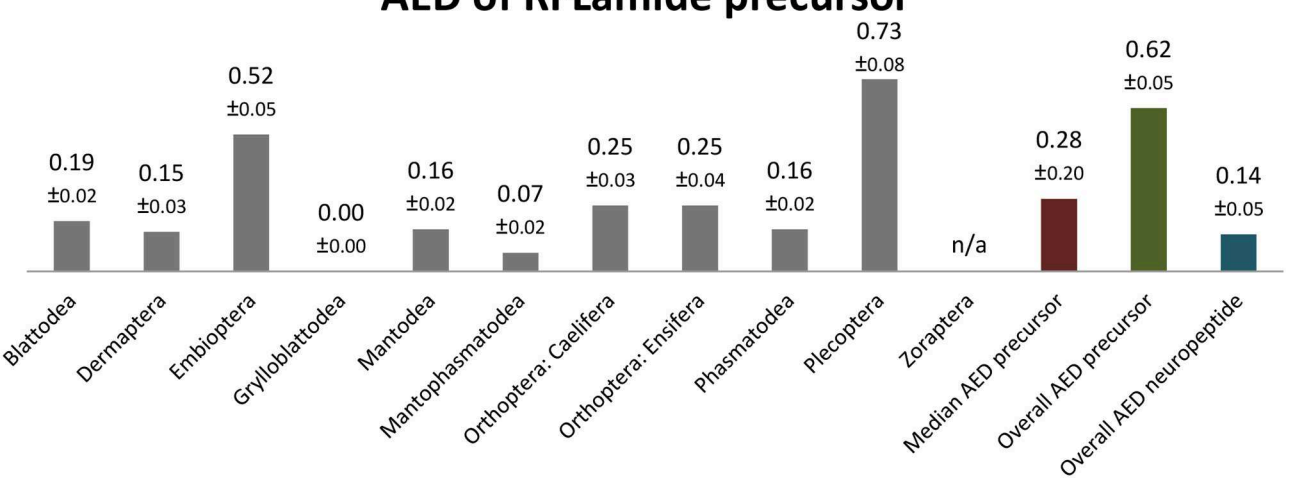



**FIGURE 8 |** Bar chart depicting the average evolutionary divergence (AED) of the RFLamide precursors for the different polyneopteran lineages with standard error estimates (S.E.). In addition, the median AED and the AED values for the available set of complete RFLamide precursors and predicted RFLamide sequences of Polyneoptera are given.

consecutive sNPFs in the precursor. The sequence of the N-terminal sNPF of the Caelifera with multiple sNPFs is highly conserved and closely resembles the orthologous sNPFs of the other polyneopteran orders. Therefore, it was included in our analyses. The sNPF sequence is always located in the middle of the precursor, N-terminally flanked by a dibasic RK cleavage site, and terminates upstream of a dibasic RR cleavage site.

Short NPF sequences (in Caelifera only the N-terminal sNPF sequence) are exclusively undecapeptides with a potential secondary cleavage site ($Arg^3$). The sequence SNRSPSLRLRF-$NH_2$ which also occurs in *T. domestica*, was found in several orders of Polyneoptera (Zoraptera, Dermaptera, Plecoptera, Caelifera and Ensifera). This sequence might therefore be regarded as ancestral for all Pterygota. Apparently the

sister group of Zoraptera + Dermaptera (i.e., the remaining polyneopteran orders) originally had two alleles coding for Ser or Ala as the N-terminal AA. Several orders of this group (Plecoptera, Caelifera, Ensifera, Grylloblattodea, Phasmatodea) still have species either with $Ser^1$ or $Ala^1$, while in Mantophasmatodea, Embioptera, Mantodea, and Blattodea the sNPF with $Ala^1$ has completely replaced the original Ser at this position. Other AA substitutions are restricted to the second AA ($Gln^2$ to $Ser^2$) and have been detected in a few Grylloblattodea and Caelifera and in all Embioptera. The median AED for the sNPF precursor is 0.20. Grylloblattodea and Mantophasmatodea show the lowest sNPF precursor variation in Polyneoptera, while Embioptera and Plecoptera possess the most variable sNPF precursors. The overall AED for the sNPF precursors is 0.44, the actual neuropeptide sequence is significantly better conserved (overall AED: 0.05).

### Trissin (Additional File 3U)

A trissin precursor with a length of 88–117 AA is present in almost all polyneopteran orders. The only exceptions were found in Dermaptera and Zoraptera, where trissin precursors are absent. For two species of the genus *Xya* (Caelifera), we have identified two trissin precursors with moderately (trissin 1) or strongly modified N-termini (trissin 2). Otherwise, the trissin precursors contain usually a well-conserved trissin motif without C-terminal amidation site. The trissin sequence in the precursor follows immediately after the signal peptide and terminates upstream of a tribasic RKR cleavage site (KKR in 2 of 26 Mantodea species and dibasic KR in 2 of 6 Ensifera species).

Most trissins of Polyneoptera consist of 27 AA. For most Caelifera a truncated sequence of trissin with a single AA (instead of two) preceding the N-terminal Cys is predicted by SignalP-5.0; trissin 1 of *Xya* probably starts directly with the N-terminal Cys. Trissins 1 of *Xya variegata* and *X. japonica* additionally show an insertion of Ser downstream of the N-terminal Cys. In Ensifera, the N-terminal of trissin is not always clearly predicted; probably it starts directly with the N-terminal Cys. Generally, the N-terminal cleavage of trissins should be confirmed by peptidomics. However, trissin has not been detected biochemically from any polyneopteran species so far. In two species of Ensifera (*Ceuthophilus* sp. and *Diestrammena asynamora*) the first Arg of the C-terminal cleavage motif is replaced by Ser, which probably leads to an extended C-terminus (NYLS-OH instead of NYL-OH). All species of the ensiferan infraorder Gryllidea whose trissin sequence has been identified (*Ceutophilus* sp., *Gryllotalpa* sp., *Neonetus* sp.) show an insertion of Asp in the middle of the sequence, indicating a synapomorphy. The sequence LSCDSCGRECXXXCGTRNFRTCCFNYL-OH (XXX: no ancestral AA assigned) was found in *T. domestica* and most orders of Polyneoptera. This sequence might therefore be regarded as ancestral for all Pterygota. Amino acid substitutions in trissins of Polyneoptera are mainly limited to substitutions of AA at positions 11–13 and 18. Significant intraordinal variation is present in Plecoptera, Caelifera, Ensifera, and Blattodea. Distinct lineage-specific features are substitutions of $Ser^4$ to $Phe^4$/$Val^4$ (Caelifera) or $Ile^4$ (Ensifera), $Phe^{24}$ to $Leu^{24}$/$Tyr^{24}$ (Caelifera/Ensifera), and $Arg^{21}$ to $Val^{21}$/$His^{21}$ (Ensifera). The median AED for the trissin precursor is 0.23. Mantophasmatodea show the lowest trissin precursor variation, while Ensifera possess the most variable trissin precursor sequences. The overall AED for the trissin precursors is 0.59, the actual neuropeptide sequence is significantly better conserved (overall AED: 0.20).

## CONCLUSIONS

In our analysis we examined the single-copy precursor sequences of 21 neuropeptide genes of Polyneoptera. The neuropeptides of 17 of these precursors are C-terminally amidated (not AST-CC, elevenin, proctolin, trissin), which prevents rapid degradation by exopeptidases and thus supports their functions as hormones. Only very few neuropeptide genes coding for single-copy neuropeptides are completely missing in a given polyneopteran order. Dermaptera have no ACP, proctolin and trissin, Ensifera do not have CCHamide-1, and in most Embioptera we could not detect any SIFamide precursor (with the exception of a multiple-copy SIFa precursor in *R. virgo*, see above). Furthermore, we did not find precursors for corazonin, CRF-DH and trissin in Zoraptera, but only one species of this order could be analyzed. Therefore, the absence of the respective neuropeptide genes has yet to be confirmed for Zoraptera. For most orders and also for the individual species within these orders, we have found all single-copy precursors, a feature already documented for the "basal" hexapods, which represent the sister group of the Pterygota [winged insects; (24)]. In contrast, peptide gene losses are more frequent in the much more species-rich and ecologically significant Holometabola. The fruit fly *Drosophila melanogaster*, which is used as a model organism in molecular biology, neurobiology and also physiology, is a good example in this context as it lacks not less than 6 of the 21 peptidergic systems analyzed here (ACP, AT, Elevenin, HanSolin, NPF-2, RFLamide) (12).

The sequence conservation of the precursor sequences, including the signal peptides, varies for the different neuropeptide genes. Low overall AED values (AST-C: 0.36; sNPF, CT-DH: 0.44; CCAP, NPF-1: 0.45; see **Additional File 3**) contrast with high AED values (CNMamide: 0.95; Elevenin: 0.87; Hansolin: 0.83; see **Additional File 3**), which are significantly above the average value of 0.63 calculated for all neuropeptide precursors (**Figure 9**). As expected, the sequences of single-copy neuropeptides within the precursors are much better conserved (overall AED 0.16; **Figure 9**). However, the extent of sequence conservation across Polyneoptera is remarkably different between the different neuropeptides. Neuropeptides such as proctolin, CCAP, AST-C, sNPF, MS, and CT-DH (overall AED $\leq$ 0.05) are almost identical in all taxa and the most common sequence always represents the predicted ancestral sequence of Pterygota (sNPF, CT-DH) or even the ancestral sequence of Hexapoda (proctolin, CCAP, AST-C, MS). For all neuropeptides with very high AED values (Elevenin: 0.45; NPF-2: 0.41; Hansolin: 0.33), the sequence ancestral to Polyneoptera could not be determined. Many of the neuropeptides with high AED values have long sequences, but this does not necessarily

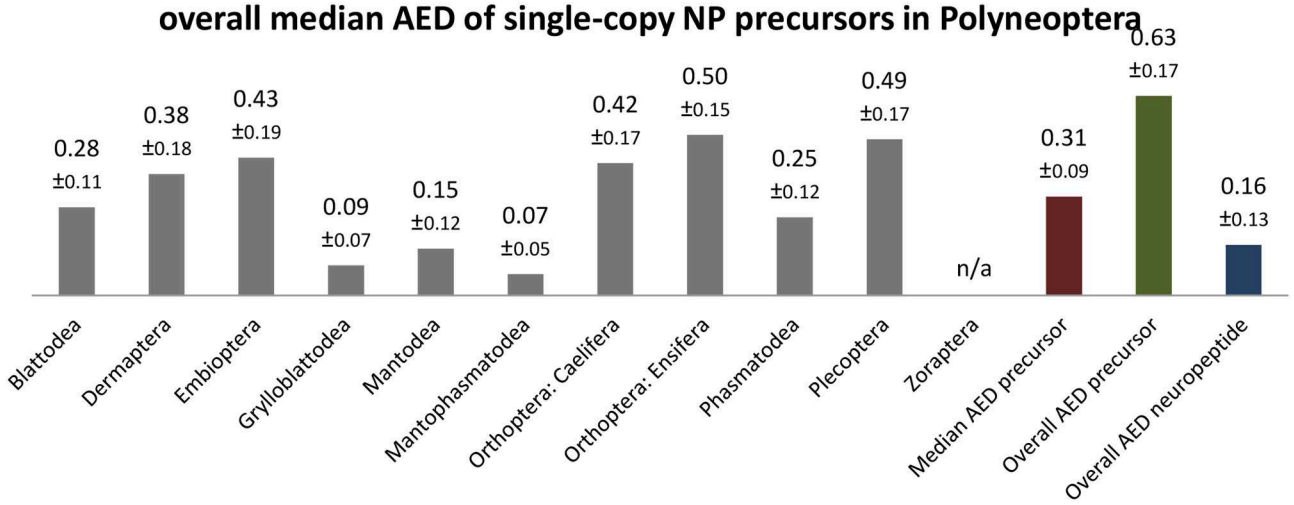



**FIGURE 9 |** Bar chart depicting the overall median average evolutionary divergence (AED) of single-copy neuropeptide (NP) precursors for the different polyneopteran lineages with standard error estimates (S.E.). In addition, the median AED and the overall AED values for the available set of complete neuropeptide precursors and predicted neuropeptide sequences of Polyneoptera are given.

lead to high AED values, as is shown for example with CT-DH (31 AA; AED: 0.05).

The overall median AEDs for the single-copy neuropeptides and precursors differ significantly between the polyneopteran orders. This was to be expected, since the different lineages evolved independently of each other over different periods of time (see **Figure 2**). In addition, several polyneopteran orders (e.g., Grylloblattodea and Mantophasmatodea) represent relict groups with only a few extant and rather closely related taxa. Thus, these orders show particularly low AEDs, while Dermaptera, Plecoptera and Orthoptera (Ensifera + Caelifera) have much higher intra-ordinal sequence diversity (**Figure 9**). The relatively high AEDs for Embioptera were somewhat unexpected in this context. Although the AEDs for the various neuropeptide precursors of the different Polyneoptera are mostly in the range of the median AED for all neuropeptide precursors, there are striking exceptions. This is especially true for Mantodea (significantly lower AEDs for NPF-1 and−2) and Dermaptera (significantly lower AED for AST-CC). A comparison of AEDs in the orthopteran sister groups Ensifera and Caelifera also shows very different AEDs for the different neuropeptide precursors, either in favor of Ensifera or Caelifera (**Additional File 3**). This means that in the evolution of the sequences of neuropeptide precursors there have been some striking increases or decreases in the AA substitution rate, which cannot be directly related to a uniform development of the peptidergic system of a given taxon or to a specific neuropeptide gene.

A number of derived neuropeptide sequences were found, showing sequence motifs (= synapomorphies) typical only for representatives of a specific polyneopteran lineage. This has to be separated from intra-ordinal variation. Within the respective lineages, the derived sequences are often well-conserved (**Additional File 3**). However, surprisingly few examples of derived sequences have been found that are typical of two or more polyneopteran orders. One clear example is the substitution within the C-terminal cleavage motif of RLFamides, which probably occurred in the last common ancestor of Mantophasmatodea and Grylloblattodea. This substitution prevents the C-terminal amidation and is typical of all Mantophasmatodea and Grylloblattodea. Furthermore, the absence of trissin in both Dermaptera and Zoraptera (here, however, only a single transcriptome was available) indicates that the loss of this neuropeptide already occurred in the last common ancestor of these two lineages. Typical for most Dictyoptera (Mantodea + Blattodea) is $Gln^{11}$ of trissin, which is only found in this taxon.

Overall, the single-copy neuropeptide precursors of the Polyneoptera show a relatively high degree of sequence conservation. Basic features of these precursors in this very heterogeneous insect group are explained here in detail for the first time. Further insights into the evolution of neuropeptides can be expected from future analyzes of the much more variable multiple-copy neuropeptides.

## DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/**Supplementary Material**.

## AUTHOR CONTRIBUTIONS

MB and RP contributed to the conception and design of the study. MB mined the transcriptomes for neuropeptide precursors and wrote the first draft of the manuscript. All authors contributed to final version of the manuscript and approved the submitted manuscript.

## FUNDING

This study was funded by the Deutsche Forschungsgemeinschaft (RP 766/11-1).

## ACKNOWLEDGMENTS

We are grateful to the entire 1KITE consortium, especially to Bernhard Misof, Alexander Donath, Benjamin Wipfler (ZFMK Bonn, Germany) and Karen Meusemann (University of Freiburg, Germany) for providing early access to the 1KITE transcriptome assemblies and raw data. Furthermore, we thank Sander Liessem and Lapo Ragionieri (University of Cologne, Germany) for supporting the transcriptome analysis.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fendo.2020.00197/full#supplementary-material

## REFERENCES

1. Veenstra JA. The contribution of the genomes of a termite and a locust to our understanding of insect neuropeptides and neurohormones. *Front Physiol*. (2014) 5:454. doi: 10.3389/fphys.2014.00454
2. Liessem S, Ragionieri L, Neupert S, Büschges A, Predel R. Transcriptomic and neuropeptidomic analysis of the stick insect, *Carausius morosus*. *J Proteome Res*. (2018) 17:2192–204. doi: 10.1021/acs.jproteome.8b00155
3. Bläser M, Misof B, Predel R. The power of neuropeptide precursor sequences to reveal phylogenetic relationships in insects: a case study on Blattodea. *Mol Phylogenet Evol.* (2020) 66:479–506. doi: 10.1016/j.ympev.2019.106686
4. Elphick MR, Mirabeau O, Larhammar D. Evolution of neuropeptide signalling systems. *J Exp Biol*. (2018) 221:jeb151092. doi: 10.1242/jeb.151092
5. Jékely G. Global view of the evolution and diversity of metazoan neuropeptide signaling. *Proc Natl Acad Sci USA*. (2013) 110:8702–7. doi: 10.1073/pnas.1221833110
6. Mirabeau O, Joly J-S. Molecular evolution of peptidergic signaling systems in bilaterians. *Proc Natl Acad Sci USA*. (2013) 110:E2028–37. doi: 10.1073/pnas.1219956110
7. Bauknecht P, Jékely G. Large-scale combinatorial deorphanization of platynereis neuropeptide GPCRs. *Cell Rep*. (2015) 12:684–93. doi: 10.1016/j.celrep.2015.06.052
8. Nachman RJ, Holman GM, Haddon WF, Ling N. Leucosulfakinin, a sulfated insect neuropeptide with homology to gastrin and cholecystokinin. *Science*. (1986) 234:71–3. doi: 10.1126/science.3749893
9. Blair JE, Hedges SB. Molecular phylogeny and divergence times of deuterostome animals. *Mol Biol Evol*. (2005) 22:2275–84. doi: 10.1093/molbev/msi225
10. Simakov O, Kawashima T, Marlétaz F, Jenkins J, Koyanagi R, Mitros T, et al. Hemichordate genomes and deuterostome origins. *Nature*. (2015) 527:459–65. doi: 10.1038/nature16150
11. Schoofs L, De Loof A, Van Hiel MB. Neuropeptides as regulators of behavior in insects. *Annu Rev Entomol*. (2017) 62:35–52. doi: 10.1146/annurev-ento-031616-035500
12. Nässel DR, Zandawala M. Recent advances in neuropeptide signaling in Drosophila, from genes to physiology and behavior. *Prog Neurobiol*. (2019) 179:101607. doi: 10.1016/j.pneurobio.2019.02.003
13. Li B, Predel R, Neupert S, Hauser F, Tanaka Y, Cazzamali G, et al. Genomics, transcriptomics, and peptidomics of neuropeptides and protein hormones in the red flour beetle *Tribolium castaneum*. *Genome Res*. (2008) 18:113–22. doi: 10.1101/gr.6714008
14. Ons S, Richter F, Urlaub H, Pomar RR. The neuropeptidome of Rhodnius prolixus brain. *Proteomics*. (2009) 9:788–92. doi: 10.1002/pmic.200800499
15. Huybrechts J, Bonhomme J, Sebastian M, Prunier-Leterme N, Dombrovsky A, Abdel-Latief M, et al. Neuropeptide and neurohormone precursors in the pea aphid *Acyrthosipon pisum*. *Insect Mol Biol*. (2010) 19(Suppl 2):87–95. doi: 10.1111/j.1365-2583.2009.00951.x
16. Predel R, Neupert S, Garczynski SF, Crim JW, Brown MR, Russell WK, et al. Neuropeptidomics of the mosquito *Aedes aegypti*. *J Proteome Res*. (2010) 9:2006–15. doi: 10.1021/pr901187p
17. Predel R, Neupert S, Derst C, Reinhardt K, Wegener C. Neuropeptidomics of the bed bug *Cimex lectularius*. *J Proteome Res*. (2018) 17:440–54. doi: 10.1021/acs.jproteome.7b00630
18. Pandit AA, Ragionieri L, Marley R, Yeoh JGC, Inward DJG, Davies S-A, et al. Coordinated RNA-Seq and peptidomics identify neuropeptides and G-protein coupled receptors (GPCRs) in the large pine weevil *Hylobius abietis*, a major forestry pest. *Insect Biochem Mol Biol*. (2018) 101:94–107. doi: 10.1016/j.ibmb.2018.08.003
19. Hummon AB, Richmond TA, Verleyen P, Baggerman G, Huybrechts J, Ewing MA, et al. From the genome to the proteome: uncovering peptides in the Apis brain. *Science*. (2006) 314:647–9. doi: 10.1126/science.1124128
20. Hauser F, Neupert S, Williamson M, Predel R, Tanaka Y, Grimmelikhuijzen CJ. Genomics and peptidomics of neuropeptides and protein hormones present in the parasitic wasp *Nasonia vitripennis*. *J Proteome Res*. (2010) 9:5296–310. doi: 10.1021/pr100570j
21. Ragionieri L, Predel R. The neuropeptidome of Carabus (Coleoptera, Adephaga: Carabidae). *Insect Biochem Mol Biol*. (2020) 20:103309. doi: 10.1016/j.ibmb.2019.103309
22. Brown BE, Starratt AN. Isolation of proctolin, a myotropic peptide, from *Periplaneta americana*. *J Insect Phys*. (1975) 23:1879–81. doi: 10.1016/0022-1910(75)90257-7
23. Yeoh JGC, Pandit AA, Zandawala M, Nässel DR, Davies SA, Dow JAT. DINeR: database for insect neuropeptide research. *Insect Biochem Mol Biol*. (2017) 86:9–19. doi: 10.1016/j.ibmb.2017.05.001
24. Derst C, Dircksen H, Meusemann K, Zhou X, Liu S, Predel R. Evolution of neuropeptides in non-pterygote hexapods. *BMC Evol Biol*. (2016) 16:51. doi: 10.1186/s12862-016-0621-4
25. Veenstra JA. Coleoptera genome and transcriptome sequences reveal numerous differences in neuropeptide signaling between species. *PeerJ*. (2019) 7:e7144. doi: 10.7717/peerj.7144
26. Wegener C, Gorbashov A. Molecular evolution of neuropeptides in the genus Drosophila. *Genome Biol*. (2008) 9:R131. doi: 10.1186/gb-2008-9-8-r131
27. Wipfler B, Letsch H, Frandsen PB, Kapli P, Mayer C, Bartel D, et al. Evolutionary history of Polyneoptera and its implications for our understanding of early winged insects. *Proc Natl Acad Sci USA*. (2019) 116:3024–9. doi: 10.1073/pnas.1817794116
28. Misof B, Liu S, Meusemann K, Peters RS, Donath A, Mayer C, et al. Phylogenomics resolves the timing and pattern of insect evolution. *Science*. (2014) 346:763–67. doi: 10.1126/science.1257570
29. Artimo P, Jonnalagedda M, Arnold K, Baratin D, Csardi G, Castro E, et al. ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Res*. (2012) 40:W597–603. doi: 10.1093/nar/gks400
30. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. (2013) 30:772–80. doi: 10.1093/molbev/mst010
31. Hall TA. "BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Sympos Seri*. (1999) 41:c1979–c2000.
32. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol*. (2018) 35:1547–9. doi: 10.1093/molbev/msy096
33. Zuckerkandl E, Pauling L. Evolutionary divergence and convergence in proteins. In: *Edited In Evolving Genes and Proteins*. Bryson V, Vogel HJ, editors. New York, NY: Academic Press. (1965). p. 97–166. doi: 10.1016/B978-1-4832-2734-4.50017-6

34. Crooks GE, Hon G, Chandonia J-M, Brenner SE. WebLogo: a sequence logo generator. *Genome Res*. (2004) 14:1188–90. doi: 10.1101/gr.849004
35. Simon S, Letsch H, Bank S, Buckley TR, Donath A, Liu S, et al. Old world and new world phasmatodea: phylogenomics resolve the evolutionary history of stick and leaf insects. *Front Ecol Evol*. (2019) 7:744. doi: 10.3389/fevo.2019.00345
36. Clynen E, Huybrechts J, Verleyen P, De Loof A, Schoofs L. Annotation of novel neuropeptide precursors in the migratory locust based on transcript screening of a public EST database and mass spectrometry. *BMC Genom*. (2006) 7:201. doi: 10.1186/1471-2164-7-201
37. Veenstra JA. Allatostatins C, double C and triple C, the result of a local gene triplication in an ancestral arthropod. *Gen Comp Endocr*. (2016) 230–231:153–7. doi: 10.1016/j.ygcen.2016.04.013
38. Predel R, Neupert S, Russell WK, Scheibner O, Nachman RJ. Corazonin in insects. *Peptides*. (2007) 28:3–10. doi: 10.1016/j.peptides.2006.10.011
39. Predel R, Neupert S, Huetteroth W, Kahnt J, Waidelich D, Roth S. Peptidomics-based phylogeny and biogeography of Mantophasmatodea (Hexapoda). *Syst Biol*. (2012) 61:609–29. doi: 10.1093/sysbio/sys003
40. Neupert S, Fusca D, Schachtner J, Kloppenburg P, Predel R. Toward a single-cell-based analysis of neuropeptide expression in *Periplaneta americana* antennal lobe neurons. *J Comp Neurol*. (2012) 520:694–716. doi: 10.1002/cne.22745
41. Gellerer A, Franke A, Neupert S, Predel R, Zhou X, Liu S, et al. Identification and distribution of SIFamide in the nervous system of the desert locust *Schistocerca gregaria*. *J Comp Neurol*. (2015) 523:108–25. doi: 10.1002/cne.23671

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

## 2.3 The Power of Neuropeptide Precursor Sequences to Reveal Phylogenetic Relationships in Insects: a Case Study on Blattodea.

Marcel Bläser, Bernhard Misof, Reinhard Predel

Published in Molecular Phylogenetics and Evolution 143 (2020) 106686. 

Supplementary Material is provided as Supplementary Material SM 3.

Author contributions

MB, BM and RP conceived and designed the experiments; MB mined the transcriptome dataset for neuropeptide precursor sequences and performed phylogenetic analyses; MB and RP analyzed the data; MB prepared figures and tables and wrote the first draft of the manuscript; MB, BM and RP wrote the paper.

Molecular Phylogenetics and Evolution 143 (2020) 106686

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

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

# The power of neuropeptide precursor sequences to reveal phylogenetic relationships in insects: A case study on Blattodea

Marcel Bläser[a,*], Bernhard Misof[b], Reinhard Predel[a,*]

[a] University of Cologne, Institute for Zoology, Zuelpicher Straße 47b, 50674 Cologne, Germany
[b] ZFMK, Centre for Molecular Biodiversity Research (zmb), Adenauerallee 160, 53113 Bonn, Germany

ARTICLE INFO

*Keywords:*
Insect
Phylogeny
Transcriptome
Gene family evolution
Neuropeptide
Blattodea

ABSTRACT

Recent state-of-the-art analyses in insect phylogeny have exclusively used very large datasets to elucidate higher-level phylogenies. We have tested an alternative and novel approach by evaluating the potential phylogenetic signals of identified and relatively short neuropeptide precursor sequences with highly conserved functional units. For that purpose, we examined available transcriptomes of 40 blattodean species for the translated amino acid sequences of 17 neuropeptide precursors. Recently proposed intra-ordinal relationships of Blattodea, based on the analysis of 2370 protein-coding nuclear single-copy genes (Evangelista et al., 2019), were corroborated with maximum support. The functionally different precursor units were analyzed separately for their phylogenetic information. Although the degree of information was different in the different sequence motifs, all precursor units contained phylogenetic informative data at the ordinal level, and their separate analysis did not reveal contradictory topologies. This study is the first comprehensive exploitation of complete neuropeptide precursor sequences of arthropods in such a context and demonstrates the applicability of these rather short but conserved sequences for an alternative, fast and simple analysis of phylogenetic relationships.

## 1. Introduction

Recent efforts to resolve phylogenetic relationships in insects have addressed a number of questions by using extensive molecular datasets (e.g. Misof et al., 2014, Zhang et al., 2018). This also applies to relationships within Blattodea, including the placement of termites (e.g. Klass, 1995, Inward et al., 2007, Djernæs et al., 2012, 2015, Legendre et al. 2015, Evangelista et al., 2019). The phylogenetic information on Blattodea now seems to be quite robust, and many of the previously controversial relationships have apparently been resolved (Evangelista et al., 2019). How robust are these state-of-the-art phylogenies? And are there any simpler approaches that can support these topologies and be used to solve existing ambiguities in arthropod phylogeny in future analyses? To address these questions, we tested an alternative and relatively easy approach in which we evaluated the potential phylogenetic signals of neuropeptide precursor sequences. Thus, our approach was based on the analysis of a limited set of identified conserved orthologous protein-coding genes and not on the analysis of many unidentified genes. To allow a direct comparison of our own results with those obtained from extensive molecular datasets, we re-evaluated the phylogeny of Blattodea using the comprehensive transcriptome data of the 1KITE initiative (http://www.1kite.org/), which provided unprecedented access to neuropeptide precursor sequences of insects. The taxon sampling in our study was largely consistent with the recent phylogenomic study by Evangelista et al. (2019). Altogether, we examined the transcriptomes of 40 blattodean species and the precursor genes of *Zootermopsis nevadensis* (Terrapon et al, 2014, Veenstra, 2014), and for each of these species the translated amino acid sequences of 17 neuropeptide precursors.

What is special about the evolution of neuropeptide sequences? First, the amino acid sequences of orthologous neuropeptides (orthocopies) are often well conserved. This is particularly true for single-copy neuropeptides which invariably must fit to their respective receptors to fulfill their functions (Derst et al., 2016). Mutations that alter the amino-acid sequences of single-copy neuropeptides must either be accompanied by parallel mutations that maintain the binding properties of the respective receptors or should not alter the steric properties of the peptides to maintain functionality. These requirements result in strong stabilizing selection on amino-acid sequences (Wegener and Gorbashov, 2008). Second, no fewer than 50 neuropeptide genes can be expected for a single insect species (for Polyneoptera see Veenstra, 2014, Liessem et al., 2018). Therewith, neuropeptides are among the most diverse group of intercellular signaling molecules in multi-cellular animal organisms (Metazoa) (Nässel, 1993, Larhammar, 2009). Several

* Corresponding authors.
*E-mail addresses:* mblaese1@uni-koeln.de (M. Bläser), rpredel@uni-koeln.de (R. Predel).

https://doi.org/10.1016/j.ympev.2019.106686
Received 2 July 2019; Received in revised form 9 November 2019; Accepted 13 November 2019
Available online 15 November 2019

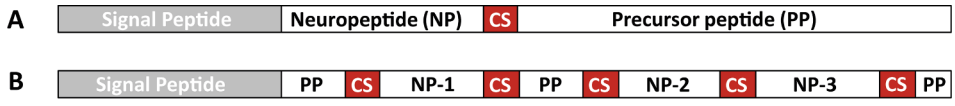



**Fig. 1.** Overview of the functionally different units of neuropeptide precursors. (A) Single-copy neuropeptide precursor. (B) Multiple-copy neuropeptide precursor with three paracopies. Signal peptides direct the precursor proteins to the secretory pathway before being cleaved off. Precursor peptides (PP) may function as recognition sites for e.g. endoproteases that use cleavage sites (CS) to cleave off the neuropeptide (NP) sequences.

neuropeptide precursors even contain multiple neuropeptide sequences (paracopies) (Wegener and Gorbashov, 2008). These paracopies evolve independently and accumulate gene sequence differences over time. Some of the genes encoding multiple neuropeptides express over 20 paracopies (Predel et al., 2004; Larhammar, 2009). Strong sequence conservation, as assumed for single-copy peptides, is less typical of these multiple-copy neuropeptides (Wegener and Gorbashov, 2008). Therefore multiple-copy neuropeptides are particularly useful for reconstructing lower-level phylogenetic relationships. Using identified single-copy and multiple-copy neuropeptides precursors together for phylogenetic purposes may provide information from species level up to the reconstruction of deep phylogenetic relationships.

Sequences of neuropeptides have been used for phylogenetic analyses in several recent studies (e.g. Roth et al., 2009, Predel et al., 2012). In these studies, mature neuropeptides of different species were identified by mass spectrometry. Here, we use transcriptome data to retrieve complete neuropeptide precursor transcripts. This provides additional sequence motifs, namely signal peptides and other precursor peptides (Fig. 1), which may evolve at very different rates and have previously not been in the focus of phylogenetic studies or studies on the evolution of neuropeptide precursors. Our approach provides highly informative data for the reconstruction of phylogenetic relationships in Blattodea and can therefore generally be used as a complementary strategy in insect phylogeny. We also analyzed in detail to what extent the functionally different sequence motifs provide phylogenetic information to recover inter-familial relationships.

## 2. Materials and methods

### 2.1. RNA isolation, transcriptome sequencing and assembly

Most data were obtained from a dataset originally created as part of the 1KITE project (Genbank Umbrella Bioproject ID PRJNA183205); the respective specimens from different Blattodea species were initially preserved in RNAlater. RNA isolation, cDNA preparation, transcriptome sequencing and assembly of raw RNA-Seq reads were carried out as described in Misof et al. (2014). A tblastn search was performed using SOAP assemblies (Luo et al., 2012) version e1 (Misof et al., 2014). For the outgroup species *Perla marginata* (Plecoptera) and *Miomantis brunni* (Mantodea) we generated new transcriptomes using tissue samples from the central nervous system (CNS) of single adult individuals (*Miomantis*) or nymphs (*Perla*) (full species list see Table S1). For both species, the CNS of a single specimen was dissected and cleaned from adjacent tissues such as fat body or trachea. During preparation, the tissue was covered with chilled insect saline of the following composition (in mM): NaCl 128, KCl 2.7, $CaCl_2$ 2, and $NaHCO_3$ 1.2; pH 7.25. Dissected tissue was transferred into 1.5 mL reaction tubes filled with RNAlater (Thermo Fisher Scientific, Waltham, MA, USA) to minimize RNA degradation. Subsequent RNA isolation, cDNA preparation and Illumina Next Generation Sequencing were carried out by BGI (Beijing Genomics Institute, New Territories, HongKong, China) as described in Ragionieri et al. (2017). Transcriptome sequencing and *de novo* assembly using Trinity (v2.2.0) (Grabherr et al., 2011) also followed existing protocols (Ragionieri et al., 2017, Liessem et al., 2018).

### 2.2. Orthology assessment and alignment of neuropeptide precursor sequences

We used amino-acid sequences from neuropeptide precursor sequences of *Carausius morosus* (Phasmatodea; Liessem et al., 2018) as blast queries to search for candidate sequences in transcriptomes. Assembled transcripts were analyzed with the tblastn algorithms provided by NCBI (Camacho et al., 2009, https://blast.ncbi.nlm.nih.gov/Blast.cgi). Identified candidate nucleotide precursor gene sequences were translated into amino-acid sequences using the ExPASy Translate tool (Artimo et al. 2012; http://web.expasy.org/translate/) with the standard genetic code. Orthology of precursor sequences was confirmed by the position of conserved neuropeptide sequences with their respective cleavage sites in the precursor. The number of neuropeptide precursors in the 1KITE dataset was insufficient for *Deropeltis erythrocephala*, *Gyna lurida*, *Diploptera punctata* and *Nyctibora* sp.. For this reason, these species have not been included in our phylogenetic analysis. Orthologous neuropeptide precursors of the remaining 40 species were aligned using the MAFFT-L-INS-i algorithm (Katoh and Standley, 2013) (dvtditr (aa) Version 7.299b alg = A, model = BLOSUM62, 1.53, −0.00, −0.00, noshift, amax = 0.0). For *Z. nevadensis* (Isoptera), annotated precursor sequences (Veenstra, 2014) were included in the alignments after their annotations were manually checked. Incomplete neuropeptide precursors containing multiple paracopies (CAPA, FXPRLamides, tachykinin-related peptides) were partially complemented with nucleotide sequences contained in the raw data (after translation into amino-acid sequences); determined amino-acid sequence motifs from sister taxa were used as queries in these cases. From these raw data, only those sequences were used whose complete sequence motif matched the query sequences. In addition, amino-acid sequences of several previously published mature neuropeptides (e.g. Roth et al., 2009) have been added to the alignments of these species if the transcriptome data did not reveal the complete neuropeptide precursor sequence. Orthology of these peptides was verified in addition to the always distinct sequences by the specific part of the CNS used for mass spectrometric analyses. Products of different neuropeptide genes accumulate in different parts of the nervous system, which then facilitates the assessment of orthology (Predel, 2001). Prediction of signal peptides from amino-acid sequences was carried out using the SignalP 4.1 Server (Petersen et al., 2011; www.cbs.dtu.dk/services/SignalP/) with D-cutoff values set as default and Eukaryotes selected as organism group. Alignments generated with the MAFFT-L-INS-i algorithm were then manually checked for misaligned sequences using N-termini of signal peptides and conserved amino-acid residues (cleavage signals, Cys as target for disulfide bridges) as anchor points. Incompletely translated transcripts of neuropeptide precursors were labeled with question marks at the respective amino-acid positions. Individual amino-acid alignments of each group of orthologous neuropeptide precursors were concatenated in BioEdit 7.2.5 (Hall, 1999). For additional analyses of the functionally different units, the amino-acid alignments of the precursor proteins (preprohormones) were manually partitioned into alignments containing only the signal peptide sequences, the prohormone sequences, sequences of putative bioactive neuropeptides, and the remaining precursor peptides. All amino-acid matrices are deposited at Mendeley Data (doi: https://doi.org/10.17632/cdwzkcdcz3.1). The Average evolutionary divergence over all sequence pairs (Table 1) was calculated using the JTT matrix-based model (Jones et al., 1992) in MEGA6 (Tamura et al., 2013).

**Table 1**
Neuropeptide precursors used in our analyses and their coverage in 40 blattodean species. In addition, the average evolutionary divergence over all sequence pairs is given for each neuropeptide precursor.

| | Complete | Partial | Missing | Length [aa Positions] | Average evolutionary divergence |
|---|---|---|---|---|---|
| *Single-copy precursor* | | | | | |
| Allatostatin C | 22 | 14 | 4 | 94–96 | 0.1086 |
| Allatotropin | 31 | 8 | 1 | 122–128 | 0.2294 |
| Crustacean cardioactive peptide | 28 | 11 | 1 | 154–174 | 0.2682 |
| CCHamide-2 | 23 | 15 | 2 | 123–142 | 0.4489 |
| CNMamide-A | 33 | 7 | 0 | 126–143 | 0.6630 |
| Corazonin | 23 | 15 | 2 | 118–140 | 0.4689 |
| CRF-like Diuretic hormone | 20 | 16 | 4 | 189–203 | 0.2784 |
| Diuretic hormone | 35 | 5 | 0 | 115–117 | 0.1446 |
| Elevenin | 27 | 12 | 1 | 120–145 | 0.5117 |
| Myosuppressin | 30 | 8 | 2 | 92–101 | 0.2632 |
| Neuropeptide F 1 | 38 | 2 | 0 | 85–91 | 0.2710 |
| Neuropeptide F 2 | 24 | 13 | 3 | 116–128 | 0.3624 |
| Proctolin | 33 | 6 | 1 | 80–93 | 0.2542 |
| Short Neuropeptide F | 31 | 9 | 0 | 100–105 | 0.1798 |
| *Multiple-copy precursor* | | | | | |
| CAPA | 9 | 22 | 9 | 191–239 | 0.7698 |
| FXPRLamide/Pyrokinin | 22 | 17 | 1 | 183–196 | 0.4654 |
| Tachykinin-related Peptide | 15 | 22 | 3 | 317–405 | 0.3861 |

### 2.3. Phylogenetic analysis

FASTA files of aligned amino acid sequences were converted into the relaxed PHYLIP format using Geneious 10.0.8 (http://www.geneious.com) and Nexus format using PGDSpider version 2 (Lischer and Excoffier, 2012). After defining the N-terminal of each neuropeptide precursor as starting partition, best-fit partitioning schemes and substitution models for subsequent phylogenetic analyses were predicted with PartitionFinder2 (Lanfear et al., 2016), implementing the greedy algorithm (Lanfear et al., 2012), in combination with RaxML (Stamatakis, 2014). Branch lengths were set as unlinked and the criterion for model selection set to AICc. Bayesian analyses were run with MrBayes, with two runs, using four chains and a sample frequency of 1000 until convergence was achieved (PSFRvalue between 1.00 and 1.02) with a minimum of 5,000,000 generations (Ronquist et al., 2012). Trees were visualized using FigTree 1.4.2 (http://tree.bio.ed.ac.uk/) and Adobe Illustrator CS6. In addition, two maximum likelihood analyses were carried out using IQ-TREE 1.6 (Nguyen et al., 2015). The mtMet + F + I + G4 model was determined with the ModelFinder option of IQ-TREE (Kalyaanamoorthy et al., 2017). Branch support (Hoang et al. 2017) and SH-like approximate likelihood ratio test (Guindon et al., 2010) was assessed using 1000 ultra-fast bootstrap replicates for each. Likewise, a second maximum likelihood analysis was run using a site-heterogeneous mixture model (CAT20) in IQ-Tree implementing the same conditions.

### 2.4. Sequence logo generation

Sequence logos of orthologous and aligned neuropeptide precursor sequences were generated using the tool WebLogo version 2.8.2 (Crooks et al., 2004; http://weblogo.berkeley.edu/logo.cgi). Each stack represents one position in the multiple sequence alignment. The overall height of a stack indicates the sequence conservation at this amino-acid position; the height of letters within the stack indicates the relative frequency of each amino acid at that position. For the color scheme of amino-acid residues, the default settings were selected.

## 3. Results and discussion

### 3.1. Higher-level Blattodea phylogeny reconstructed with concatenated neuropeptide precursor sequences

A dataset of 45 neuropeptide precursors from the stick insect *Carausius morosus*, not including protein hormone precursors, was used to search for orthologous sequences in the Blattodea transcriptome assemblies of the 1KITE initiative. Transcriptome data from 40 blattodean species covering the eight major lineages of cockroaches (see Beccaloni and Eggleton, 2013) and five families of Isoptera were of sufficient quality to obtain multiple neuropeptide precursor sequences. A total of 17 neuropeptide precursors was identified with orthologs present in most blattodean species that were analyzed (Table 1). In addition, precursor sequences of *Z. nevadensis* have been included with orthology to the above-mentioned 17 precursor sequences of Blattodea. These sequences were derived from the genome of *Z. nevadensis* (Terrapon et al, 2014, Veenstra, 2014). The average precursor length was 146 amino acids with the shortest precursor containing 80 amino acids (*Ectobius sylvestris* Proctolin precursor) and the longest of these precursors containing 405 amino acids (*Eucorydia yasumatsui* tachykinin-related peptide precursor). The concatenated matrix of the 17 precursors from all species contained 3157 aligned amino-acid positions of which 1453 were parsimony informative (46%) and 18% missing data. Likely, two reasons were responsible for missing data in precursor sequences. First, neuropeptide gene expression is restricted to a limited number of cells in the nervous system and midgut. Therefore, whole-body transcriptomes such as those generated within the 1KITE initiative typically provide less information than transcriptomes of e. g. the central nervous system. Second, the Blattodea sequences of the 1KITE initiative are from 2014 or earlier. Improved sample preparation and new sequencing efforts would result in much better coverage and a larger number of precursor sequences that can be included in future analyses. The 17 neuropeptide precursors selected here for our analyses include 14 precursors coding for single-copy neuropeptides and three precursors coding for multiple-copy neuropeptides (see Table 1). Orthologous sequences of these precursors were subsequently identified for three outgroup species: *Thermobia domestica* (Zygentoma; Derst et al., 2016), *P. marginata* (transcriptome CNS, this study), and *M. brunni* (transcriptome CNS, this study).

Our Bayesian phylogenetic inference of the concatenated neuropeptide precursor dataset (Fig. 2; Supplementary Fig. 1A and B for ML analysis) recovered Blaberoidea as monophyletic. Within Blaberoidea, Blaberidae is monophyletic. Ectobiidae, as traditionally defined (von Wattenwyl, 1865, Beccaloni and Eggleton, 2011), is not a monophyletic taxon (see also Inward et al., 2007; Djernæs et al., 2015; Legendre et al., 2015). In this context, downgrading of Blaberidae to a subfamily (= Blaberinae) of Ectobiidae is likely to be a simple solution that avoids further changes to the existing general classification. Particularly interesting within Blaberoidea is the position of the genus *Ectobius* as sister to the rest (see Fig. 2). This phylogenetic relationship was first recovered by Wang et al. (2017). Species of the subfamily Blattellinae were determined as sister to Blaberidae in our analysis. *Anallecta methanoides*, which current taxonomic position is within Blattellinae was recovered as sister to a clade consisting of several taxa of Pseudophyllodromiinae (see also Bourguignon et al., 2018). *Anallacta* + Pseudophyllodromiinae represent the sister clade to the remaining Blattellinae + Blaberidae. The Blattoidea represented the sister to Corydioidea and encompassed here the taxa Tryonicidae + Blattidae and Lamproblattidae + [Cryptocercidae + Isoptera]. A close relationship of Blattidae and Tryonicidae was proposed several times (e.g. McKittrick, 1965, Murienne, 2009, Legendre et al., 2015, Evangelista et al., 2019) but recent phylogenetic analyses placed Tryonicidae alternatively as sister to Cryptocercidae + Isoptera (Djernæs et al., 2015) or as sister to Lamproblattidae (Wang et al., 2017). The close relationship which was recovered in our study between Cryptocercidae

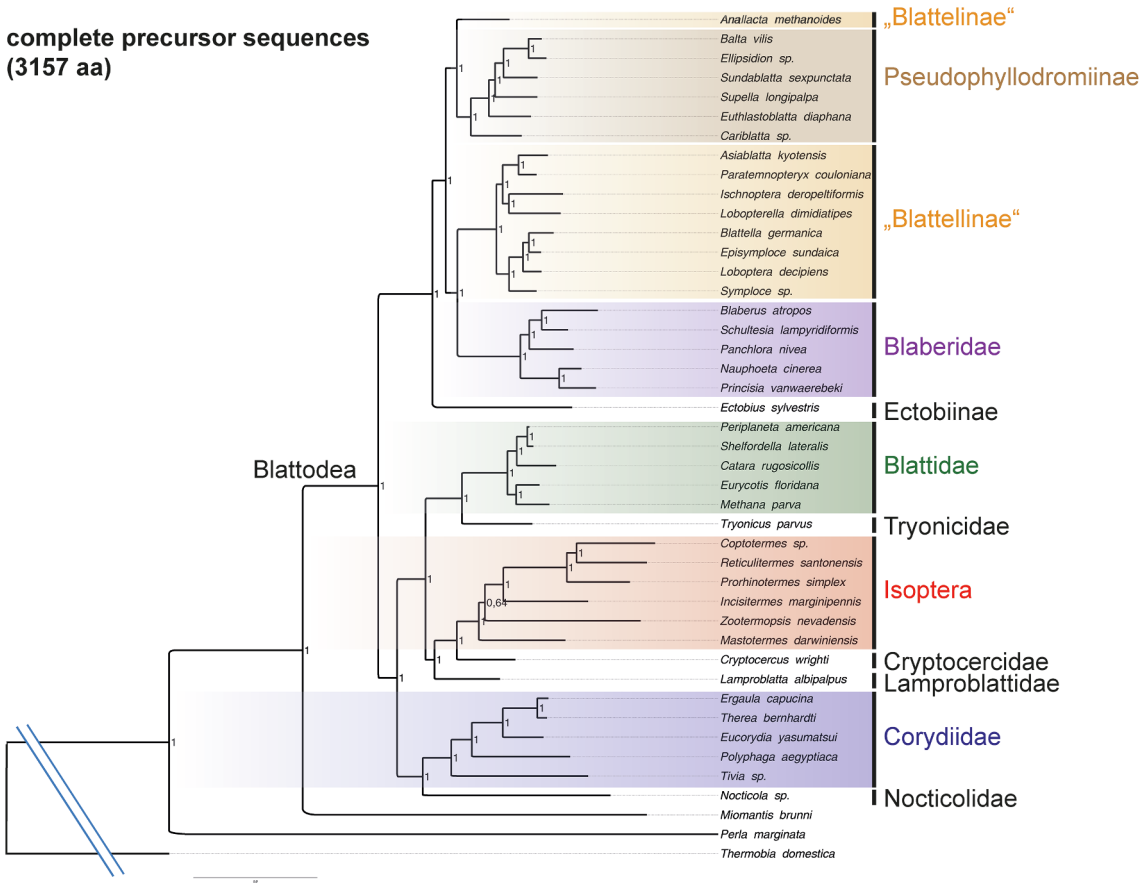



**Fig. 2.** Bayesian Posterior Probability (BPP) phylogenetic analysis of a concatenated dataset of 17 neuropeptide precursors for 41 species of Blattodea and three hexapod outgroups (model = GTR + I + G + F). Nodal support values (BPP values) are mostly with maximal support; only the position of *Z. nevadensis* was not fully solved with our data. Traditionally accepted families are indicated by colors. Instead of the paraphyletic family Ectobiidae, subfamilies are given. The paraphyletic subfamily Blattelinae is in quotation marks.

and Isoptera (Tutricablattae in Evangelista et al. 2019) has been established already in earlier studies (e.g. Klass, 1995, Lo et al., 2003, Inward et al., 2007). The position of *Z. nevadensis* was not fully solved with our data, and this remained the only ambiguity in our analysis. In the study by Evangelista et al. (2019) using the same taxa, the placement of *Z. nevadensis* was identical to ours and was also not very well supported. Most likely, the taxon sampling was not adequate for Isoptera, as some critical taxa (e. g. *Hodotermes*, Termitidae) were not included.

Overall, our analysis revealed a topology with maximum support (posterior probability, pp = 1; except for *Z. nevadensis*) identical to the topology proposed by Evangelista et al. (2019), excluding the four species that were not considered in our analysis due to poor coverage of neuropeptide precursors. This confirms on the one hand the results of Evangelista et al. (2019) and on the other hand the excellent applicability of neuropeptide precursor sequences for phylogenetic questions.

### *3.2. Which parts of the neuropeptide precursors contain the phylogenetic informative sequences?*

The concatenated dataset with 17 neuropeptide precursors was sufficient to provide a phylogenetic tree of Blattodea whose nodes showed maximal support, except for the placement of the termite *Z. nevadensis*. A neuropeptide precursor consists of sequence motifs with functional differentiation, which likely results in a different degree of sequence conservation. Hence, different parts of the precursor sequence could provide phylogenetic information that is more appropriate for closely-related or higher ranking taxa. The dataset used in our study provided an opportunity to investigate that topic in detail.

Neuropeptide precursors of Metazoa generally consist of a signal peptide sequence at the N-terminus, predictable neuropeptide sequences which are flanked by mono- or dibasic cleavage signals, and remaining precursor sequences which function e. g. as recognition sites for endoproteases (see Fig. 1). Particularly the genuine neuropeptide sequences that co-evolve with their peptide-specific receptors are potentially well conserved. This conservation and the possibility to elucidate peptides directly from nerve tissue samples by MALDI-TOF mass spectrometry without purification steps made neuropeptides attractive for phylogenetic analyses even before the availability of comprehensive genomic and transcriptomic data. First comprehensive tests with such a peptidomic approach were performed on cockroaches using a concatenated dataset of less than 100 amino acids (CAPA-peptides, adipokinetic hormones and sulfakinins; Roth et al., 2009). Although this was not sufficient to solve the higher-ranking clades with maximal support, more closely related taxa were always grouped together up to family level. Incorporation of additional neuropeptide orthocopies to solve the phylogenetic relationships within Mantophasmatodea (Insecta: Polyneoptera) already significantly improved the node support (Predel et al., 2012). In the study at hand, concatenated sequences of the putative neuropeptides from 17 precursors of each of the 41 Blattodea species (see Table 1) resulted in about 576 amino acids per species. These sequences, which represent about 18% of the complete precursor sequences, yielded already a well resolved topology of Blattodea, although not all clades were supported with maximal support (Fig. 3). A consistent and well-supported topology was also retrieved with the remaining precursor sequences (without predicted neuropeptide sequences and signal peptides), but with *Miomantis* nesting within Blattodea (Supplementary Fig. 2). Partitioning this dataset in three

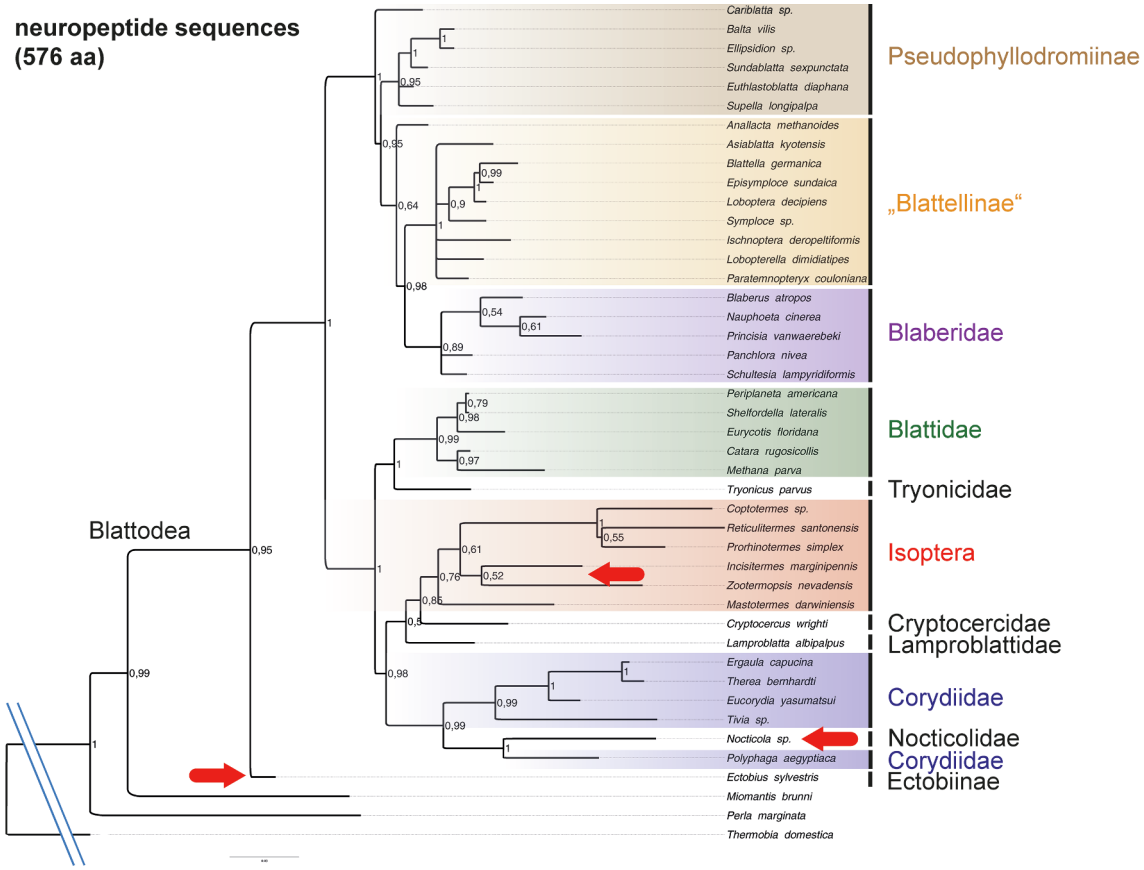



**Fig. 3.** Bayesian Posterior Probability (BPP) phylogenetic analysis of a concatenated dataset of predicted neuropeptide sequences from 17 neuropeptide precursors for 41 species of Blattodea and three hexapod outgroups (model = DAYHOFF + G) with BPP values. Note the position of *Ectobius sylvestris* as sister group to all remaining Blattodea and the placement of Nocticolidae inside of Corydiidae. The position of *Z. nevadensis* differs from the analysis of the complete precursor (see Fig. 2). Otherwise all major lineages could be recovered. Traditionally accepted families are indicated by colors. Instead of the paraphyletic family Ectobiidae, subfamilies are given. The paraphyletic subfamily Blattelinae is in quotation marks.

units of 576, 576 and 936 amino acids yielded cladograms with poorly supported topologies (Supplementary Fig. 3). Thus, the genuine neuropeptide sequences indeed contained more phylogenetic information per sequence unit. The higher number of amino acids in the remaining precursor peptide sequences, however, compensated for the lower phylogenetic information. A very similar topology with remarkably good node support was obtained using the short sequences of signal peptides (Fig. 4). These sequences direct the precursor proteins into the endoplasmic reticulum and, hence, to the secretory pathway. The signal peptides of the different precursors from a single species fulfill identical functions. Therefore, it was very surprising that, at the ordinal level, the concatenated sequences of the 17 signal peptides (about 494 amino acids per species) recovered the blattodean phylogeny with relatively well supported nodes. Only the position of *Miomantis* was questionable, as it was nested within Blattodea, but with little nodal support. As a consequence out of this, the three functionally different sequence parts of neuropeptide precursors can be combined to optimize the phylogenetic analysis of taxa at the ordinal level. This is true, although the sequence conservation may be remarkably different for the different units of precursor proteins (Fig. 5).

### 3.3. Does the degree of phylogenetic information differ between the neuropeptide precursors?

The selection of precursor sequences in our study was based solely on the availability of the respective sequences in the transcriptomes of the different species of Blattodea. Therefore, precursor sequences not considered here may contain similar or even more phylogenetic information and could be included if the respective sequences are available. Within the 17 precursors analyzed, the degree of sequence conservation and hence, the phylogenetic information is very different. On the one hand, sequences of several precursors (e.g. MS, AST-C, DH-31, Proctolin, sNPF) are highly conserved, and the phylogenetic information at the ordinal level is rather low in these cases. These precursors, however, can provide information to clarify relationships at the inter-ordinal level. On the other hand, a number of precursors contain sequence motifs that caused ambiguities in the alignments. Such ambiguities have mostly been found in multiple-copy precursors which code for a variable number of paracopies. Precursors that cause ambiguities cannot be meaningfully aligned at the inter-ordinal level and were not taken into account in our analyses. At the intra-ordinal level, however, the remaining multiple-copy precursors apparently contain very suitable sequence information. In our study, we included three multiple-copy precursors (tachykinin-related peptides, CAPA, FXPRLa) and the phylogenetic tree obtained from these sequences (Supplementary Fig. 4) recovered the same topology as the tree obtained from the 14 single-copy peptide precursors (Supplementary Fig. 5). Only the support of the higher-level nodes decreased considerably.

## 4. Conclusions

Although only one third of the neuropeptide precursors generally present in insects could be considered due to incomplete data sets, these sequences already provided a very well-supported topology of Blattodea; including the placement of *Lamproblatta*, *Ectobius*, *Tryonicus*,

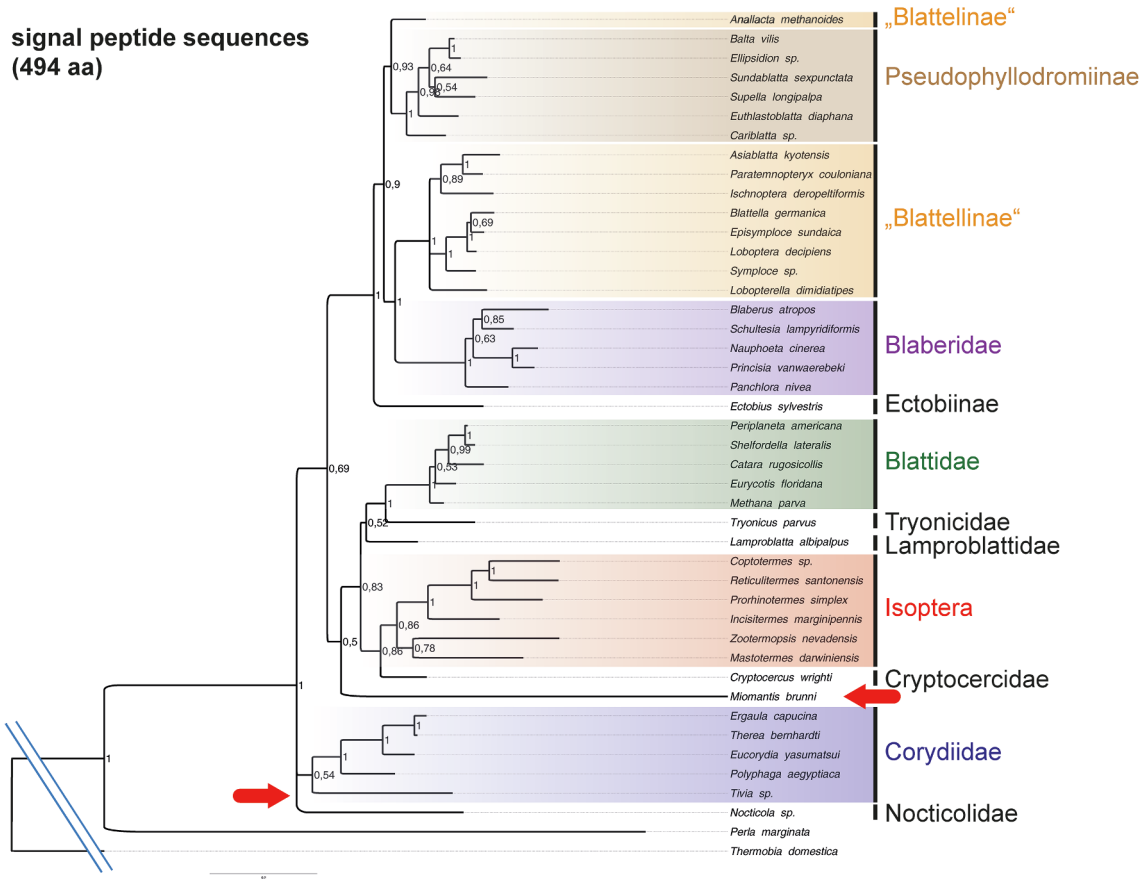



**Fig. 4.** Bayesian Posterior Probability (BPP) phylogenetic analysis of a concatenated dataset of signal peptide sequences from 17 neuropeptide precursors for 41 species of Blattodea and three hexapod outgroups (model = GTR + G + F) with BPP values. Note the placement of *Miomantis brunni* inside of Blattodea. The position of Corydiidae and Nocticolidae is not resolved with this dataset. Traditionally accepted families are indicated by colors. Instead of the paraphyletic family Ectobiidae, subfamilies are given. The paraphyletic subfamily Blattelinae is in quotation marks.

and *Nocticola*. In fact, all nodes except for the placement of the termite *Z. nevadensis* were confirmed with maximum support. The results are congruent with a recent nuclear phylogenomic study that used a largely identical taxon sampling (Evangelista et al., 2019). Since none of the precursor sequences used in our work was used by Evangelista et al. (2019), the topology we obtained independently confirms the previous data. At the same time, the results show the very good applicability of neuropeptide precursors for phylogenetic questions. One of the reasons for this is the conservation of sequences that co-evolve with the corresponding receptors. The proof that phylogenetic questions can be answered effectively and meaningfully with a few conserved and easily identifiable neuropeptide precursor sequences was a central point of our study. The results are based on only 17 neuropeptide coding genes in contrast to 2370 protein-coding nuclear single-copy genes used by Evangelista et al. (2019). As confirmed by separate analyses of the functionally distinct units of the neuropeptide precursors, all units contained phylogenetic informative data at the ordinal level and their analysis did not reveal contradictory topologies. While the applicability of the actual neuropeptide sequences for phylogenetic questions was already known, the presented work shows an almost equally good applicability of the remaining sequences of the corresponding precursors. Especially surprising for us is the topology obtained using the short signal peptides. The signal peptides of the different precursors all have basically the same function. It is possible that the phylogenetic information contained therein is due to co-evolution with channel proteins of the endoplasmic reticulum, which the signal peptides have to recognize in order to enter the lumen of the ER.

An advantage of using neuropeptide precursor sequences for phylogenetic analysis is the opportunity to review the alignments manually. Mainly using known N-terminals of signal peptides (Met), position and cleavage sites of orthocopies and positions of Cys (half of disulfide bridges) as anchor points. Automatically generated alignments contained a number of obvious misalignments of highly conserved amino acids positions. In some precursors, slightly different transcripts occurred, and, BLAST searches in the assemblies selected different transcripts for different species, depending on their relative abundances. Such false positives, which are easily detectable by manually checking the alignments, can create artificial relationships. Nevertheless, the phylogenetic relationships in our analyses remained unchanged when the concatenated dataset was used with the not manually curated alignments (not shown). This indicates saturation of node support in the concatenated dataset. Phylogenetic analyses of individual precursor alignments (not shown) confirmed that the various neuropeptide precursors have different phylogenetic information and therefore provided different contributions to the topology presented. While some of the single-copy precursors can only be used to differentiate higher ranking clades, aligning of multiple-copy precursors can sometimes be problematic because of varying number of paracopies. The latter precursors are, however, well suited to clarify relationships of closely related taxa; e.g. at the intra-familial level.

With the methodological approach presented here also ambiguous sister-group relationships in other arthropod groups can be analyzed. Future studies will show to what extent precursor sequences can be used to evaluate phylogenetic relationships even at the inter-ordinal level. In this case, alignment ambiguities will definitively reduce the number of potential precursor genes. This problem can be solved by considering additional precursors. In the end, the quality of the transcriptomes is essential.


**Fig. 5.** Comparison of sequence conservation of the functionally distinct units of a neuropeptide precursor; illustrated by allatotropin (AT) precursors of Blattodea. The sequence logo shows the variation for each AA position across 34 orthologous AT precursors. The overall height of a stack indicates the sequence conservation at that position, while the height of letters within the stack indicates the relative frequency of each amino acid at a particular position. Gaps in the sequences are depicted with an "X". The AT precursor belongs to the single-copy precursors, containing highly conserved neuropeptide sequences. Note the different degrees of sequence conservation in the different units of the precursor. (A) Signal peptide sequence with a high degree of sequence variation. (B) The part of the AT precursor C-terminal of the signal peptide, containing the AT neuropeptide sequence (red box). This sequence is particularly well conserved. (C) N-terminal part of the precursor peptide showing a high degree of sequence variation. (D) C-terminal part of the precursor peptide showing a highly conserved sequence. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements

This work was supported by the German Research Foundation (PR766/11-1). We thank Jennifer Schmitz, Rene Köhler and Lapo Ragionieri (University of Cologne, Germany) for dissection of specimens and help with transcriptome generation. We are grateful for the help of Alexander Donath (ZFMK Bonn, Germany) and Karen Meusemann (University of Freiburg, Germany) for providing early access to the 1KITE transcriptome assemblies and raw data.

## Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ympev.2019.106686.

## References

Artimo, P., Jonnalagedda, M., Arnold, K., Baratin, D., Csardi, G., de Castro, E., Duvaud, S., Flegel, V., Fortier, A., Gasteiger, E., Grosdidier, A., Hernandez, C., Ioannidis, V., Kuznetsov, D., Liechti, R., Moretti, S., Mostaguir, K., Redaschi, N., Rossier, G., Xenarios, I., Stockinger, H., 2012. ExPASy: SIB bioinformatics resource portal. Nucleic Acids Res. 40 (W1), 597–603. https://doi.org/10.1093/nar/gks400.

Beccaloni, G.W., Eggleton, P., 2011. Order Blattodea Brunner von Wattenwyl, 1882. In: Zhang, Z.Q. (Ed.). Animal biodiversity: An outline of higher-level classification and survey of taxonomic richness. Zootaxa. 13148, pp. 199.

Beccaloni, G., Eggleton, P., 2013. Order Blattodea. In: Zhang, Z.Q. (Ed.) Animal Biodiversity: An outline of higher-level classification and survey of taxonomic richness (Addenda 2013). Zootaxa. 3703, pp. 46–48.

Bourguignon, T., Tang, Q., Ho, S.Y.W., Juna, F., Wang, Z., Arab, D.A., Cameron, S.L., Walker, J., Rentz, D., Evans, T.A., Lo, N., 2018. Transoceanic dispersal and plate tectonics shaped global cockroach distributions: evidence from mitochondrial phylogenomics. Mol. Biol. Evol. 35 (4), 970–983. https://doi.org/10.1093/molbev/msy013.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., Madden, T.L., 2009. BLAST+: architecture and applications. BMC Bioinf. 10, 421. https://doi.org/10.1186/1471-2105-10-421.

Crooks, G.E., Hon, G., Chandonia, J.M., Brenner, S.E., 2004. WebLogo: a sequence logo generator. Genome Res. 14, 1188–1190. https://doi.org/10.1101/gr.849004.

Derst, C., Dircksen, H., Meusemann, K., Zhou, X., Liu, S., Predel, R., 2016. Evolution of neuropeptides in non-pterygote hexapods. BMC Evol. Biol. 16, 51. https://doi.org/10.1186/s12862-016-0621-4.

Djernæs, M., Klass, K.-D., Picker, M.D., Damgaard, J., 2012. Phylogeny of cockroaches (Insecta, Dictyoptera, Blattodea), with placement of aberrant taxa and exploration of out-group sampling. Syst. Entomol. 37, 65–83. https://doi.org/10.1111/j.1365-3113.2011.00598.x.

Djernæs, M., Klass, K.D., Eggleton, P., 2015. Identifying possible sister groups of Cryptocercidae+Isoptera: a combined molecular and morphological phylogeny of Dictyoptera. Mol. Phylogenet. Evol. 84, 284–303. https://doi.org/10.1016/j.ympev.2014.08.019.

Evangelista, D.A., Wipfler, B., Béthoux, O., Donath, A., Fujita, M., Kohli, M.K., Legendre, F., Liu, S., Machida, R., Misof, B., Peters, R.S., Podsiadlowski, L., Rust, J., Schuette, K., Tollenaar, W., Ware, J.L., Wappler, T., Zhou, X., Meusemann, K., Simon, S., 2019. An integrative phylogenomic approach illuminates the evolutionary history of cockroaches and termites (Blattodea). Proc. R. Soc. B. 286 (1895), 20182076. https://doi.org/10.1098/rspb.2018.2076.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B.W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., Regev, A., 2011. Full-length transcriptome assembly from RNA-seq data without a reference genome. Nat. Biotechnol. 29, 644–652. https://doi.org/10.1038/nbt.1883.

Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O., 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. 59 (3), 307–321. https://doi.org/10.1093/sysbio/syq010.

Hall, T.A., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp. Ser. 41 (41), 95–98. https://doi.org/10.14601/Phytopathol_Mediterr-14998u1.29.

Hoang, D.T., Chernomor, O., von Haeseler, A., Minh, B.Q., Vinh, L.S., 2017. UFBoot2: Improving the ultrafast bootstrap approximation. Mol. Biol. Evol. 35 (2), 518–522. https://doi.org/10.1093/molbev/msx281.

Inward, D., Beccaloni, G., Eggleton, P., 2007. Death of an order: a comprehensive

molecular phylogenetic study confirms that termites are eusocial cockroaches. Biol. Lett. 3, 331–335.
Jones, D.T., Taylor, W.R., Thornton, J.M., 1992. The rapid generation of mutation data matrices from protein sequences. Comput. Appl. Biosci. 8, 275–282. https://doi.org/10.1093/bioinformatics/8.3.275.
Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A., Jermiin, L.S., 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat. Methods 14, 587–589. https://doi.org/10.1038/nmeth.4285.
Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30, 772–780. https://doi.org/10.1093/molbev/mst010.
Klass, K.D., 1995. Die Phylogenie der Dictyoptera, first ed. Cuvillier, Göttingen, Germany.
Lanfear, R., Calcott, B., Ho, S.Y., Guindon, S., 2012. PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. Mol. Biol. Evol. 29 (6), 1695–1701. https://doi.org/10.1093/molbev/mss020.
Lanfear, R., Frandsen, P.B., Wright, A.M., Senfeld, T., Calcott, B., 2016. PartitionFinder 2: new methods for selecting partitioned models of evolution formolecular and morphological phylogenetic analyses. Mol. Biol. Evol. 34 (3), 772–773. https://doi.org/10.1093/molbev/msw260.
Larhammar, D., 2009. Neuropeptides phylogeny and evolution. In: Squire, L.R. (Ed.), Encyclopedia of Neuroscience. Academic Press, Oxford, Oxford, pp. 873–879.
Legendre, F., Nel, A., Svenson, G.J., Robillard, T., Pellens, R., Grandcolas, P., 2015. Phylogeny of Dictyoptera: dating the origin of cockroaches, praying mantises and termites with molecular data and controlled fossil evidence. PLoS ONE 10, e0130127. https://doi.org/10.1371/journal.pone.0130127.
Liessem, S., Ragionieri, L., Neupert, S., Büschges, A., Predel, R., 2018. Transcriptomic and neuropeptidomic analysis of the stick insect, *Carausius morosus*. J. Proteome Res. 17 (6), 2192–2204. https://doi.org/10.1021/acs.jproteome.8b00155.
Lischer, H.E.L., Excoffier, L., 2012. PGDSpider: An automated data conversion tool for connecting population genetics and genomics programs. Bioinformatics 28, 298–299. https://doi.org/10.1093/bioinformatics/btr642.
Lo, N., Bandi, C., Watanabe, H., Nalepa, C., Beninati, T., 2003. Evidence for cocladogenesis between diverse dictyopteran lineages and their intracellular endosymbionts. Mol. Biol. Evol. 20, 907–913. https://doi.org/10.1093/molbev/msg097.
Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., He, G., Chen, Y., Pan, Q., Liu, Y., Tang, J., Wu, G., Zhang, H., Shi, Y., Liu, Y., Yu, C., Wang, B., Lu, Y., Han, C., Cheung, D.W., Yiu, S.M., Peng, S., Xiaoqian, Z., Liu, G., Liao, X., Li, Y., Yang, H., Wang, J., Lam, T.W., Wang, J., 2012. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. GigaScience 1 (1), 18. https://doi.org/10.1186/2047-217X-1-18.
McKittrick, F.A., 1965. A contribution to the understanding of cockroach-termite affinities. Ann. Entomol. Soc. Am. 58, 18–22. https://doi.org/10.1093/aesa/58.1.18.
Misof, B., Liu, S., Meusemann, K., Peters, R.S., Donath, A., Mayer, C., Frandsen, P.B., Ware, J., Flouri, T., Beutel, R.G., Niehuis, O., Petersen, M., Izquierdo-Carrasco, F., Wappler, T., Rust, J., Aberer, A.J., Aspöck, U., Aspöck, H., Bartel, D., Blanke, A., Berger, S., Böhm, A., Buckley, T.R., Calcott, B., Chen, J., Friedrich, F., Fukui, M., Fujita, M., Greve, C., Grobe, P., Gu, S., Huang, Y., Jermiin, L.S., Kawahara, A.Y., Krogmann, L., Kubiak, M., Lanfear, R., Letsch, H., Li, Y., Li, Z., Li, J., Lu, H., Machida, R., Mashimo, Y., Kapli, P., McKenna, D.D., Meng, G., Nakagaki, Y., Navarrete-Heredia, J.L., Ott, M., Ou, Y., Pass, G., Podsiadlowski, L., Pohl, H., von Reumont, B.M., Schütte, K., Sekiya, K., Shimizu, S., Slipinski, A., Stamatakis, A., Song, W., Su, X., Szucsich, N.U., Tan, M., Tan, X., Tang, M., Tang, J., Timelthaler, G., Tomizuka, S., Trautwein, M., Tong, X., Uchifune, T., Walzl, M.G., Wiegmann, B.M., Wilbrandt, J., Wipfler, B., Wong, T.K.F., Wu, Q., Wu, G., Xie, Y., Yang, S., Yang, Q., Yeates, D.K., Yoshizawa, K., Zhang, Q., Zhang, R., Zhang, W., Zhang, Y., Zhao, J., Zhou, C., Zhou, L., Ziesmann, T., Zou, S., Li, Y., Xu, X., Zhang, Y., Yang, H., Wang, J., Wang, J., Kjer, K.M., Zhou, X., 2014. Phylogenomics resolves the timing and pattern of insect evolution. Science 346, 763–767. https://doi.org/10.1126/science.1257570.
Murienne, J., 2009. Molecular data confirm family status for the *Tryonicus-Lauraesilpha* group (Insecta: Blattodea: Tryonicidae). Organ. Divers. Evol. 9, 44–51. https://doi.org/10.1016/j.ode.2008.10.005.
Nässel, D.R., 1993. Neuropeptides in the insect brain: a review. Cell Tissue Res. 273 (1), 1–29. https://doi.org/10.1007/BF00304608.
Nguyen, L.T., Schmidt, H.A., von Haeseler, A., Minh, B.Q., 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-mikelihood phylogenies. Mol. Biol. Evol. 32 (1), 268–274. https://doi.org/10.1093/molbev/msu300.
Petersen, T.N., Brunak, S., von Heijne, G., Nielsen, H., 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat. Methods 8, 785–786. https://doi.org/10.1038/nmeth.1701.
Predel, R., 2001. Peptidergic neurohemal system of an insect: mass spectrometric morphology. J. Comp. Neurol. 436 (3), 363–375. https://doi.org/10.1002/cne.1073.
Predel, R., Wegener, C., Russell, W.K., Tichy, S.E., Russell, D.H., Nachman, R.J., 2004. Peptidomics of CNS-associated neurohemal systems of adult *Drosophila melanogaster*: a mass spectrometric survey of peptides from individual flies. J. Comp. Neurol. 474 (3), 379–392. https://doi.org/10.1002/cne.20145.
Predel, R., Neupert, S., Huetteroth, W., Kahnt, J., Waidelich, D., Roth, S., 2012. Peptidomics-based phylogeny and biogeography of Mantophasmatodea (Hexapoda). Syst. Biol. 61 (4), 609–629. https://doi.org/10.1093/sysbio/sys003.
Ragionieri, L., Özbagci, B., Neupert, S., Salts, Y., Davidovitch, M., Altstein, M., Predel, R., 2017. Identification of mature peptides from *pban* and *capa* genes of the moths *Heliothis peltigera* and *Spodoptera littoralis*. Peptides 94, 1–9. https://doi.org/10.1016/j.peptides.2017.05.004.
Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M.A., Huelsenbeck, J.P., 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst. Biol. 61 (3), 539–542. https://doi.org/10.1093/sysbio/sys029.
Roth, S., Fromm, B., Gäde, G., Predel, R., 2009. A proteomic approach for studying insect phylogeny: CAPA peptides of ancient insect taxa (Dictyoptera, Blattoptera) as a test case. BMC Evol. Biol. 9 (1), 50. https://doi.org/10.1186/1471-2148-9-50.
Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30 (9), 1312–1313. https://doi.org/10.1093/bioinformatics/btu033.
Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S., 2013. MEGA6: molecular evolutionary genetics Analysis version 6.0. Mol. Biol. Evol. 30, 2725–2729. https://doi.org/10.1093/molbev/mst197.
Terrapon, N., Li, C., Robertson, H.M., Ji, L., Meng, X., Booth, W., Chen, Z., Childers, C.P., Glastad, K.M., Gokhale, K., Gowin, J., Gronenberg, W., Hermansen, R.A., Hu, H., Hunt, B.G., Huylmans, A.K., Khalil, S.M.S., Mitchell, R.D., Munoz-Torres, M.C., Mustard, J.A., Pan, H., Reese, J.T., Scharf, M.E., Sun, F., Vogel, H., Xiao, J., Yang, W., Yang, Z., Yang, Z., Zhou, J., Zhu, J., Brent, C.S., Elsik, C.G., Goodisman, M.A.D., Liberles, D.A., Roe, R.M., Vargo, E.L., Vilcinskas, A., Wang, J., Bornberg-Bauer, E., Korb, J., Zhang, G., Liebig, J., 2014. Molecular traces of alternative social organization in a termite genome. Nat. Commun. 5, 3636. https://doi.org/10.1038/ncomms4636.
Veenstra, J.A., 2014. The contribution of the genomes of a termite and a locust to our understanding of insect neuropeptides and neurohormones. Front. Physiol. 5, 454. https://doi.org/10.3389/fphys.2014.00454.
von Wattenwyl, K.B., 1865. Nouveau système des Blattaires. G. Braumüller, Vienne. doi:10.5962/bhl.title.5652.
Wang, Z., Shi, Y., Qiu, Z., Che, Y., Lo, N., 2017. Reconstructing the phylogeny of Blattodea: robust support for interfamilial relationships and major clades. Sci. Rep. 7, 1–8. https://doi.org/10.1038/s41598-017-04243-1.
Wegener, C., Gorbashov, A., 2008. Molecular evolution of neuropeptides in the genus *Drosophila*. Genome Biol. 9 (8), 131. https://doi.org/10.1186/gb-2008-9-8-r131.
Zhang, S.Q., Che, L.H., Li, Y., Pang, H., Ślipiński, A., Zhang, P., 2018. Evolutionary history of Coleoptera revealed by extensive sampling of genes and species. Nat. Commun. 9 (1), 205. https://doi.org/10.1038/s41467-017-02644-4.

## Glossary

***Cleavage site:*** Mostly mono- or dibasic amino acid motifs that serve as recognition sites for enzymes (proprotein convertases) that cleave proprotein sequences to release neuropeptide sequences.

***Multiple-copy precursor:*** A neuropeptide precursor containing multiple neuropeptide sequences (= paracopies).

***Neuropeptide precursor:*** Protein sequence following neuropeptide gene transcription and mRNA translation. After directing the precursor protein (preproprotein) into the endoplasmic reticulum, the N-terminal signal peptide is cleaved and the remaining proprotein is processed by enzymes in biologically active neuropeptides and other precursor-derived peptides.

***Orthocopies:*** Neuropeptide sequences of different species that are derived from a common ancestor.

***Paracopies:*** Multiple peptide copies within a neuropeptide precursor sequence of a certain species.

***Single-copy precursor:*** A neuropeptide precursor containing a single neuropeptide (= single copy) sequence.

# 3. Discussion

The study at hand was designed to investigate the evolution of neuropeptide precursor sequences in the insect group Polyneoptera. Through a multitude of methods, we examined lineage specific differences in polyneopteran neuropeptide composition and sequences. These differences were then used as a basis for phylogenetic analysis inside Blattodea, one order of Polyneoptera.

## 3.1 Identification and Distribution of Products from Novel *tryptopyrokinin* Genes in the Locust, *Locusta migratoria*

The first study followed a discovery of a newly described neuropeptide gene, which was found in two polyneopteran species. The newly described *tryptoPK* gene belongs to the neuropeptide family of FXPRLamides/pyrokinins and CAPA (Yaginuma and Niimi, 2016). This family represents one of the best studied neuropeptide families in insects due to their important role in the regulation of development, mating, muscle contraction and diuresis. FXPRLamides have a well conserved C-terminus and were first described from Blattodea, more specifically from the cockroach, *L. maderae* (Holman et al., 1986), but were later found in other insect orders as well, e.g., Coleoptera, Diptera, Lepidoptera and Hymenoptera (Jurenka and Nusawardani, 2011). Similarly, CAPA peptides were also first described from a cockroach, *P. americana* (Predel et al., 1995), and were later found in many other insect orders. With the uprising of molecular techniques and the rising number of genomes, it was discovered that FXPRLamides and CAPAs are processed from two different genes, the *pk* gene, named *hugin* in *D. melanogaster*, and the *capa* gene (Predel and Wegener, 2006; Jurenka and Nusawardani, 2011; reviewed in Derst et al., 2016). It has been shown that these two genes originated as a single gene with two different putative bioactive ligands (Derst et al., 2016). In the ancestor of hexapods, a third putative bioactive ligand (tryptoPK) with a unique, respective receptor arose (Derst et al, 2016). Tryptopyrokinins were known to be co-expressed on both pyrokinin and

CAPA precursors. Following the release of the *L. migratoria* genome (Wang et al., 2014), four new genes encoding only for tryptopyrokinins were postulated (Veenstra, 2014). We were able to identify products from two of these new genes by mass spectrometry for the first time. Furthermore, I could show that the postulated *tryptoPK* gene 3 and *tryptoPK* gene 4 were in fact partial sequences of a single gene, *tryptoPK* gene 3. Products of *tryptoPK* gene 3, however, could not be identified. For *tryptoPK* genes 1 and 2 we were able to discover the expression patterns using both mass spectrometry and immunocytochemistry. Both genes are expressed in ventral posterior median neurons (VPMNs) in the labial neuromere of the subesophageal ganglion (SEG). From there, axonal projections carry products of *tryptoPK* genes 1 and 2 anteriad into the retrocerebral complex via the *nervi corporis cardiaci*-3 (NCC-3) and distal parts of the *nervi corporis allati*-2 (NCA-2) and posteriad into the lateral cardiac nerves.

Tryptopyrokinins can also be found in VPMNs and homologous cells in the SEG in other insect orders. *Periplaneta* PK-6 and 7, for example, were found in neurosecretory cells in the SEG (Predel et al., 2007). *Periplaneta* PK-6 is a product of the *pk* gene of *P. americana*, whereas *Periplaneta* PK-7 on the other hand is part of a *P. americana tryptoPK* gene (unpublished). In contrast to *L. migratoria*, *tryptoPK* and *pk* genes are expressed together in these neurosecretory SEG cells. Additionally, in *D. melanogaster* tryptoPKs are expressed in SEG cells homologous to the locust VPMNs (Neupert et al., 2009). This indicates a conserved expression pattern of tryptoPKs in cells in the SEG that is present throughout insects and is not driven by a type of gene, but rather the type of expressed neuropeptide.

In his first description of the novel *tryptoPK* genes in *L. migratoria*, Jan Veenstra also proposed a *tryptoPK* gene for another polyneopteran species, the termite *Z. nevadadensis* (Veenstra, 2014). This hinted at the presence of *tryptoPK* genes in additional orders of Polyneoptera. In fact, since this study was published, *tryptoPK* genes have been found in two species of Phasmatodea (Liessem et al., 2018; Veenstra, 2019b). In addition, I was able to extract tryptoPK precursors from transcriptomes of almost all polyneopteran lineages (unpublished), and Jana Redeker was able to identify the respective products of these genes in Blattodea, Dermaptera, Mantophasmatodea, Orthoptera and Phasmatodea in her Master's thesis (Figure 3). These findings suggest that the duplication of either a *pk* or *capa* gene that resulted in the presence of a *tryptoPK* gene already occurred in the last common ancestor of Polyneoptera. There

are no *tryptoPK* genes known from species outside of Polyneoptera, however, indicating that *tryptoPK* genes are a feature which exists exclusively in Polyneoptera.

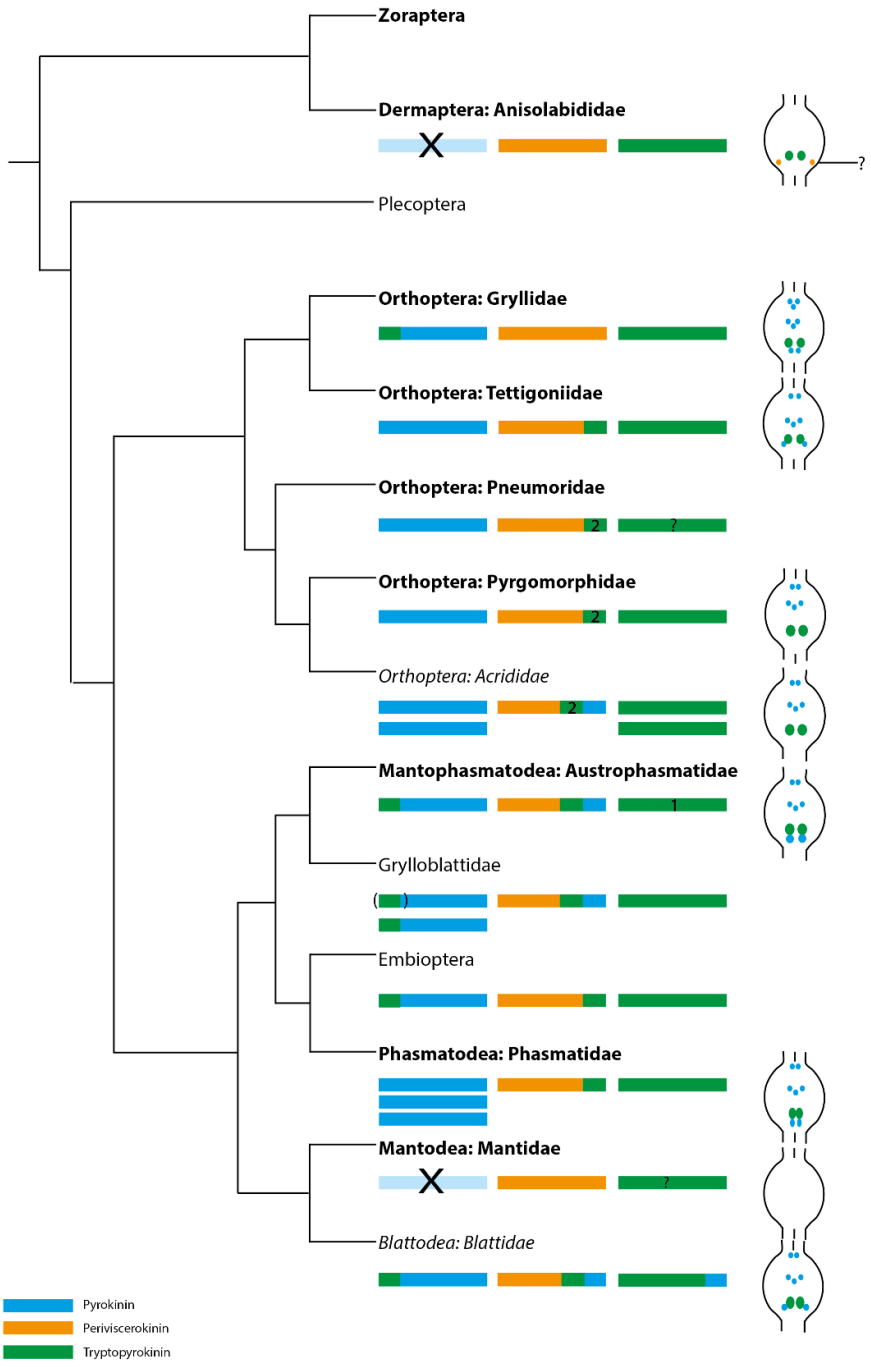



**Figure 3: Schematic overview of the occurrence of *tPK* (green), *capa* (orange) and *pk* (blue) genes in polyneopteran lineages and their putative distribution in cells of the SEG. In some lineages of Orthoptera, two tPKs are present on the *capa* gene (depicted with the number 2). In Grylloblattidae, some species do not possess a tPK sequence on one of their *pk* genes (depicted as brackets). A question mark depicts genes that were found in the transcriptome but could not be identified via mass spectrometry. Courtesy of Jana Redeker (Münster, Germany).**

partially known sequences of the *pk* and *capa* genes (SM 1.1), and we also studied their expression pattern. Especially noteworthy was the discovery of a second pyrokinin precursor, containing only two PKs, *pk-like*. The presence of a second pk precursor was hitherto unknown in insects. Meanwhile, two similar precursors were found in another polyneopteran species, *C. morosus* (Liessem et al., 2018). My study of further polyneopteran transcriptomes, provided by the 1KITE initiative, also revealed pk-like precursors across multiple orders of Polyneoptera (unpublished). Their sequences often share relatively low sequence similarities except for the pyrokinins. The coverage of these pk-like precursors across Polyneoptera however is rather low. Considering the age of the 1KITE transcriptomes and the fact that full-body transcriptomes were analyzed rather than transcriptomes derived from nervous system tissue alone, the low coverage might be artificial. Further studies and new transcriptomes will be needed to provide further insight into the evolutionary origin and expression patterns of these novel genes.

## 3.2 Evolution of Neuropeptide Precursors in Polyneoptera (Insecta)

While the first study was focused on a single neuropeptide family in a single species, in the second study we widened the scope. Here, we analyses orthologs of 21 neuropeptide precursors from over 220 species of Polyneoptera. This may possibly represent the largest in-depth analysis of neuropeptide precursors in insects. The second study aimed at finding lineage specific patterns and evolutionary trends in single-copy neuropeptide composition of Polyneoptera. I could demonstrate that the genes for single-copy neuropeptides are well-conserved in Polyneoptera. Only a very limited set of precursors was completely missing in a given order of Polyneoptera. In most orders I was able to find the complete set of neuropeptide precursors in every species, a feature that was also discovered in apterygote hexapods (Derst et al., 2016). This stands in contrast to the seemingly frequent loss of complete neuropeptide precursors in holometabolous insects. In the fruit fly, *Drosophila melanogaster*, six of the 21 precursors analyzed in our study were absent (Nässel and Zandawala, 2019). The loss of several neuropeptides precursors, including the single-copy corazonin, is also reported from Coleoptera (Pandit et al., 2018; Veenstra, 2019a; Ragionieri and Predel, 2020). Complete losses of multiple neuropeptide precursors were also recently reported from parasitoid wasps

(Hymenoptera) (Chang et al., 2018). Here, the single-copy precursor proctolin, besides others, is missing in all parasitoid wasps and ants. Furthermore, multiple precursors were missing in single lineages of parasitoid wasps, e.g., the single-copy precursors elevenin and trissin were absent in all species of Chalcidoidea among others. Since these frequent losses of complete neuropeptide precursors are not present in apterygote and polyneopteran lineages, it is reasonable to conclude that this specific evolutionary trend developed after the divergence of Condylognatha and Holometabola. It is, however, so far unknown how the loss of function that is connected to the loss of a neuropeptide is recovered in these species. There is no obvious increase in newly arisen neuropeptide genes in these lineages as counteraction to the losses and one would therefore speculate that other previously existing neuropeptides have taken over the functions of the lost genes.

The second study was further designed to find lineage specific patterns and evolutionary trends in neuropeptide evolution of Polyneoptera. Across all 21 neuropeptide precursors from ten polyneopteran orders, we found a few examples of lineage specificity between two or more orders. The most prominent example is the substitution of the first Arg of the C-terminal cleavage site to a Met in RFLamides of Xenonomia (i.e., Mantophasmatodea and Grylloblattodea). RFLamides of these species are therefore enlarged by two AAs compared to other polyneopteran sequences and are not amidated (2.2 Figure 7). These changes only occur in the two genera Mantophasmatodea and Grylloblattodea, but are omnipresent within these orders. Thus, this substitution occurred within the last common ancestor of Xenonomia. Another example is the absence of trissin in Dermaptera and Zoraptera. This absence in both orders also suggests that the loss of the *trissin* gene occurred in the last common ancestor of the two groups. We only analyzed one species of Zoraptera, so further investigations of zorapteran species are needed to confirm this finding. A third example we were able to identify was the presence of Gln at the eleventh position of the peptide in most species of Dictyoptera (i.e., Mantodea and Blattodea). This amino acid is not found at all in trissin sequences of Polyneoptera outside of Dictyoptera.

More prominent than examples of lineage specificity between two or more orders were traits specific to individual orders. This included to the sparse examples of neuropeptide precursor losses mentioned above as well as AA substitutions found only in a single lineage, e.g., substitution of $Gln^1$ to $Pro^1$ (or $Thr^1$) in myosuppressin of Caelifera or substitution of $Gly^1$ to

Gln[1] in allatotropin of Dermaptera. Notably, the most prominent lineage specific traits that were observable in our dataset were changes of the average evolutionary divergence (AED) of the different precursors. Sequence conservation differed remarkably between polyneopteran orders. These differences are noticeable in the AEDs of respective precursors (SM 2.3) as well as the overall median AEDs calculated in our study (2.2 Figure 9). The latter, in particular, highlights differences in the neuropeptide precursor sequence conservation across Polyneoptera. Some orders of Polyneoptera (such as Dermaptera, Plecoptera and especially Orthoptera) show high intra-ordinal sequence diversity, whereas others, such as Grylloblattodea, Mantodea and Mantophasmatodea, exhibit low AEDs. In the case of Grylloblattodea and Mantophasmatodea, these low AEDs can be explained by the fact that these orders represent relict groups with only a few extant and rather closely related taxa. Mantodea however is a highly diverse group with more than 1800 extant species that dates back ~300 Mya (Resh and Cardé, 2009; Legendre et al., 2015), which makes their strong neuropeptide precursor sequence conservation remarkable. Similarly, the AED of the highly diverse Phasmatodea with over 3000 extant species in more than 480 genera (Bradler and Buckley, 2018) is below the average AED of single-copy neuropeptide precursors in Polyneoptera. Considering the coverage of different lineages in the 1KITE dataset, these results are unlikely to be biased by taxon sampling.

The AEDs of respective single-copy neuropeptide precursors are mostly in the range of the median of the individual order. There are exceptions however. The AED of AST-CC in Dermaptera, for example, is significantly lower (0.14) than the average AED of Dermaptera (0.38). In Blattodea the AED of AST-CCC is likewise significantly lower (0.08) than the average AED of Blattodea (0.28), whereas the AED of CNMamide, conversely, is significantly higher (0.53). Thus, the evolution of neuropeptide precursor sequences underwent significant increases and decreases in the AA substitution rate and there is no uniform evolutionary speed in either single neuropeptide genes or peptidergic systems of a given taxon.

Across all 21 single-copy neuropeptides analyzed in our study, the overall AED of the neuropeptide sequence within the precursor is lower than the overall AED of the complete precursor sequence (SM 2.3). This showcases that there are different levels of sequence conservation present on a single neuropeptide precursor and that neuropeptide sequences are in fact more strongly conserved than the remaining precursor sequences. In some cases (CCAP,

Proctolin) not a single AA substitution between all analyzed species could be detected (SM 2.3 Figure E and Figure Q). The low sequence variation allowed us to postulate the individual neuropeptide sequences that were present at the time of the insect emergence more than 400 million years ago for the majority of single-copy neuropeptides.

This study also includes first extensive knowledge about the distribution of the two newly described neuropeptides, HanSolin and RFLamide, in Polyneoptera. Both were present in all orders of Polyneoptera. Although a consensus ancestral sequence of HanSolin for Polyneoptera could not be extracted from our data due to sequence variation, the overall sequence conservation of the HanSolin precursor across Polyneoptera is within range of the median AED of all polyneopteran single-copy precursors (2.2 Figure 6). I was able to determine that the AED of the polyneopteran RFLamide precursor is even slightly below the median AED of all polyneopteran single-copy precursors (2.2 Figure 8). For RFLamide, an ancestral state of the peptide sequence could be identified (2.2 Figure 9), although the RFLamide of Xenonomia is distinctly altered, as explained above.

## 3.3 The Power of Neuropeptide Precursor Sequences to Reveal Phylogenetic Relationships in Insects: a Case Study on Blattodea.

In my third study, I used the knowledge of lineage specific difference of neuropeptide sequences to perform a phylogenetic analysis. Limited datasets of neuropeptide sequences of mature peptides, measured via mass spectrometry, have already been successfully used to reconstruct phylogenies at the intra-familial level (e.g., Roth et al., 2009, Predel et al. 2012). In this study, we provided the first use of complete neuropeptide precursor sequences of Blattodea for the reconstruction of the phylogenetic relationships in this large insect group. Although only about a third of all neuropeptide precursors generally present in insects had a sufficient coverage in our dataset, we were able to demonstrate that these short but well conserved sequences are useful for family-level phylogenetic reconstruction. We wanted to establish this investigation as a proof of principle study for the potential of neuropeptide precursors as a basis for further phylogenetic analysis. Thus, we chose the order Blattodea, in which a subgroup of the 1KITE

initiative was simultaneously working on a blattodean phylogeny based on the same transcriptome dataset, but with a conventional phylogenetic methodology. Our results are consistent with the tree by Evangelista et al. (2019), based on 2,370 protein-coding nuclear single-copy genes and published prior to our study. This highlights the efficacy of our method.

In addition, the novel use of complete neuropeptide precursor sequences allowed for a comparison of the phylogenetic information of functionally different parts of a precursor. A neuropeptide precursor consists of functionally different subunits which serve different purposes (2.3 Figure 1) and are therefore subject to different degrees of sequence conservation (2.3 Figure 5). The signal peptide at the N-terminus of the precursor is needed to direct the preprohormone into the endoplasmic reticulum and therefore into the secretory pathway (see Figure 1). These short sequences fulfill the identical role in every neuropeptide of every species. Nevertheless, we could show that the concatenated sequences of 17 signal peptides (494 amino acids per species) recovered the blattodean phylogeny surprisingly well at the ordinal level (2.3 Figure 4). Concatenated sequences of the putative neuropeptides of these 17 precursors (576 amino acids per species) also yielded a well-resolved topology of Blattodea, although not all clades were maximally supported (2.3 Figure 3). These findings are especially noteworthy since three datasets of comparative and longer alignment lengths (units of 576, 576 and 936 amino acids) of the remaining precursor sequences without predicted neuropeptide sequences and signal peptides resulted in poorly resolved topologies (SM 3.3). Only by combining these three datasets and therefore drastically enlarging the alignment to about four times that of the individual analyses of signal peptide and putative neuropeptides (2088 amino acids per species) were we able to reconstruct a consistent and well-supported topology from the remaining precursor sequences (SM 3.2). Hence, the sequences of both the signal peptides and the putative neuropeptides contain more phylogenetic information per site than the remaining precursor sequences. The phylogenetic information thereby correlates to the functional importance of the unit on the precursor. The signal peptide, as mentioned above, is needed to direct the preprohormone into the secretory pathway and its sequence might be under compensatory evolution with the respective channel proteins of the endoplasmic reticulum (Williams et al., 2000). The recognition of these sequences by those channel proteins is essential and could be a stabilizing factor in the evolution of these sequences. Neuropeptides are irredeemable to trigger the biological function of the underlying gene and are known to undergo co-evolution with their respective receptors (e.g., Park

et al., 2002; Janssen et al., 2010; Kim et al., 2012). The function of the precursor peptide sequences, however, is so far unknown and can only be speculated on (Wegener and Gorbashov, 2008). Our results here indicate that their function is less crucial to the flawless synthesis and function of the neuropeptide than the actual neuropeptide sequence and the signal peptide sequence.

The degree of phylogenetic information not only differed between the functional units of the precursor, but also between the different neuropeptide precursors themselves. I was able to establish that the AEDs of the respective precursors varied between 0.11 (AST-C*) and 0.77 (CAPA) (see 2.3 Table 1). Several precursors with low AEDs, e.g., Ast-C and DH-31, are highly conserved and showed few substitutions across the 40 blattodean species we analysed. Precursors, like CNMamide and CAPA, nevertheless have a higher AED and show more frequent substitutions. Noticeably, two of the three multiple-copy neuropeptide precursors included in this study are among the precursors with the five highest AEDs. Concatenated sequences of these three multiple-copy precursors yielded a topology almost equal to the topology recovered from concatenated sequences of the remaining 14 single-copy neuropeptide sequences (SM 3.4, SM 3.5). The single-copy precursor dataset was twice the length of the multiple-copy dataset, yet only the support of higher level nodes decreased between the results of those two datasets. Multiple-copy neuropeptide precursors have thus proven to possess a high degree of phylogenetic information at the intra-ordinal level. Their use at the inter-ordinal level, however, could create ambiguities in the alignment since the number of copies on an orthologous multiple-copy neuropeptide precursor can be different and evolutionary origins of the internal copies are often unclear. A misalignment of these copies would therefore create bias in the dataset.

In conclusion, to optimize the reconstruction of phylogenetic relationships based on neuropeptide precursor sequences, the sequences of all different functional units, i.e. the complete precursor, and both single-copy and multiple-copy neuropeptide precursor sequences, should be integrated. Hereby, phylogenetic information of strongly and loosely conserved sites is

* Please note that the neuropeptide precursor called Allatostatin-C in this study (Bläser et al., 2020) is actually Allatostatin-CCC following Veenstra (2016). In the second study of this thesis (Bläser and Predel 2020), this precursor is correctly labeled as Allatostatin-CCC. For continuity with the underlying publication, this precursor will still be falsely addressed as Allatostatin-C in this chapter.

combined, thus enabling the simultaneous analysis of phylogenetic relationships between closely as well as distantly related lineages.

Neuropeptide precursor sequences have generally hereby been proven to be a reliable basis for phylogenetic analyses. Although of the around 50 neuropeptide precursors than can be expected in a polyneopteran species (e.g., Veenstra, 2014; Liessem et al., 2018), only 17 possessed a sufficient coverage in all 40 blattodean species to be used in this analysis, we were able to confirm all nodes except the placement of *Z. nevadensis* with maximum support. Furthermore, our results coincide with the findings of Evangelista and colleagues based on a much larger set of nuclear markers (Evangelista et al., 2019). Both studies are based on the same transcriptomes, but since none of the sequences used in our study were used by Evangelista and colleagues (2019), our analyses are independent from each other. The results confirm the applicability of neuropeptide precursor sequences in future phylogenetic analyses. The inclusion of more neuropeptide precursors enabled by newer transcriptomes made from nervous system tissue will only increase the power of neuropeptide precursor sequences in elucidating ambiguous phylogenetic relationships.

# 4. Concluding Remarks and Future Prospects

The investigations carried out during my thesis focused on different aspects of the evolution of neuropeptide precursor sequences. I was able to demonstrate that neuropeptide precursor sequence evolution does not follow a uniform rate, but rather it is unique in every precursor and even differs between orthologs of different lineages. Even on a single neuropeptide precursor, the degree of sequence conservation differs between the functional units of the precursor. The findings of the first study show that in some genes, upsurges of neuropeptide evolution can occur (Redeker, Bläser et al., 2017). These are mostly the result of gene duplications, as evidenced by the arising of the novel *tryptoPK* genes and the novel *pk-like* gene found in *L. migratoria*. The second study indicates that the set of neuropeptide precursors that was present at the emergence of Hexapoda is still mostly present in polyneopteran orders, and that gene loss is not a frequent occurrence. The rise of new neuropeptide genes, however, seems to be a trait more frequently found in Polyneoptera than expected. This observation is supported by later findings, such as the two novel *pk-like* genes (Liessem et al., 2018), and the presence of second copies of multiple single-copy neuropeptide precursors in Polyneoptera (Bläser and Predel, 2020), e.g., the second precursor with different corazonin sequences in *Nippancistroger testaceus* (Ensifera) and *Medauroidea extradentata* (Phasmatodea). Apart from these sudden jumps in neuropeptide precursor evolution, neuropeptide precursors of Polyneoptera show a relatively high degree of sequence conservation.

The analyses of the average evolutionary divergence in both the second and third study allowed for an easy comparison of the degree of sequence conservation in different neuropeptide precursors and orders of Polyneoptera (Bläser et al., 2020; Bläser and Predel, 2020). This statistical analysis should be included in future studies concerning neuropeptide sequence evolution and moreover, is useful in the selection of neuropeptide precursors for phylogenetic analyses. With this tool, we were able to identify generally more slowly evolving, i.e., strongly conserved neuropeptide precursor sequences, like calcitonin-like diuretic hormone (CT-DH/DH-31), which congruently had the lowest median AED in both studies (see SM 2.3 Figure K and 2.3 Figure 1), albeit significant differences between the different orders. These sequences can be used to infer phylogenetic analyses of distantly related lineages. In fact, we used the AED scores of

neuropeptide precursor sequences as a troubleshooting tool in an ongoing phylogenetic analysis of genera of Tenebrionidae (Coleoptera) from the Chilean Atacama Desert (Ragionieri et al., unpublished).

This study is part of the CRC 1211 project (Earth - Evolution at the Dry Limit; https://sfb1211.uni-koeln.de/) in which we are studying the evolutionary history of Tenebrionidae from the hyperarid Chilean Atacama Desert. This marks the first use of neuropeptide precursor sequences for phylogenetic use after the proof-of-principle study in Blattodea (Bläser et al., 2020). We selected 22 neuropeptide precursors from transcriptomes extracted in our laboratory for this purpose. In the beginning of the study, we included 28 neuropeptide precursors into our analyses but could not yield consistent topologies. After closely examining the AEDs of the selected precursors, we identified six precursors with the highest AED values. After excluding these six precursors, the remaining dataset yielded consistent topologies with high node support in various analyses. The precursors with the highest AED values likely caused biases in our dataset and led to inconsistencies in the phylogenetic analyses. Through the AED analysis tool, we were able to exclude these biases from our dataset, highlighting the investigative power of AED analysis for phylogenetic studies based on neuropeptide precursor sequences.

Prior to my thesis, phylogenetic analyses with neuropeptide sequences were based solely on the sequences of mature neuropeptides identified via mass spectrometry (e.g., Roth et al., 2009; Predel et al., 2012). In my third study (Bläser et al., 2020), we were able to show that the sequences of mature neuropeptides are indeed capable of resolving phylogenetic relationships to a certain degree. As demonstrated in my second study, however, the sequences of mature neuropeptides are more strongly conserved than the sequence of the remaining neuropeptide precursor (Bläser and Predel, 2020). This conservation in some neuropeptides is so strong that there are no AA substitutions across all polyneopteran species. Further analysis revealed that this is indeed the case far beyond Polyneoptera. The sequence of the CCAP peptide of Polyneoptera is identical to the sequence of many further insects such as Hemiptera (*Rhodnius proxilus*) and multiple holometabolous lineages, including Hymenoptera (19 species), Lepidoptera (*Bombyx mori*), Diptera (*D. melanogaster*) and Coleoptera (*Tribolium castaneum*) (Chang et al., 2018). Similarly, the ancestral sequence of MS which we found in most orders of Polyneoptera was also present in *Daphnia pulex* and *Xibalbanus tulumensis* (Crustacea), as well as in every apterygote

species (Derst et al., 2016). Additionally, the identical sequence was ubiquitously present in parasitoid wasps and other Hymenoptera (Chang et al., 2018). These examples furthermore showcase that mature neuropeptide sequences are limited in their use as an exclusive basis for phylogenetic analyses. The utilization of complete neuropeptide precursor sequences as implemented in my third study should be the focus of future studies in this regard.

Another aspect that the results of my thesis showcase is that the quality of the underlying transcriptomes is essential. In the first study, we were able to not only find complete transcripts of the tryptoPKs, but were also able to reveal that the precursor fragments previously assigned to different precursors (tryptoPK 3 and tryptoPK 4) were in fact partial sequences of a single precursor, tryptoPK 3. Moreover, we could complete the previously only partially known capa and pk precursors. Furthermore, a novel pk-like precursor was found. Not all of these findings are possible with a transcriptome of lower quality. For comparison, I investigated the sequences mentioned above in the transcriptome of *L. migratoria* provided by the 1KITE initiative and mostly found partial sequences. The pk-like precursor was completely absent. I was able to make a similar observation with the two novel pk-like precursors found in *C. morosus* (Liessem et al., 2018), which I was not able to recover in 1KITE transcriptome of this species.

In the second study, 31% of all precursor sequences we analyzed were only present as partial sequences, and 28% of precursor information across all lineages was missing. Only by combining the information from multiple species per order in our large dataset were we able to formulate statements about absence of certain precursors. Analyses of smaller orders, however, are left uncertain. The most prominent example in our study is the order Zoraptera, where only one species was analyzed. Here, 18% of precursors were partial and 13% were missing. The 13% of missing data are of special interest, since they could possibly indicate lineage specificity between two orders (see 3.2). If we compare these numbers to the transcriptomes of the order Mantophasmatodea, which we prepared ourselves, the differences become more obvious. In Mantophasmatodea, only 1% of sequences were partial, and 5.6% were missing. In this order, four species were sufficient to formulate qualified statements about the neuropeptidome of Mantophasmatodea. We currently cannot assess how many evolutionary trends or intra-ordinal traits we did not see in our dataset due to poor transcriptome quality.

In the third study, only a third of the neuropeptides expected in a polyneopteran species had sufficient coverage to be included in the analysis. Of these precursors selected, 30% were only present as partial sequences (2.3 Table 1). This coverage was sufficient for the intra-ordinal analysis of phylogenetic relationships in Blattodea, but when inter-ordinal relationships are analyzed, there should be a larger number of complete neuropeptide precursors. Alignment ambiguities in multiple-copy neuropeptide precursors of distantly related lineages could reduce the number of neuropeptide precursors suitable for each respective analysis. Here, an increase of neuropeptide precursors in the repertoire of suitable candidate precursors is vital to offset this reduction. In the analysis of the tenebrionid genera from the Chilean Atacama Desert introduced above, we generated our own transcriptomes, specifically, in order to yield the maximum amount of neuropeptide information possible and were able to lower the percentage of partial sequences to about 13% while simultaneously increasing the number of neuropeptide precursors (Ragionieri et al., unpublished).

We were able to increase the set of neuropeptide precursor sequences even further in an ongoing analysis of the *Periplaneta* cluster inside of Blattodea. Here, we found in the literature and in our own analyses that the species of three genera, *Periplaneta*, *Blatta* and *Shelfordella*, are constantly nested together in phylogenies (e.g., Inward et al., 2007; Murienne, 2009; Roth et al., 2009; Legendre et al., 2015). The individual sister group relationships, however, differ between analyses and multiple studies result in these genera being paraphyletic. We generated our own transcriptomes of eight species from these three genera and included six additional blattodean outgroup species to reconstruct their phylogenetic relationships based on neuropeptide precursor sequences. In total, we analyzed 47 neuropeptide precursors, including 11 multiple-copy neuropeptide precursors. This dataset represents the largest set of neuropeptide precursor sequences used for phylogenetic analyses to date. Our results indicate that the genus *Periplaneta* is indeed paraphyletic and that the famous species *Periplaneta americana* belongs in fact to the genus *Blatta*, in which it was originally described (Linné, 1758). Furthermore, the monotypic genus *Shelfordella* also groups together with *Blatta* and *Periplaneta americana*. The remaining species of *Periplaneta* analyzed here form the sister group to this clade. Hence, we will suggest to the International Commission on Zoological Nomenclature that *Periplaneta americana* and *Shelfordella lateralis* be transferred into the genus *Blatta* and that the genus *Periplaneta* be defined by a new holotype without *Periplaneta americana*. In addition to the phylogenetic data,

morphological features to support this taxonomic revision will be needed. The analysis of the male post-abdomen showed some promise in this regard in preliminary literature studies.

As mentioned before, the second study gave an overview over the distribution of the newly described neuropeptides, HanSolin and RFLa, in Polyneoptera for the first time. Both were present in all polyneopteran orders. In additional BLAST searches in 1KITE transcriptomes of apterygote insects, I was able to confirm the presence of both HanSolin and RFLa in species of Diplura and Zygentoma (unpublished). In Protura, I only detected HanSolin but not RFLa and both precursors were absent in Collembola. Moreover, I found both precursors in our own transcriptomes of two species of Odonata, but not in the ephemeropteran species *Ephemera danica*. HanSolin and RFLa were, furthermore, found in Coleoptera (Veenstra, 2019a). The sequences of these new neuropeptide precursors both show a considerable amount of sequence variation across insects (data not shown). Their function, as well as localization in the nervous system, is unknown as of this time. I have started a project in cooperation with Sander Liessem (Büschges Lab, University of Cologne, Cologne, Germany) to investigate the evolution and localization of both neuropeptides in the nervous system of different species of Polyneoptera. For this project, we are including transcriptomics and peptidomics similar to the methodology of my first study. The transcriptome analyses enabled us to produce new antisera against *C. morosus* HanSolin and RFla. These antisera are tested on different parts of the nervous system, and tissue of interest is further analyzed via mass spectrometry or compared to existing mass spectrometry data (Liessem et al., 2018). Preliminary results for HanSolin show a wide but very unique distribution of HanSolin in the nervous system of *C. morosus* and *P. americana* with intense staining in different parts of the brain including the central complex. These preliminary results indicate that HanSolin might be involved in the modulation of locomotor behavior. Future experiments will continue to give further insight into the localization and distribution of both HanSolin and RFLa and likely allude to the function of these novel neuropeptides.

Ongoing studies also include analyses of neuropeptide precursor sequence evolution outside of Insecta. In an ongoing cooperation with Tom Illife (Texas A&M University, Galveston, USA), we are studying the distribution and localization of the neuropeptide family of FXPRLamides/pyrokinins and CAPA in the cave-dwelling Remipedia. This class of crustaceans

is currently established as the sister group of Hexapoda (Reumont et al., 2012; Misof et al., 2014). In Remipedia, the *pk* gene and the *capa* gene have not yet separated (Derst et al., 2016). Thus, studying the localization of the common ancestor of these genes in Remipedia allows us further insight into the evolution of this peptidergic system.

In a second ongoing study, we are analyzing the neuropeptide sequences in early arthropod lineages, namely Onychophora and Tardigrada. This study is carried out in collaboration with the Mayer Lab (University Kassel, Kassel, Germany) and is intended to compare the neuropeptidome of these ancient lineages with the set of neuropeptides commonly found in insects. Preliminary BLAST searches revealed a surprisingly large overlap of neuropeptide precursors between these lineages and insects. Future analyses of these data will yield further insight into the evolution of neuropeptide precursors and their sequences during the evolution of arthropods and their transition onto land.

# 5. Bibliography

Beutel, R. G., Friedrich, F., Ge, S.-Q., and Yang, X.-K. (2014). *Insect morphology and phylogeny: A textbook for students of entomology*. Berlin, Boston: de Gruyter.

Bläser, M., Misof, B., and Predel, R. (2020). The power of neuropeptide precursor sequences to reveal phylogenetic relationships in insects: a case study on Blattodea. *Mol Phylogenet Evol*, 106686.

Bläser, M., and Predel, R. (2020). Evolution of Neuropeptide Precursors in Polyneoptera (Insecta). *Front Endocrinol (Lausanne)* 11, 197.

Blobel, G., and Dobberstein, B. (1975). Transfer of proteins across membranes. I. Presence of proteolytically processed and unprocessed nascent immunoglobulin light chains on membrane-bound ribosomes of murine myeloma. *J Cell Biol* 67, 835–851.

Bradler, S., and Buckley, T. R. (2018). "Biodiversity of Phasmatodea," in *Insect biodiversity: Science and society*, eds. R. Foottit, and P. H. Adler (Hoboken, NJ: John Wiley & Sons), 281–313.

Brownstein, M. J., Russell, J. T., and Gainer, H. (1980). Synthesis, transport, and release of posterior pituitary hormones. *Science* 207, 373–378.

Caers, J., Verlinden, H., Zels, S., Vandersmissen, H. P., Vuerinckx, K., and Schoofs, L. (2012). More than two decades of research on insect neuropeptide GPCRs: an overview. *Front Endocrinol (Lausanne)* 3, 151.

Chang, J., Zhao, J., and Tian, X. (2018). In silico prediction of neuropeptides in Hymenoptera parasitoid wasps. *PLoS ONE* 13, e0193561.

Coast, G. M., and Schooley, D. A. (2011). Toward a consensus nomenclature for insect neuropeptides and peptide hormones. *Peptides* 32, 620–631.

Conklin, B. R., Herzmark, P., Ishida, S., Voyno-Yasenetskaya, T. A., Sun, Y., Farfel, Z., et al. (1996). Carboxyl-terminal mutations of Gq alpha and Gs alpha that alter the fidelity of receptor activation. *Mol Pharmacol* 50, 885–890.

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

Eckert, M., Herbert, Z., Pollák, E., Molnár, L., and Predel, R. (2002). Identical cellular distribution of all abundant neuropeptides in the major abdominal neurohemal system of an insect (Periplaneta americana). *J Comp Neurol* 452, 264–275.

Elphick, M. R., Mirabeau, O., and Larhammar, D. (2018). Evolution of neuropeptide signalling systems. *J Exp Biol*. 221. doi: 10.1242/jeb.151092 [Epub ahead of print].

Evangelista, D. A., Wipfler, B., Béthoux, O., Donath, A., Fujita, M., Kohli, M. K., et al. (2019). An integrative phylogenomic approach illuminates the evolutionary history of cockroaches and termites (Blattodea). *Proc Biol Sci* 286, 20182076.

Fricker, L. D. (2012). Neuropeptides and Other Bioactive Peptides: From Discovery to Function. *Colloquium Series on Neuropeptides* 1, 1–122.

Garden, R. W., Moroz, T. P., Gleeson, J. M., Floyd, P. D., Li, L., Rubakhin, S. S., et al. (1999). Formation of N-pyroglutamyl peptides from N-Glu and N-Gln precursors in Aplysia neurons. *J Neurochem* 72, 676–681.

Grimmelikhuijzen, C. J. P., and Hauser, F. (2012). Mini-review: the evolution of neuropeptide signaling. *Regulatory Peptides* 177 Suppl, S6-9.

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.

Hauser, F., Cazzamali, G., Williamson, M., Blenau, W., and Grimmelikhuijzen, C. J. P. (2006). A review of neurohormone GPCRs present in the fruitfly Drosophila melanogaster and the honey bee Apis mellifera. *Prog Neurobiol* 80, 1–19.

Holman, G. M., Cook, B. J., and Nachman, R. J. (1986). Primary structure and synthesis of a blocked myotropic neuropeptide isolated from the cockroach, Leucophaea maderae. *Comparative Biochemistry and Physiology Part C: Comparative Pharmacology* 85, 219–224.

Hook, V., Funkelstein, L., Lu, D., Bark, S., Wegrzyn, J., and Hwang, S.-R. (2008). Proteases for processing proneuropeptides into peptide neurotransmitters and hormones. *Annu Rev Pharmacol Toxicol* 48, 393–423.

Huybrechts, J., Bonhomme, J., Minoli, S., Prunier-Leterme, N., Dombrovsky, A., Abdel-Latief, M., et al. (2010). Neuropeptide and neurohormone precursors in the pea aphid, Acyrthosiphon pisum. *Insect Mol Biol* 19 Suppl 2, 87–95.

Inward, D., Beccaloni, G., and Eggleton, P. (2007). Death of an order: a comprehensive molecular phylogenetic study confirms that termites are eusocial cockroaches. *Biol Lett* 3, 331–335.

Janssen, T., Lindemans, M., Meelkop, E., Temmerman, L., and Schoofs, L. (2010). Coevolution of neuropeptidergic signaling systems: from worm to man. *Ann N Y Acad Sci* 1200, 1–14.

Jékely, G. (2013). Global view of the evolution and diversity of metazoan neuropeptide signaling. *Proc Natl Acad Sci U S A* 110, 8702–8707.

Jurenka, R., and Nusawardani, T. (2011). The pyrokinin/ pheromone biosynthesis-activating neuropeptide (PBAN) family of peptides and their receptors in Insecta: evolutionary trace indicates potential receptor ligand-binding domains. *Insect Mol Biol* 20, 323–334.

Kim, D.-K., Cho, E. B., Moon, M. J., Park, S., Hwang, J.-I., Do Rego, J.-L., et al. (2012). Molecular Coevolution of Neuropeptides Gonadotropin-Releasing Hormone and Kisspeptin with their Cognate G Protein-Coupled Receptors. *Front Neurosci* 6, 3.

Kim, K.-H., and Seong, B. L. (2001). Peptide amidation: Production of peptide hormonesin vivo andin vitro. *Biotechnol. Bioprocess Eng.* 6, 244–251.

Kleine, B., and Rossmanith, W. G. (2016). *Hormones and the Endocrine System*. Cham: Springer International Publishing.

Larhammar, D. (2009). "Neuropeptides Phylogeny and Evolution," in *Encyclopedia of neuroscience*, ed. L. R. Squire (London: Academic Elsevier), 873–880.

Legendre, F., Nel, A., Svenson, G. J., Robillard, T., Pellens, R., and Grandcolas, P. (2015). Phylogeny of Dictyoptera: Dating the Origin of Cockroaches, Praying Mantises and Termites with Molecular Data and Controlled Fossil Evidence. *PLoS ONE* 10, e0130127.

Li, B., Predel, R., Neupert, S., Hauser, F., Tanaka, Y., Cazzamali, G., et al. (2008). Genomics, transcriptomics, and peptidomics of neuropeptides and protein hormones in the red flour beetle Tribolium castaneum. *Genome Res* 18, 113–122.

Li, C., and Kim, K. (2008). Neuropeptides. *WormBook*, 1–36.

Li, L., Garden, R. W., Romanova, E. V., and Sweedler, J. V. (1999). In situ sequencing of peptides from biological tissues and single cells using MALDI-PSD/CID analysis. *Anal Chem* 71, 5451–5458.

Liessem, S., Ragionieri, L., Neupert, S., Büschges, A., and Predel, R. (2018). Transcriptomic and Neuropeptidomic Analysis of the Stick Insect, Carausius morosus. *J Proteome Res* 17, 2192–2204.

Linné, C. von (1758). *Caroli Linnaei…Systema naturae per regna tria naturae :secundum classes, ordines, genera, species, cum characteribus, differentiis, synonymis, locis*. Holmiae: Impensis Direct. Laurentii Salvii.

Loof, A. de, and Schoofs, L. (1990). Homologies between the amino acid sequences of some vertebrate peptide hormones and peptides isolated from invertebrate sources. *Comparative Biochemistry and Physiology Part B: Comparative Biochemistry* 95, 459–468.

Mendoza, A. de, Sebé-Pedrós, A., and Ruiz-Trillo, I. (2014). The evolution of the GPCR signaling system in eukaryotes: modularity, conservation, and the transition to metazoan multicellularity. *Genome Biol Evol* 6, 606–619.

Merighi, A., ed (2011). *Neuropeptides: Methods and Protocols*. Totowa, NJ: Humana Press.

Merkler, D. J. (1994). C-Terminal amidated peptides: Production by the in vitro enzymatic amidation of glycine-extended peptides and the importance of the amide to bioactivity. *Enzyme and Microbl Technol* 16, 450–456.

Mirabeau, O., and Joly, J.-S. (2013). Molecular evolution of peptidergic signaling systems in bilaterians. *Proc Natl Acad Sci U S A* 110, E2028-37.

Misof, B., Liu, S., Meusemann, K., Peters, R. S., Donath, A., Mayer, C., et al. (2014). Phylogenomics resolves the timing and pattern of insect evolution. *Science* 346, 763–767.

Murienne, J. (2009). Molecular data confirm family status for the Tryonicus–Lauraesilpha group (Insecta: Blattodea: Tryonicidae). *Org Divers Evol* 9, 44–51.

Murphy, D., and Gainer, H., eds (2016). *Molecular neuroendocrinology: From genome to physiology*. Oxford, Chichester, Hoboken, NJ: Wiley Blackwell.

Nachman, R. J., Holman, G. M., Haddon, W. F., and Ling, N. (1986). Leucosulfakinin, a sulfated insect neuropeptide with homology to gastrin and cholecystokinin. *Science* 234, 71–73.

Nagata, S. (2016). "FMRFamides," in *Handbook of hormones: Comparative endocrinology for basic and clinical research*, eds. Y. Takei, H. Ando, and K. Tsutsui (Oxford: Academic Press), 432–433.

Nässel, D. R. (1993). Neuropeptides in the insect brain: a review. *Cell Tissue Res* 273, 1–29.

Nässel, D. R., and Winther, A. M. E. (2010). Drosophila neuropeptides in regulation of physiology and behavior. *Prog Neurobiol* 92, 42–104.

Nässel, D. R., and Zandawala, M. (2019). Recent advances in neuropeptide signaling in Drosophila, from genes to physiology and behavior. *Prog Neurobiol* 179, 101607.

Neupert, S., Huetteroth, W., Schachtner, J., and Predel, R. (2009). Conservation of the function counts: homologous neurons express sequence-related neuropeptides that originate from different genes. *J Neurochem* 111, 757–765.

Neupert, S., Marciniak, P., Köhler, R., Nachman, R. J., Suh, C. P.-C., and Predel, R. (2018). Different processing of CAPA and pyrokinin precursors in the giant mealworm beetle Zophobas atratus (Tenebrionidae) and the boll weevil Anthonomus grandis grandis (Curculionidae). *Gen Comp Endocrinol* 258, 53–59.

Ons, S., Richter, F., Urlaub, H., and Pomar, R. R. (2009). The neuropeptidome of Rhodnius prolixus brain. *Proteomics* 9, 788–792.

Owji, H., Nezafat, N., Negahdaripour, M., Hajiebrahimi, A., and Ghasemi, Y. (2018). A comprehensive review of signal peptides: Structure, roles, and applications. *Eur J Cell Biol* 97, 422–441.

Pandit, A. A., Ragionieri, L., Marley, R., Yeoh, J. G. C., Inward, D. J. G., Davies, S.-A., et al. (2018). Coordinated RNA-Seq and peptidomics identify neuropeptides and G-protein coupled receptors (GPCRs) in the large pine weevil Hylobius abietis, a major forestry pest. *Insect Biochem Mol Biol* 101, 94–107.

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.

Pauls, D., Chen, J., Reiher, W., Vanselow, J. T., Schlosser, A., Kahnt, J., et al. (2014). Peptidomics and processing of regulatory peptides in the fruit fly Drosophila. *EuPA Open Proteomics* 3, 114–127.

Pollák, E., Eckert, M., Molnár, L., and Predel, R. (2005). Differential sorting and packaging of capa-gene related products in an insect. *J Comp Neurol* 481, 84–95.

Predel, R. (2001). Peptidergic neurohemal system of an insect: Mass spectrometric morphology. *J Comp Neurol* 463, 363–375.

Predel, R., Brandt, W., Kellner, R., Rapus, J., Nachman, R. J., and Gäde, G. (1999). Post-translational modifications of the insect sulfakinins: sulfation, pyroglutamate-formation and O-methylation of glutamic acid. *Eur J Biochem* 263, 552–560.

Predel, R., Linde, D., Rapus, J., Vettermann, S., and Penzlin, H. (1995). Periviscerokinin (Pea-PVK): A novel myotropic neuropeptide from the perisympathetic organs of the American cockroach. *Peptides* 16, 61–66.

Predel, R., Neupert, S., Derst, C., Reinhardt, K., and Wegener, C. (2018). Neuropeptidomics of the Bed Bug Cimex lectularius. *J Proteome Res* 17, 440–454.

Predel, R., Neupert, S., Garczynski, S. F., Crim, J. W., Brown, M. R., Russell, W. K., et al. (2010). Neuropeptidomics of the mosquito Aedes aegypti. *J Proteome Res* 9, 2006–2015.

Predel, R., Neupert, S., Huetteroth, W., Kahnt, J., Waidelich, D., and Roth, S. (2012). Peptidomics-based phylogeny and biogeography of Mantophasmatodea (Hexapoda). *Syst Biol* 61, 609–629.

Predel, R., Neupert, S., Russell, W. K., Scheibner, O., and Nachman, R. J. (2007). Corazonin in insects. *Peptides* 28, 3–10.

Predel, R., and Wegener, C. (2006). Biology of the CAPA peptides in insects. *Cell Mol Life Sci* 63, 2477–2490.

Ragionieri, L., and Predel, R. (2020). The neuropeptidome of Carabus (Coleoptera, Adephaga: Carabidae). *Insect Biochem Mol Biol* 118, 103309.

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

Resh, V. H., and Cardé, R. T. (2009). *Encyclopedia of insects*. Amsterdam: Academic Press.

Reumont, B. M. von, Jenner, R. A., Wills, M. A., Dell'ampio, E., Pass, G., Ebersberger, I., et al. (2012). Pancrustacean phylogeny in the light of new phylogenomic data: support for Remipedia as the possible sister group of Hexapoda. *Mol Biol Evol* 29, 1031–1045.

Roth, S., Fromm, B., Gäde, G., and Predel, R. (2009). A proteomic approach for studying insect phylogeny: CAPA peptides of ancient insect taxa (Dictyoptera, Blattoptera) as a test case. *BMC Evol Biol* 9, 50.

Schneider, L. E., and Taghert, P. H. (1988). Isolation and characterization of a Drosophila gene that encodes multiple neuropeptides related to Phe-Met-Arg-Phe-NH2 (FMRFamide). *Proc Natl Acad Sci U S A* 85, 1993–1997.

Schoofs, L., Loof, A. de, and van Hiel, M. B. (2017). Neuropeptides as Regulators of Behavior in Insects. *Annu Rev Entomol* 62, 35–52.

Simonds, W. F., Goldsmith, P. K., Woodard, C. J., Unson, C. G., and Spiegel, A. M. (1989). Receptor and effector interactions of G s Functional studies with antibodies to the α s carboxyl-terminal decapeptide. *FEBS Lett* 249, 189–194.

Starratt, A. N., and Brown, B. E. (1975). Structure of the pentapeptide proctolin, a proposed neurotransmitter in insects. *Life Sciences* 17, 1253–1256.

Sullivan, K. A., Miller, R. T., Masters, S. B., Beiderman, B., Heideman, W., and Bourne, H. R. (1987). Identification of receptor contact site involved in receptor-G protein coupling. *Nature* 330, 758–760.

Taussig, R., and Scheller, R. H. (1986). The Aplysia FMRFamide gene encodes sequences related to mammalian brain peptides. *DNA* 5, 453–461.

van den Pol, A. N. (2012). Neuropeptide transmission in brain circuits. *Neuron* 76, 98–115.

van Veelen, P. A., Jiménez, C. R., Li, K. W., Geraerts, W.P.M., Tjaden, U. R., and van der Greef, J. (1994). New advanced analytical tools in peptide/protein research: matrix assisted laser desorption mass spectrometry. *J Control Release* 29, 223–229.

Vaudry, H., and Seong, J. Y. (2014). Neuropeptide GPCRs in Neuroendocrinology. *Front Endocrinol (Lausanne)* 5, 41.

Veenstra, J. A. (2000). Mono- and dibasic proteolytic cleavage sites in insect neuroendocrine peptide precursors. *Arch. Insect Biochem.* 43, 49–63.

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

Veenstra, J. A. (2019a). Coleoptera genome and transcriptome sequences reveal numerous differences in neuropeptide signaling between species. *PeerJ* 7, e7144.

Veenstra, J. A. (2019b). Two Lys-vasopressin-like peptides, EFLamide, and other phasmid neuropeptides. *Gen Comp Endocrinol* 278, 3–11.

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.

Wegener, C., and Gorbashov, A. (2008). Molecular evolution of neuropeptides in the genus Drosophila. *Genome Biology* 9, R131.

Wied, D. de (1971). Long term effect of vasopressin on the maintenance of a conditioned avoidance response in rats. *Nature* 232, 58–60.

Wied, D. de (2000). Peptide hormones and neuropeptides: birds of a feather. *Trends in Neurosciences* 23, 113.

Williams, E. J.B., Pal, C., and Hurst, L. D. (2000). The molecular evolution of signal peptides. *Gene* 253, 313–322.

Wipfler, B., Letsch, H., Frandsen, P. B., Kapli, P., Mayer, C., Bartel, D., et al. (2019). Evolutionary history of Polyneoptera and its implications for our understanding of early winged insects. *Proc Natl Acad Sci U S A* 116, 3024–3029.

Yaginuma, T., and Niimi, T. (2016). "FXPRLamide Peptide Family," in *Handbook of hormones: Comparative endocrinology for basic and clinical research*, eds. Y. Takei, H. Ando, and K. Tsutsui (Oxford: Academic Press), 395–396.

Zupanc, G. K.H. (1996). Peptidergic transmission: From morphological correlates to functional implications. *Micron* 27, 35–91.

# 6. Teilpublikationen

## 6.1 Research Articles

Identification and distribution of products from novel tryptopyrokinin genes in the locust, *Locusta migratoria*

Jana Redeker*, **Marcel Bläser***, Susanne Neupert, Reinhard Predel, 2017

* gemeinsame Erstautorenschaft

*Biochem Biophys Res Commun (**published** doi:10.1016/j.bbrc.2017.02.135)*

The power of neuropeptide precursor sequences to reveal phylogenetic relationships in insects: a case study on Blattodea

**Marcel Bläser**, Bernhard Misof, Reinhard Predel, 2019

*Mol Phylogenet Evol (**published** doi:10.1016/j.ympev.2019.106686)*

Evolution of neuropeptide precursors in Polyneoptera (Insecta)

**Marcel Bläser**, Reinhard Predel , 2020

*Front Endocrinol (**published** doi:10.3389/fendo.2020.00197)*

## 6.2 Invited Talks

**14th Rauischholzhausen Seminar of Development and Plasticity of the insect nervous system, Rauischolzhausen, Germany, May 2016**

Title: Exploration of Polyneopteran Transcriptomes for Evolution of Neuropeptides in Insects, especially Polyneoptera

**8th Dresden Meeting Insect Phylogeny, Dresden, Germany, Sep 2017**

Title: Neuropeptide Precursor Sequences Provide Comprehensive Data to Elucidate Phylogenetic Relationships within Blattodea

**111th Annual Conference of the German Zoological Society (DZG), Greifswald, Germany, Sep 2018**
Title: Neuropeptide Precursor Sequences Provide Comprehensive Data to Elucidate Phylogenetic Relationships within Blattodea

## 6.3 Poster Abstracts

**4th Satellite Symposia of Arthropod Neural Network (ANN) at the 109th Annual Conference of the German Zoological Society (DZG), Kiel, Germany, Sep 2016**
Title: Analysis of the evolution of Polyneoptera (Insecta) by using precursor sequences of neuropeptides

**109th Annual Meeting of the German Zoological Society (DZG), Kiel, Germany, Sep 2016**
Title: Analysis of the evolution of Polyneoptera (Insecta) by using precursor sequences of neuropeptides

**1st Arthropod Neural Network (ANN) Spring Meeting, Altleiningen, Germany, Mar 2018**
Title: Evolution of Neuropeptides over the last 300 mya

**112th Annual Conference of the German Zoological Society (DZG), Jena, Germany, Sep 2019**
Title: Evolution Neuropeptide Precursors in polyneopteran insects

# 7. Acknowledgments

First of I want to thank Prof. Dr. Reinhard Predel, not only for providing me with the opportunity to write my thesis in this interesting topic, but for his support and scientific guidance since my Bachelor thesis. I will never forget the trips to Chile and South Africa.

I am grateful to PD Dr. Benjamin Altenhein for kindly agreeing to be a tutor in my thesis advisory committee and taking on the task as second reviewer of this thesis.

Likewise, I am grateful to Prof. Dr. Bernhard Misof for kindly agreeing to be a tutor in my thesis advisory committee and taking time out of his busy schedule to co-author on one of my publications and giving me solid advice in phylogenetic questions.

I thank all current and former members of the AG Predel for the scientific discussions and their support particularly in the early days of my thesis. Many thanks especially for the always pleasant interactions in the office, during Mensa breaks, and during after-hours beers.

I am grateful to the 1KITE initiative, especially Dr. Karen Meusemann (University of Freiburg, Germany) and Dr. Alexander Donath (ZFMK Bonn, Germany), for providing early access to the 1KITE transcriptome assemblies and raw data. Furthermore, I would like to thank all other collaboration partners of ongoing projects.

I especially thank Susanne Neupert, Jana Redeker, Sander Liessem, Ludwig Bläser and Heather D. Humphrey for their time and expertise in proofreading this thesis.

Finally, I thank my family and friends for their support throughout 10 years of study.

# 8. Supplementary Material

**Table of contents:**

## SM 1 Identification and Distribution of Products from Novel *tryptopyrokinin* Genes in the Locust, *Locusta migratoria*

**SM 1.1** Supplementary Material 1 Sequences of L. migratoria tryptoPK, CAPA, PK, and PKL precursors as obtained from trancriptome analysis of CNS tissue. Amino acids different from those already published for L. m. manilensis [12] are marked in bold green. Predicted signal peptides (highlighted in grey), amidation signals (italic), cleavage signals (italics, red) and supposed bioactive mature peptides (underscored) are indicated. The tryptoPK precursors 3 and 4, published in Veenstra [12], likely represent partial sequences of a single precursor which is the tryptoPK 3 precursor presented here. Allelic differences in the CAPA precursor of L. m. migratoria specimens, which were analyzed in our study, are highlighted in yellow.

pp. 72-73

**SM 1.2** Supplementary Material 2 MALDI-TOF/TOF fragmentation spectrum of PK-$4_{2\text{-}9}$ [pQ] at m/z 966.55, which was regularly detected in mass spectra of preparations of Mx Cells, Md Cells and the RC. Prominent y- and b-fragments, as well as internal fragments (mainly proline directed) are labeled. The fragments were analyzed manually and confirmed that this sequence is part of a longer peptide from the PK precursor, PK-4.

p. 74

**Supplementary Material 1** Sequences of *L. migratoria* tryptoPK, CAPA, PK, and PKL precursors as obtained from trancriptome analysis of CNS tissue. Amino acids different from those already published for *L. m. manilensis* [12] are marked in bold green. Predicted signal peptides (highlighted in grey), amidation signals (italic), cleavage signals (italics, red) and supposed bioactive mature peptides (underscored) are indicated. The tryptoPK precursors 3 and 4, published in Veenstra [12], likely represent partial sequences of a single precursor which is the tryptoPK 3 precursor presented here. Allelic differences in the CAPA precursor of *L. m. migratoria* specimens, which were analyzed in our study, are highlighted in yellow.

*>Locusta m. migratoria* tryptoPK1 precursor

MPRGAQLFLLLALVTAARVLDARAESSR**P**GSDAQSSHQDGRGA**A**HDRNDESNELNDENRSDGDAQDATFRR
GRE**V**QEF**G**SH**VA**GPIG**S**GEED**P**WL**T**LADGSYIPAEAVRELVQ*R*G**S**TLSENTG**V**WFGPRY*GRR*TSCGENVPL
KWLSQVE*KR*AAKQPALWFGPRV*GR*SLDEEPKGEEWRDDDKGLK**G**GSSQRQD*R*SAQPPGLWFGPRV*GRR*SDA
QVDDMLWFGPR**P*G***RSVDTDKQDLYDDEVAMRDQ*R*GAKHPGLWFGPRF*GR*-

*>Locusta m. migratoria* tryptoPK2 precursor

MESSMDVLPVALLTLAALLFSVETSEVLH**D**VND**H**NGSNRD**FV**EQTPSAVLSRS*KR*NNSG**P**HV**N**KWFRPDTG
AGEDYWNALFREDAEDHLSKPDASTVDRLENVVPTAE**A**PRLSVAQH**EL**RSEGSH**V**DH**S**DSEPEVWIGQHSG
*R*SIPEPGT**L**FGPRI*GR*SHTE**A**G**V**WFGPRY*GR*SYPEPGMWFGPRV*GR***SQA**EPG**V**WFG**A**R**I***GR***N**Q**P**EPG**T**WFG
ARI*GR***SH**PEPGMWFGPRV*GR*SHPEPGTWFGPRI*GR*SHSEPELRI-

*>Locusta m. migratoria* tryptoPK3 precursor (new)

..RRTHSSLAVAVTVVAIFAFAVSAAKPREMHSKNVVVQRCIQTGFSDGGSI*R*SVPETSLWFGPRI*GR*SNL
ETALLFGRCVGCSHPGTSQCCEPLA*RR*SGSEDSLGYSSPEPPHDNRKISPESGLWFVDRN*GRR*QPDTALWF
GPRV*GRR*MQHIQPEASEYFRPDV*RR*SGPEHSLWSDTQLRQGDPEPPLHVRISNPKQNLWFRPDT*RR*RQSEN
ILQFGPQV*GRR*NSESALWFGPRV*GR*SYLESNLSDERYVQHRNLEPGLWSNNLGQIQTTPSMDVTNRFP*GR*D
LWSGTDLDRTKHETALWFGPRI*GR*SNPETNLWFGPRV*GR*SHPETSQSFGQYARHSNVETGLWPDDFVQSRP
TTLEDVT*KR*NPERNWWSVNGV*GR*TKHETTLWFGPRI*GR*SNPEINLWFGPRV*GR*SHPETSQSFGPYS*RR*NNV
ESGLWPDNLGQSRPTTWEDVT*KR*YPGHNWWSVNGL*GR*TKHETALWFGPRI*GR*SNPDTHLWFGPRV*GR*SQSL
ASEQLR-

*>Locusta m. migratoria* PK precursor (completed)

MAPAVAAALLLLCSASVAAAHGGGGSWVS*RR*EGDFTPRL*GR*ESAEQGGGVSAWQGGEPQQEEQVLAGPFVP
RL*GR*GAVPAAQFSPRL*GRR*DPPVDGPLVWLPLQVSPRLA*RR*RQQPFVPRL*GR*DSGDEWPQQPFVPRL*GRR*L
HQNGMPFSPRL*GR*DAAEQQPADE-

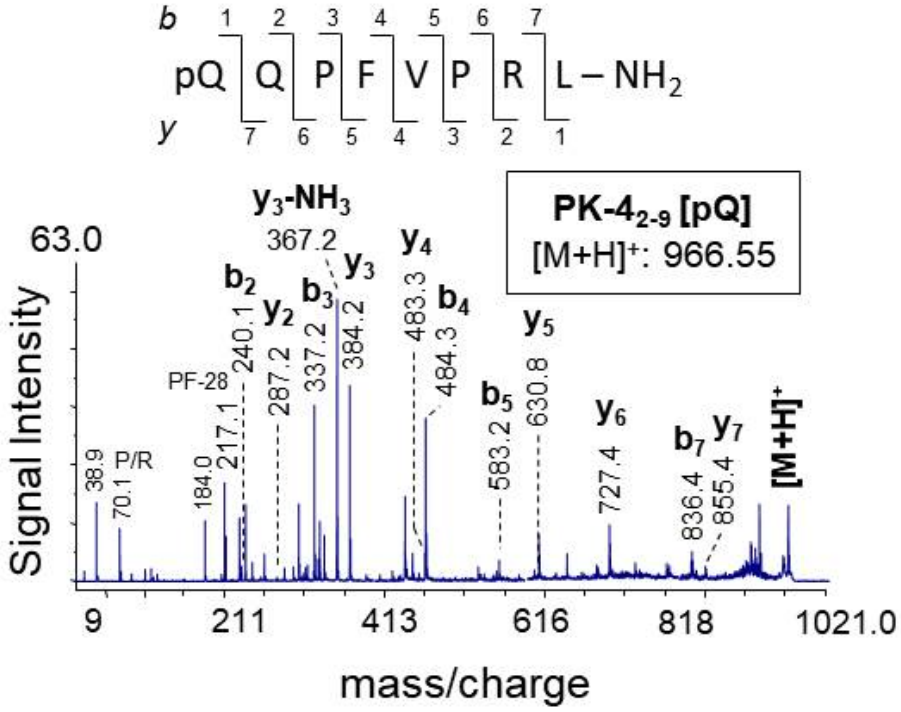
*>Locusta m. migratoria* PK-like precursor (new)

MLPLPALWASVLAAAGSFPAGAASASAW*RR*QSVPTFTPRL*GR*DSAGDELAEEDAVDDGGDGLPQPQLAPPF
WPRP*GRR*APPTRGAAPPTS-

*>Locusta m. migratoria* CAPA precursor (completed)

MAAPSTSRAAAAALLAAAALALLAAAAAQDGDKGISKLKKTSSLFPHPRI*GR*SDYMNLGAG/-
EVGE*KR*AAGLFQFPRV*GR*ALVHGPLPFPLGVFSPLQLAPHQPADTAADLDDGNVSEQQQQPPPSALGH*KR*K
GLVASARV*GRR*DGGQ/EPAAPLWFGPRV*GR*SDLLGLGAAHAQEARAGT*KR*RGLLAFPRV*GR*SGSDGDRDRD
RQGLWFGR/PRV*GRR*E*RR*SLRLRLPAAAWLAAGDVGGTKGDFTPRLGRESGEDEAAVLLVGDGGAEGFDDA
DIDTEER-

**Supplementary Material 2** MALDI-TOF/TOF fragmentation spectrum of PK-$4_{2\text{-}9}$ [pQ] at m/z 966.55, which was regularly detected in mass spectra of preparations of Mx Cells, Md Cells and the RC. Prominent y- and b-fragments, as well as internal fragments (mainly proline directed) are labeled. The fragments were analyzed manually and confirmed that this sequence is part of a longer peptide from the PK precursor, PK-4.



## SM 2 Evolution of Neuropeptide Precursors in Polyneoptera (Insecta)

**Additional File 1:** Coverage of neuropeptide precursors.

| ACP | complete | partial | missing |
|---|---|---|---|
| Blattodea | 7 | 14 | 20 |
| Dermaptera | 0 | 0 | 16 |
| Embioptera | 1 | 1 | 3 |
| Grylloblattodea | 2 | 4 | 0 |
| Mantodea | 0 | 18 | 34 |
| Mantophasmatodea | 4 | 0 | 0 |
| Orthoptera: Caelifera | 10 | 5 | 12 |
| Orthoptera: Ensifera | 3 | 6 | 7 |
| Phasmatodea | 1 | 4 | 32 |
| Plecoptera | 1 | 3 | 11 |
| Zoraptera | 0 | 1 | 0 |
| Overall | 29 | 56 | 135 |
| Overall percentages | 13.18 | 25.45 | 61.36 |

| AST-CC | complete | partial | missing |
|---|---|---|---|
| Blattodea | 24 | 12 | 5 |
| Dermaptera | 11 | 4 | 1 |
| Embioptera | 5 | 1 | 0 |
| Grylloblattodea | 2 | 3 | 1 |
| Mantodea | 37 | 14 | 1 |
| Mantophasmatodea | 4 | 0 | 0 |
| Orthoptera: Caelifera | 18 | 6 | 3 |
| Orthoptera: Ensifera | 4 | 6 | 6 |
| Phasmatodea | 16 | 12 | 9 |
| Plecoptera | 10 | 3 | 2 |
| Zoraptera | 1 | 0 | 0 |
| Overall | 132 | 61 | 28 |
| Overall percentages | 59.73 | 27.60 | 12.67 |

| AST-CCC | complete | partial | missing |
|---|---|---|---|
| Blattodea | 4 | 22 | 15 |
| Dermaptera | 5 | 7 | 4 |
| Embioptera | 2 | 2 | 1 |
| Grylloblattodea | 1 | 3 | 2 |
| Mantodea | 16 | 24 | 12 |
| Mantophasmatodea | 3 | 1 | 0 |
| Orthoptera: Caelifera | 1 | 9 | 17 |
| Orthoptera: Ensifera | 1 | 10 | 5 |
| Phasmatodea | 3 | 19 | 15 |
| Plecoptera | 7 | 7 | 1 |
| Zoraptera | 1 | 0 | 0 |
| Overall | 44 | 104 | 72 |
| Overall percentages | 20.00 | 47.27 | 32.73 |

| AT | complete | partial | missing |
|---|---|---|---|
| Blattodea | 33 | 6 | 2 |
| Dermaptera | 3 | 7 | 6 |
| Embioptera | 4 | 1 | 0 |
| Grylloblattodea | 3 | 3 | 0 |
| Mantodea | 21 | 22 | 9 |
| Mantophasmatodea | 4 | 0 | 0 |
| Orthoptera: Caelifera | 16 | 8 | 3 |
| Orthoptera: Ensifera | 4 | 6 | 6 |
| Phasmatodea | 19 | 14 | 4 |
| Plecoptera | 5 | 5 | 5 |
| Zoraptera | 1 | 0 | 0 |
| Overall | 113 | 72 | 35 |
| Overall percentages | 51.36 | 32.73 | 15.91 |

| CCAP | complete | partial | missing |
|---|---|---|---|
| Blattodea | 29 | 11 | 1 |
| Dermaptera | 12 | 3 | 1 |
| Embioptera | 5 | 0 | 0 |
| Grylloblattodea | 1 | 4 | 1 |
| Mantodea | 20 | 20 | 12 |
| Mantophasmatodea | 4 | 0 | 0 |
| Orthoptera: Caelifera | 7 | 12 | 8 |
| Orthoptera: Ensifera | 2 | 9 | 5 |
| Phasmatodea | 13 | 14 | 10 |
| Plecoptera | 6 | 6 | 3 |
| Zoraptera | 0 | 1 | 0 |
| Overall | 99 | 80 | 41 |
| Overall percentages | 45.00 | 36.36 | 18.64 |

| CCHamide-1 | complete | partial | missing |
|---|---|---|---|
| Blattodea | 10 | 26 | 5 |
| Dermaptera | 1 | 6 | 9 |
| Embioptera | 3 | 0 | 2 |
| Grylloblattodea | 0 | 3 | 3 |
| Mantodea | 1 | 19 | 32 |
| Mantophasmatodea | 4 | 0 | 0 |
| Orthoptera: Caelifera | 1 | 12 | 14 |
| Orthoptera: Ensifera | 0 | 0 | 16 |
| Phasmatodea | 3 | 15 | 19 |
| Plecoptera | 2 | 8 | 5 |
| Zoraptera | 1 | 0 | 0 |
| Overall | 26 | 89 | 105 |
| Overall percentages | 11.82 | 40.45 | 47.73 |

| CCHamide-2 | complete | partial | missing |
|---|---|---|---|
| Blattodea | 22 | 16 | 3 |
| Dermaptera | 4 | 10 | 2 |
| Embioptera | 4 | 1 | 0 |
| Grylloblattodea | 4 | 2 | 0 |
| Mantodea | 16 | 21 | 15 |
| Mantophasmatodea | 4 | 0 | 0 |
| Orthoptera: Caelifera | 5 | 15 | 7 |
| Orthoptera: Ensifera | 0 | 9 | 7 |
| Phasmatodea | 16 | 17 | 4 |
| Plecoptera | 7 | 6 | 2 |
| Zoraptera | 1 | 0 | 0 |
| Overall | 83 | 97 | 40 |
| Overall percentages | 37.73 | 44.09 | 18.18 |

| CNMamide | complete | partial | missing |
|---|---|---|---|
| Blattodea | 29 | 8 | 4 |
| Dermaptera | 13 | 3 | 0 |
| Embioptera | 3 | 1 | 1 |
| Grylloblattodea | 3 | 3 | 0 |
| Mantodea | 46 | 3 | 3 |
| Mantophasmatodea | 4 | 0 | 0 |
| Orthoptera: Caelifera | 18 | 9 | 0 |
| Orthoptera: Ensifera | 8 | 5 | 3 |
| Phasmatodea | 20 | 13 | 4 |
| Plecoptera | 4 | 7 | 4 |
| Zoraptera | 1 | 0 | 0 |
| Overall | 149 | 52 | 19 |
| Overall percentages | 67.73 | 23.64 | 8.64 |

| Corazonin | complete | partial | missing |
|---|---|---|---|
| Blattodea | 17 | 21 | 3 |
| Dermaptera | 3 | 8 | 5 |
| Embioptera | 1 | 4 | 0 |
| Grylloblattodea | 1 | 4 | 1 |
| Mantodea | 5 | 20 | 27 |
| Mantophasmatodea | 4 | 0 | 0 |
| Orthoptera: Caelifera | 3 | 4 | 20 |
| Orthoptera: Ensifera | 6 | 5 | 5 |
| Phasmatodea | 10 | 8 | 19 |
| Plecoptera | 6 | 7 | 2 |
| Zoraptera | 0 | 0 | 1 |
| Overall | 56 | 81 | 83 |
| Overall percentages | 25.45 | 36.82 | 37.73 |

| CRF-DH | complete | partial | missing |
|---|---|---|---|
| Blattodea | 22 | 15 | 4 |
| Dermaptera | 5 | 9 | 2 |
| Embioptera | 4 | 1 | 0 |
| Grylloblattodea | 3 | 3 | 0 |
| Mantodea | 20 | 29 | 3 |
| Mantophasmatodea | 4 | 0 | 0 |
| Orthoptera: Caelifera | 3 | 12 | 12 |
| Orthoptera: Ensifera | 6 | 7 | 3 |
| Phasmatodea | 14 | 17 | 6 |
| Plecoptera | 4 | 7 | 4 |
| Zoraptera | 0 | 1 | 0 |
| Overall | 85 | 101 | 34 |
| Overall percentages | 38.64 | 45.91 | 15.45 |

| CT-DH | complete | partial | missing |
|---|---|---|---|
| Blattodea | 36 | 5 | 0 |
| Dermaptera | 9 | 6 | 1 |
| Embioptera | 4 | 1 | 0 |
| Grylloblattodea | 5 | 1 | 0 |
| Mantodea | 35 | 12 | 5 |
| Mantophasmatodea | 4 | 0 | 0 |
| Orthoptera: Caelifera | 14 | 6 | 7 |
| Orthoptera: Ensifera | 8 | 7 | 1 |
| Phasmatodea | 23 | 14 | 0 |
| Plecoptera | 7 | 6 | 2 |
| Zoraptera | 1 | 0 | 0 |
| Overall | 146 | 58 | 16 |
| Overall percentages | 66.36 | 26.36 | 7.27 |

| Elevenin | complete | partial | missing |
|---|---|---|---|
| Blattodea | 25 | 14 | 2 |
| Dermaptera | 5 | 10 | 1 |
| Embioptera | 2 | 1 | 2 |
| Grylloblattodea | 2 | 1 | 3 |
| Mantodea | 12 | 24 | 16 |
| Mantophasmatodea | 4 | 0 | 0 |
| Orthoptera: Caelifera | 5 | 9 | 13 |
| Orthoptera: Ensifera | 3 | 6 | 7 |
| Phasmatodea | 6 | 21 | 10 |
| Plecoptera | 5 | 8 | 2 |
| Zoraptera | 1 | 0 | 0 |
| Overall | 70 | 94 | 56 |
| Overall percentages | 31.82 | 42.73 | 25.45 |

| HanSolin | complete | partial | missing |
|---|---|---|---|
| Blattodea | 20 | 15 | 6 |
| Dermaptera | 10 | 5 | 1 |
| Embioptera | 2 | 2 | 1 |
| Grylloblattodea | 1 | 1 | 4 |
| Mantodea | 2 | 9 | 41 |
| Mantophasmatodea | 4 | 0 | 0 |
| Orthoptera: Caelifera | 16 | 7 | 4 |
| Orthoptera: Ensifera | 2 | 7 | 7 |
| Phasmatodea | 15 | 13 | 9 |
| Plecoptera | 7 | 6 | 2 |
| Zoraptera | 1 | 0 | 0 |
| Overall | 80 | 65 | 75 |
| Overall percentages | 36.36 | 29.55 | 34.09 |

| NPF-1 | complete | partial | missing |
|---|---|---|---|
| Blattodea | 40 | 1 | 0 |
| Dermaptera | 11 | 4 | 1 |
| Embioptera | 2 | 2 | 1 |
| Grylloblattodea | 5 | 1 | 0 |
| Mantodea | 40 | 7 | 5 |
| Mantophasmatodea | 4 | 0 | 0 |
| Orthoptera: Caelifera | 12 | 6 | 9 |
| Orthoptera: Ensifera | 7 | 4 | 5 |
| Phasmatodea | 15 | 12 | 10 |
| Plecoptera | 7 | 4 | 4 |
| Zoraptera | 1 | 0 | 0 |
| Overall | 144 | 41 | 35 |
| Overall percentages | 65.45 | 18.64 | 15.91 |

| NPF-2 | complete | partial | missing |
|---|---|---|---|
| Blattodea | 22 | 13 | 6 |
| Dermaptera | 12 | 4 | 0 |
| Embioptera | 4 | 0 | 1 |
| Grylloblattodea | 1 | 3 | 2 |
| Mantodea | 35 | 10 | 7 |
| Mantophasmatodea | 4 | 0 | 0 |
| Orthoptera: Caelifera | 12 | 8 | 7 |
| Orthoptera: Ensifera | 5 | 5 | 6 |
| Phasmatodea | 8 | 11 | 18 |
| Plecoptera | 7 | 7 | 1 |
| Zoraptera | 1 | 0 | 0 |
| Overall | 111 | 61 | 48 |
| Overall percentages | 50.45 | 27.73 | 21.82 |

| MS | complete | partial | missing |
|---|---|---|---|
| Blattodea | 32 | 7 | 2 |
| Dermaptera | 9 | 7 | 0 |
| Embioptera | 4 | 0 | 1 |
| Grylloblattodea | 3 | 1 | 2 |
| Mantodea | 25 | 17 | 10 |
| Mantophasmatodea | 4 | 0 | 0 |
| Orthoptera: Caelifera | 11 | 8 | 8 |
| Orthoptera: Ensifera | 9 | 5 | 2 |
| Phasmatodea | 21 | 12 | 4 |
| Plecoptera | 10 | 3 | 2 |
| Zoraptera | 1 | 0 | 0 |
| Overall | 129 | 60 | 31 |
| Overall percentages | 58.64 | 27.27 | 14.09 |

| **Proctolin** | **complete** | **partial** | **missing** |
|---|---|---|---|
| Blattodea | 35 | 5 | 1 |
| Dermaptera | 0 | 0 | 16 |
| Embioptera | 3 | 1 | 1 |
| Grylloblattodea | 6 | 0 | 0 |
| Mantodea | 45 | 5 | 2 |
| Mantophasmatodea | 4 | 0 | 0 |
| Orthoptera: Caelifera | 18 | 7 | 2 |
| Orthoptera: Ensifera | 9 | 3 | 4 |
| Phasmatodea | 22 | 4 | 11 |
| Plecoptera | 11 | 3 | 1 |
| Zoraptera | 1 | 0 | 0 |
| Overall | 154 | 28 | 38 |
| Overall percentages | 70.00 | 12.73 | 17.27 |

| **SIFamide** | **complete** | **partial** | **missing** |
|---|---|---|---|
| Blattodea | 25 | 4 | 12 |
| Dermaptera | 11 | 3 | 2 |
| Embioptera | 3 | 0 | 2 |
| Grylloblattodea | 6 | 0 | 0 |
| Mantodea | 10 | 11 | 31 |
| Mantophasmatodea | 4 | 0 | 0 |
| Orthoptera: Caelifera | 6 | 1 | 20 |
| Orthoptera: Ensifera | 6 | 7 | 3 |
| Phasmatodea | 8 | 12 | 17 |
| Plecoptera | 10 | 4 | 1 |
| Zoraptera | 1 | 0 | 0 |
| Overall | 90 | 42 | 88 |
| Overall percentages | 40.91 | 19.09 | 40.00 |

| SMYamide | complete | partial | missing |
|---|---|---|---|
| Blattodea | 9 | 4 | 28 |
| Dermaptera | 0 | 0 | 16 |
| Embioptera | 2 | 0 | 3 |
| Grylloblattodea | 0 | 0 | 6 |
| Mantodea | 19 | 4 | 29 |
| Mantophasmatodea | 0 | 0 | 4 |
| Orthoptera: Caelifera | 4 | 2 | 21 |
| Orthoptera: Ensifera | 7 | 1 | 8 |
| Phasmatodea | 10 | 2 | 25 |
| Plecoptera | 0 | 0 | 15 |
| Zoraptera | 0 | 0 | 1 |
| Overall | 51 | 13 | 156 |
| Overall percentages | 23.18 | 5.91 | 70.91 |

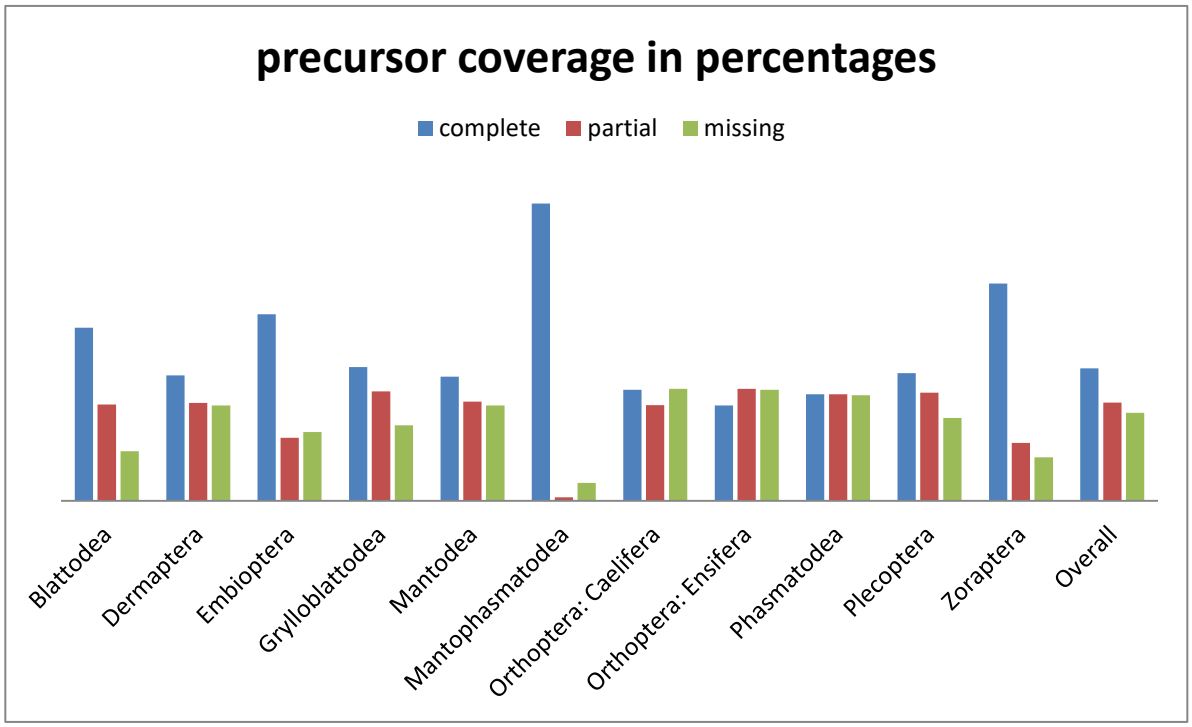
| sNPF | complete | partial | missing |
|---|---|---|---|
| Blattodea | 31 | 10 | 0 |
| Dermaptera | 13 | 3 | 0 |
| Embioptera | 2 | 2 | 1 |
| Grylloblattodea | 6 | 0 | 0 |
| Mantodea | 26 | 21 | 5 |
| Mantophasmatodea | 4 | 0 | 0 |
| Orthoptera: Caelifera | 14 | 9 | 4 |
| Orthoptera: Ensifera | 8 | 4 | 4 |
| Phasmatodea | 21 | 13 | 3 |
| Plecoptera | 12 | 1 | 2 |
| Zoraptera | 1 | 0 | 0 |
| Overall | 138 | 63 | 19 |
| Overall percentages | 62.73 | 28.64 | 8.64 |

| RFLamide | complete | partial | missing |
|---|---|---|---|
| Blattodea | 9 | 26 | 6 |
| Dermaptera | 1 | 9 | 6 |
| Embioptera | 3 | 0 | 2 |
| Grylloblattodea | 0 | 2 | 4 |
| Mantodea | 7 | 17 | 28 |
| Mantophasmatodea | 3 | 0 | 1 |
| Orthoptera: Caelifera | 7 | 13 | 7 |
| Orthoptera: Ensifera | 3 | 7 | 6 |
| Phasmatodea | 3 | 17 | 17 |
| Plecoptera | 1 | 5 | 9 |
| Zoraptera | 0 | 1 | 0 |
| Overall | 37 | 97 | 86 |
| Overall percentages | 16.82 | 44.09 | 39.09 |

| Trissin | complete | partial | missing |
|---|---|---|---|
| Blattodea | 8 | 17 | 16 |
| Dermaptera | 0 | 0 | 16 |
| Embioptera | 2 | 1 | 2 |
| Grylloblattodea | 0 | 3 | 3 |
| Mantodea | 3 | 25 | 24 |
| Mantophasmatodea | 4 | 0 | 0 |
| Orthoptera: Caelifera | 3 | 8 | 16 |
| Orthoptera: Ensifera | 3 | 3 | 10 |
| Phasmatodea | 2 | 5 | 30 |
| Plecoptera | 2 | 5 | 8 |
| Zoraptera | 0 | 0 | 1 |
| Overall | 27 | 67 | 126 |
| Overall percentages | 12.27 | 30.45 | 57.27 |

| Overall | complete | partial | missing |
|---|---|---|---|
| Blattodea | 489 | 272 | 140 |
| Dermaptera | 138 | 108 | 105 |
| Embioptera | 65 | 22 | 24 |
| Grylloblattodea | 55 | 45 | 31 |
| Mantodea | 441 | 352 | 339 |
| Mantophasmatodea | 82 | 1 | 5 |
| Orthoptera: Caelifera | 204 | 176 | 206 |
| Orthoptera: Ensifera | 104 | 122 | 121 |
| Phasmatodea | 269 | 269 | 266 |
| Plecoptera | 131 | 111 | 85 |
| Zoraptera | 15 | 4 | 3 |
| Overall | 1993 | 1482 | 1325 |
| Overall percentages | 41.52 | 30.88 | 27.60 |

| Overall % | complete | partial | missing |
|---|---|---|---|
| Blattodea | 54.27 | 30.19 | 15.54 |
| Dermaptera | 39.32 | 30.77 | 29.91 |
| Embioptera | 58.56 | 19.82 | 21.62 |
| Grylloblattodea | 41.98 | 34.35 | 23.66 |
| Mantodea | 38.96 | 31.10 | 29.95 |
| Mantophasmatodea | 93.18 | 1.14 | 5.68 |
| Orthoptera: Caelifera | 34.81 | 30.03 | 35.15 |
| Orthoptera: Ensifera | 29.97 | 35.16 | 34.87 |
| Phasmatodea | 33.46 | 33.46 | 33.08 |
| Plecoptera | 40.06 | 33.94 | 25.99 |
| Zoraptera | 68.18 | 18.18 | 13.64 |
| Overall | 41.52 | 30.88 | 27.60 |


precursor coverage in percentages
complete
partial
missing
Blattodea
Dermaptera
Embioptera
Grylloblattodea
Mantodea
Mantophasmatodea
Orthoptera: Caelifera
Orthoptera: Ensifera
Phasmatodea
Plecoptera
Zoraptera
Overall

**Additional File 2:** Alignments of all neuropeptide precursor sequences identified in this study, sorted by the different polyneopteran lineages.

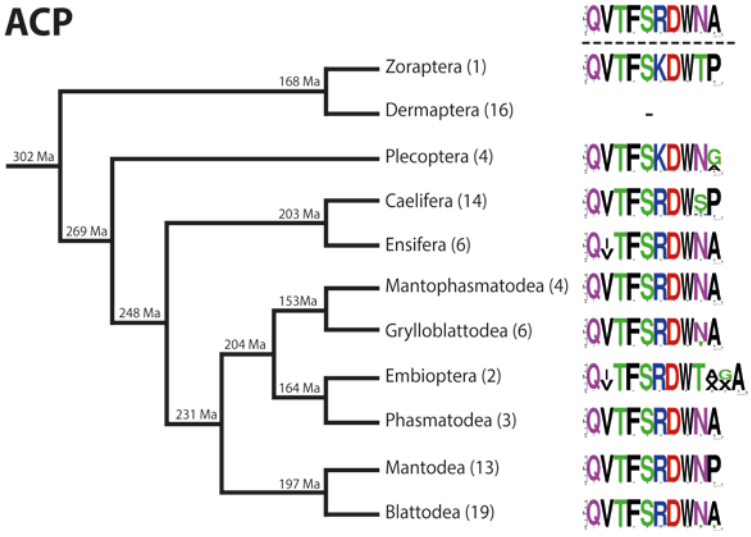
The alignments can be found as Data_Sheet_2 online at:

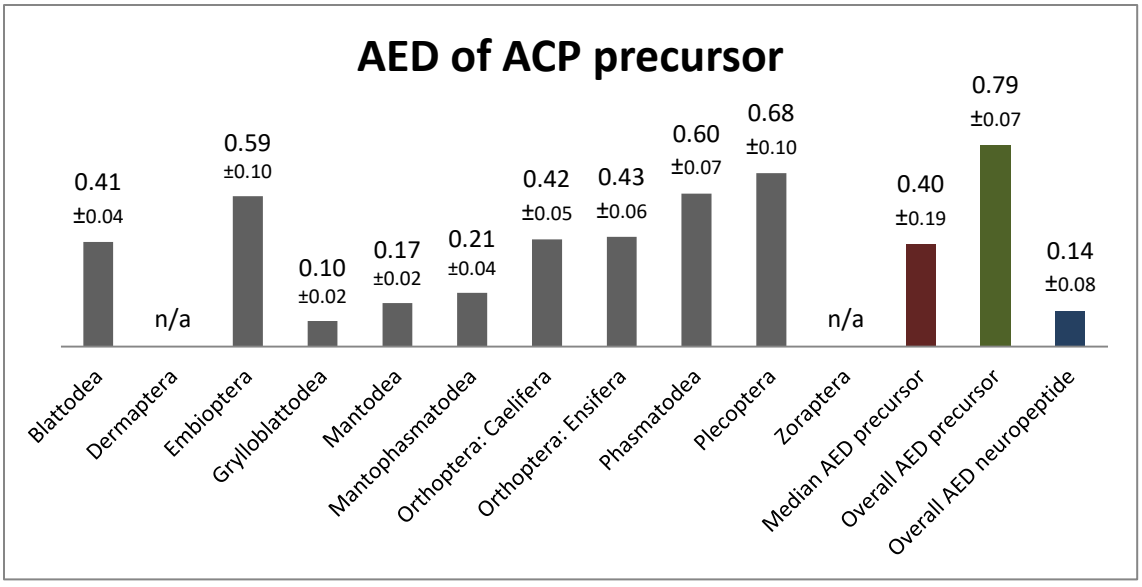
https://www.frontiersin.org/articles/10.3389/fendo.2020.00197/full#supplementary-material

**Additional File 3**: Statistical overview of the neuropeptide precursors in Polyneoptera with information on number of transcripts, sequence length, and position of neuropeptide sequences in the precursors (top), sequence logo representations showing the degree of amino acid sequence conservation of the neuropeptides (middle), and bar charts depicting the average evolutionary divergence (AED) of the neuropeptide precursors for the different polyneopteran lineages with standard error estimates (bottom). For the sequence logo representations, only the completely obtained precursor sequences were considered, the respective number is given in parentheses for each taxon. An “X” in the sequence represents a gap. The hypothetical ancestral state of the ACP sequence in Polyneoptera is listed at the top.

*A:*

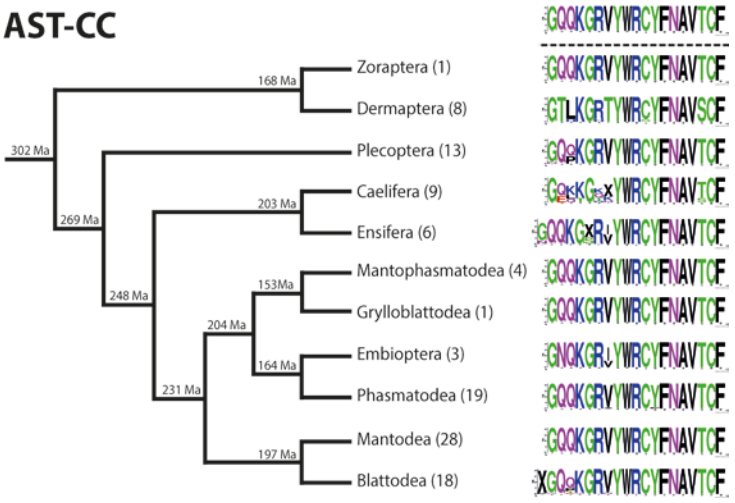
| ACP | presence | transcripts | position neuropeptide | length |
|---|---|---|---|---|
| Blattodea | + | 1 | N-terminal | 96-109 |
| Dermaptera | - | n/a | n/a | n/a |
| Embioptera | + | 1 | N-terminal | 109 |
| Grylloblattodea | + | 1 | N-terminal | 98-99 |
| Mantodea | + | 1 | N-terminal | n/a |
| Mantophasmatodea | + | 1 | N-terminal | 88-90 |
| Orthoptera: Caelifera | + | 1 | N-terminal | 85-92 |
| Orthoptera: Ensifera | + | 1 | N-terminal | 93-103 |
| Phasmatodea | + | 1 | N-terminal | 94 |
| Plecoptera | + | 1 | N-terminal | 106 |
| Zoraptera | + | 1 | N-terminal | n/a |

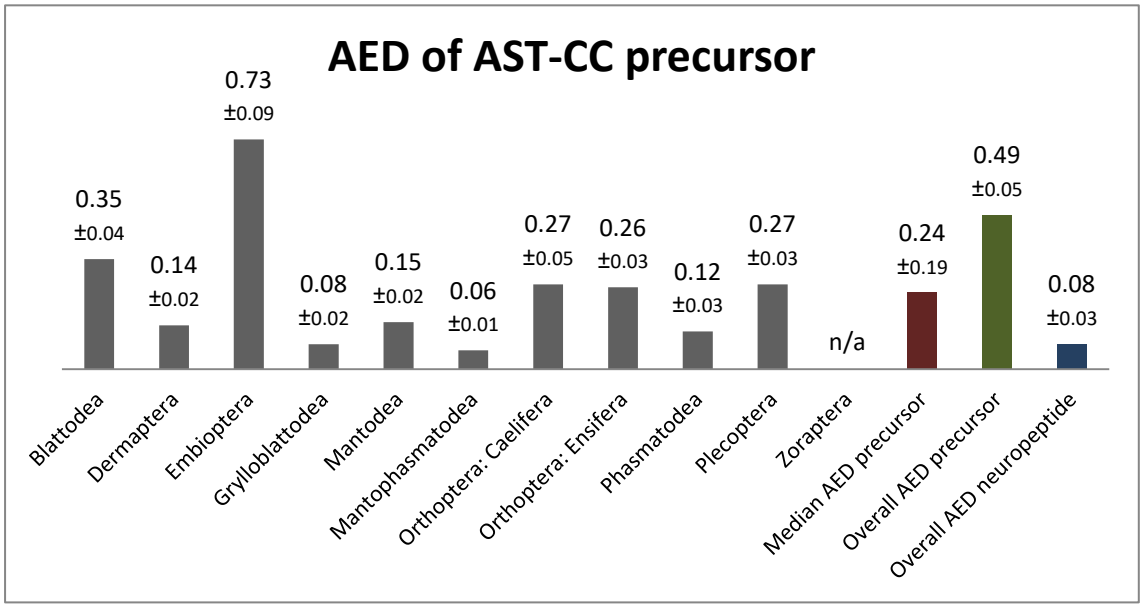


*B:*

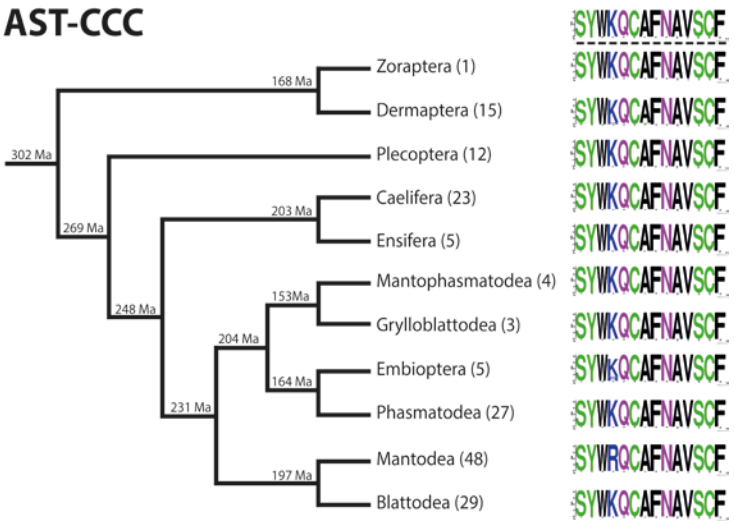
| AST-CC | presence | transcripts | position neuropeptide | length |
|---|---|---|---|---|
| Blattodea | + | 1 | C-terminal | 149-172 |
| Dermaptera | + | 1 | C-terminal | 114-116 |
| Embioptera | + | 1 | C-terminal | 130-131 |
| Grylloblattodea | + | 1 | C-terminal | 159 |
| Mantodea | + | 1 | C-terminal | 108-134 |
| Mantophasmatodea | + | 1 | C-terminal | 136-137 |
| Orthoptera: Caelifera | + | 1 | C-terminal | 142 |
| Orthoptera: Ensifera | + | 1 | C-terminal | 137 |
| Phasmatodea | + | 1 | C-terminal | 131-152 |
| Plecoptera | + | 1 | C-terminal | 133-153 |
| Zoraptera | + | 1 | C-terminal | 137 |

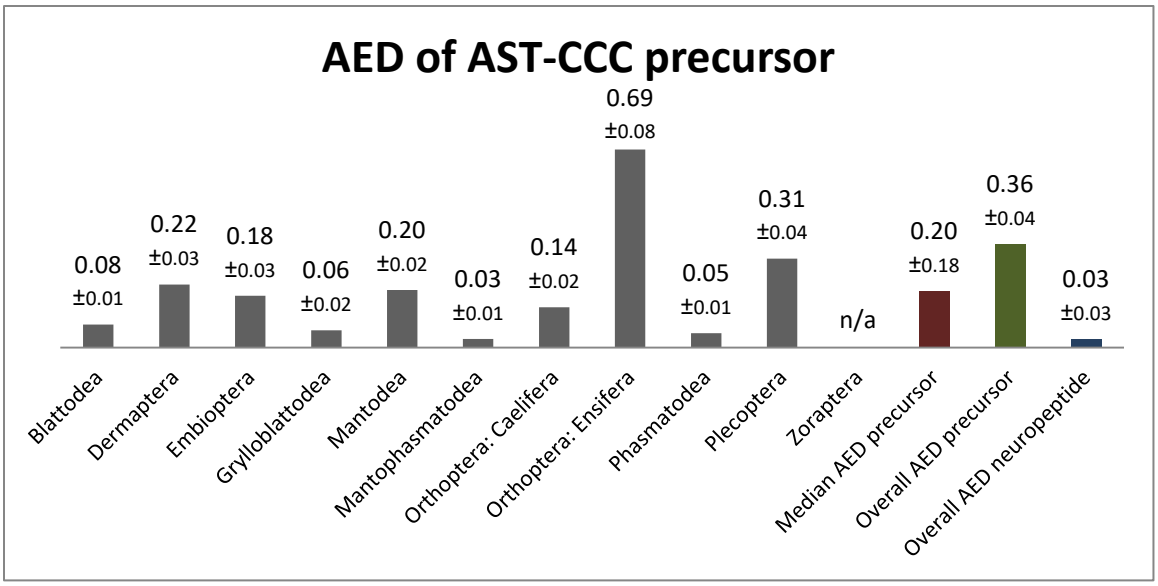


*C:*

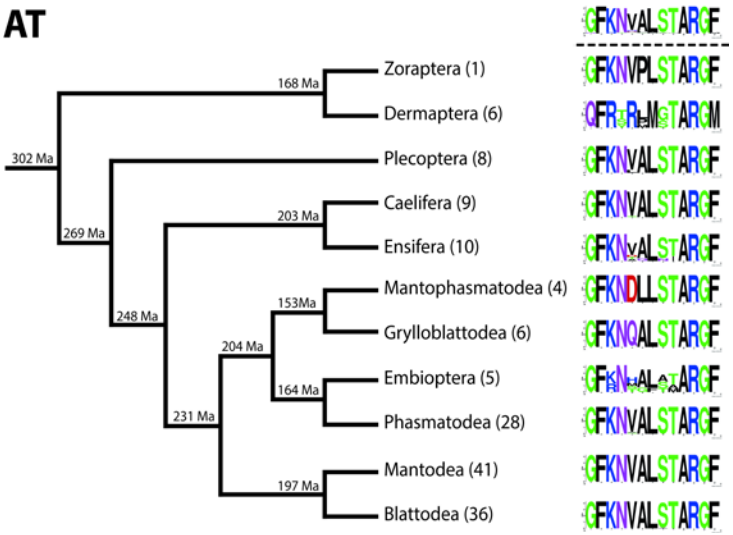
| AST-CCC | presence | transcripts | position neuropeptide | length |
|---|---|---|---|---|
| Blattodea | + | 1 | C-terminal | 96 |
| Dermaptera | + | 1 | C-terminal | 115-121 |
| Embioptera | + | 1 | C-terminal | 89-93 |
| Grylloblattodea | + | 1 | C-terminal | 96 |
| Mantodea | + | 1 | C-terminal | 94-103 |
| Mantophasmatodea | + | 1 | C-terminal | 97-98 |
| Orthoptera: Caelifera | + | 1 | C-terminal | 90-98 |
| Orthoptera: Ensifera | + | 1 | C-terminal | 97-101 |
| Phasmatodea | + | 1 | C-terminal | 97-101 |
| Plecoptera | + | 1 | C-terminal | 93-97 |
| Zoraptera | + | 1 | C-terminal | 96 |

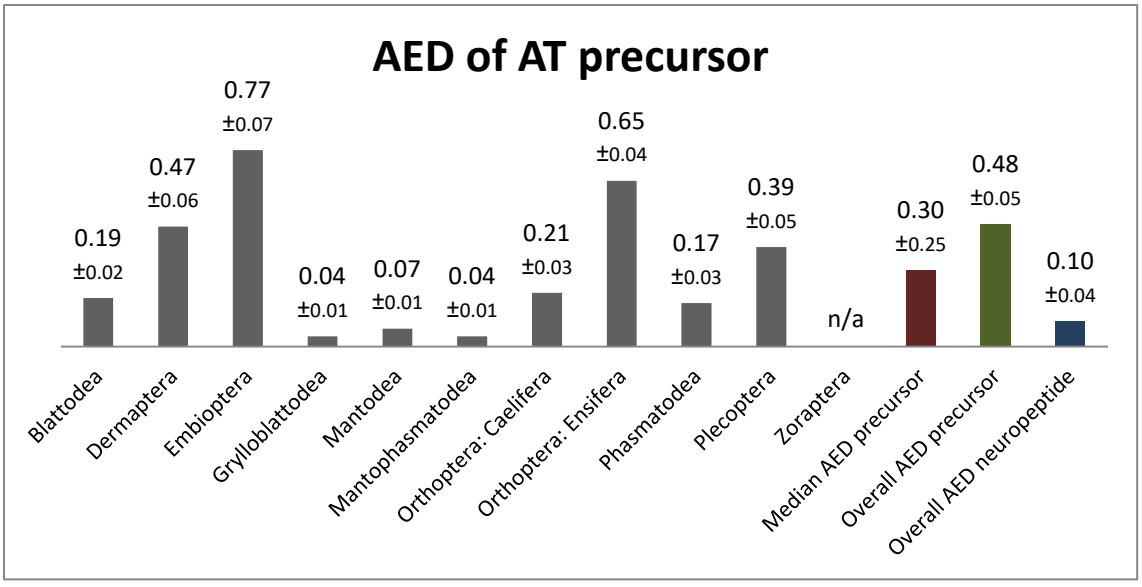


*D:*

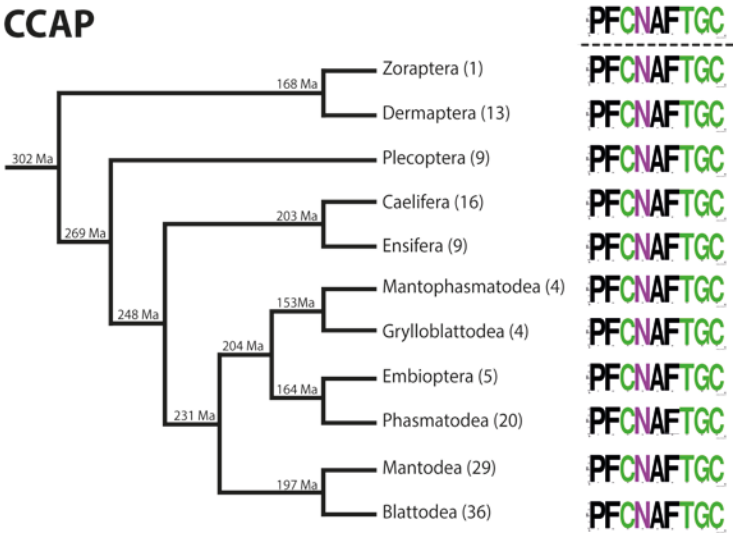
| AT | presence | transcripts | position neuropeptide | length |
|---|---|---|---|---|
| Blattodea | + | 1 | in the middle | 122-128 |
| Dermaptera | + | 1 | N-terminal | 113-123 |
| Embioptera | + | 2 | in the middle | 117-120/134-161 |
| Grylloblattodea | + | 1 | in the middle | 120 |
| Mantodea | + | 1 | in the middle | 108-118 |
| Mantophasmatodea | + | 1 | in the middle | 118-123 |
| Orthoptera: Caelifera | + | 1 | in the middle | 102-118 |
| Orthoptera: Ensifera | + | 1 | in the middle | 113-142 |
| Phasmatodea | + | 1 | in the middle | 117-120 |
| Plecoptera | + | 1 | in the middle | 121-124 |
| Zoraptera | + | 1 | in the middle | 119 |

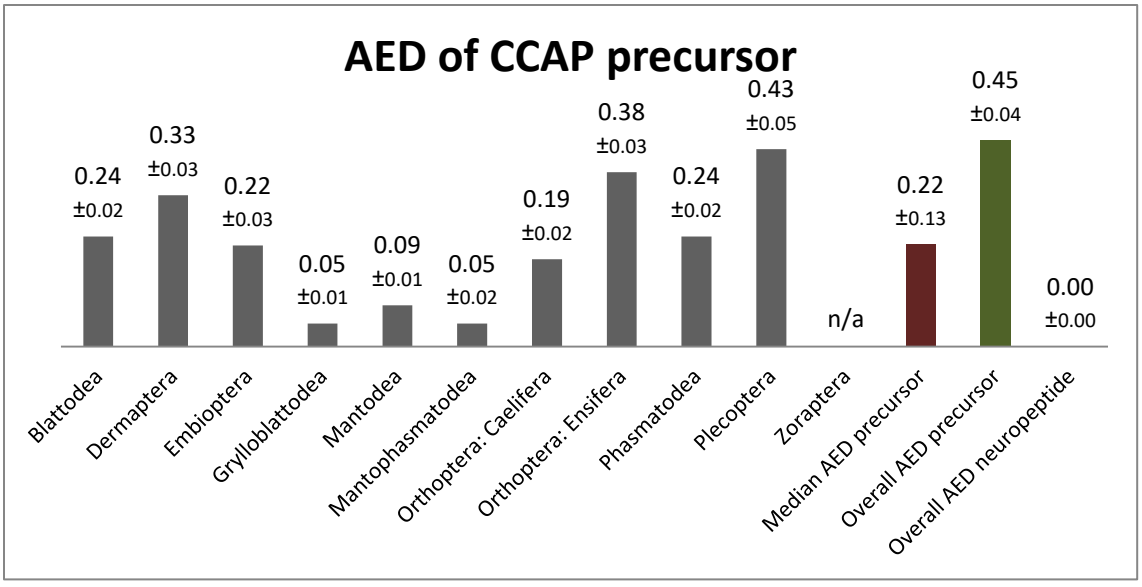


*E:*

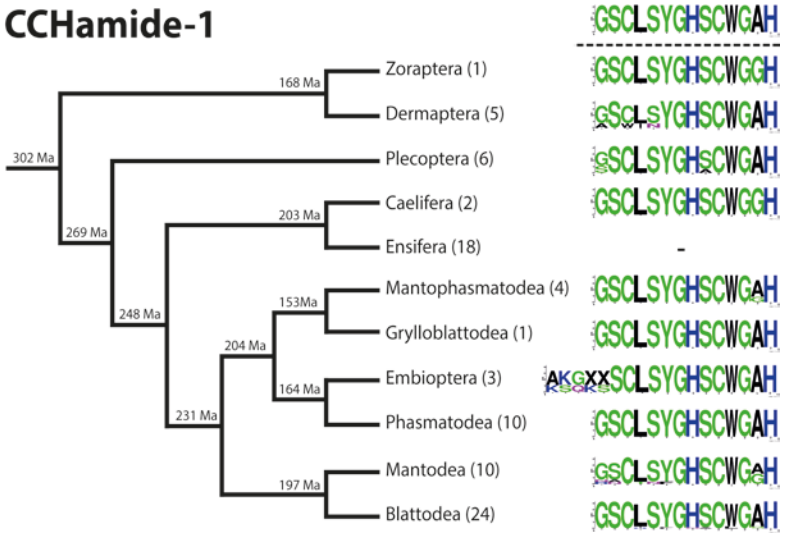
| CCAP | presence | transcripts | position neuropeptide | length |
|---|---|---|---|---|
| Blattodea | + | 1 | in the middle | 154-174 |
| Dermaptera | + | 1 | in the middle | 150-159 |
| Embioptera | + | 1 | in the middle | 161-166 |
| Grylloblattodea | + | 1 | in the middle | 161 |
| Mantodea | + | 1 | in the middle | 150-158 |
| Mantophasmatodea | + | 1 | in the middle | 148 |
| Orthoptera: Caelifera | + | 1 | in the middle | 143-156 |
| Orthoptera: Ensifera | + | 1 | in the middle | 150-163 |
| Phasmatodea | + | 1 | in the middle | 146-153 |
| Plecoptera | + | 1 | in the middle | 152-167 |
| Zoraptera | + | 1 | in the middle | 148 |

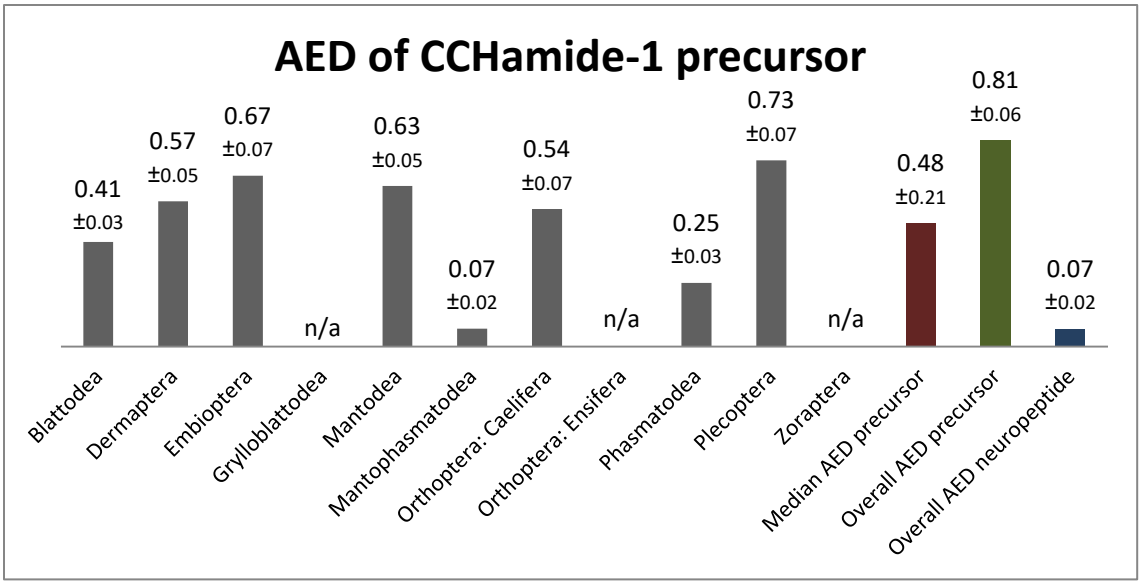


F:

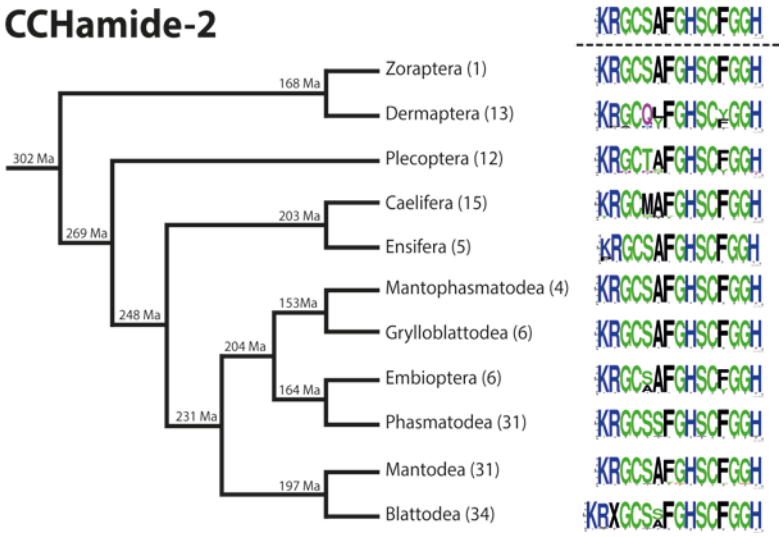
| CCHamide-1 | presence | transcripts | position neuropeptide | length |
|---|---|---|---|---|
| Blattodea | + | 1 | N-terminal | 184-206 |
| Dermaptera | + | 1 | N-terminal | 241 |
| Embioptera | + | 1 | N-terminal | 145-153 |
| Grylloblattodea | + | 1 | N-terminal | n/a |
| Mantodea | + | 1 | N-terminal | 209 |
| Mantophasmatodea | + | 1 | N-terminal | 115-116 |
| Orthoptera: Caelifera | + | 1 | N-terminal | 162-163 |
| Orthoptera: Ensifera | - | n/a | n/a | n/a |
| Phasmatodea | + | 1 | N-terminal | 160-161 |
| Plecoptera | + | 1 | N-terminal | 121-128 |
| Zoraptera | + | 1 | N-terminal | 170 |

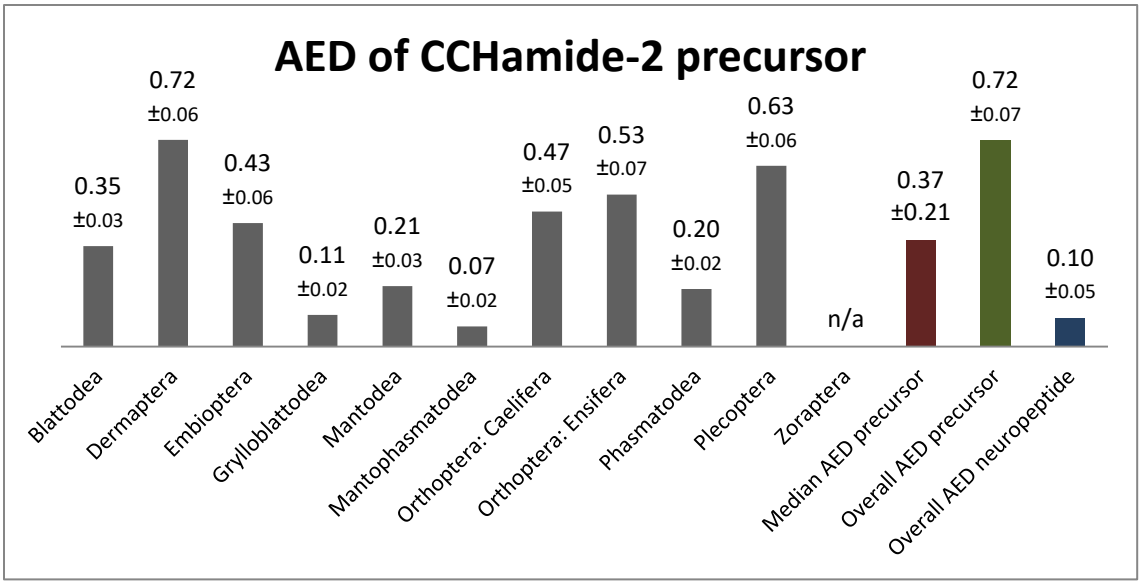


# G:

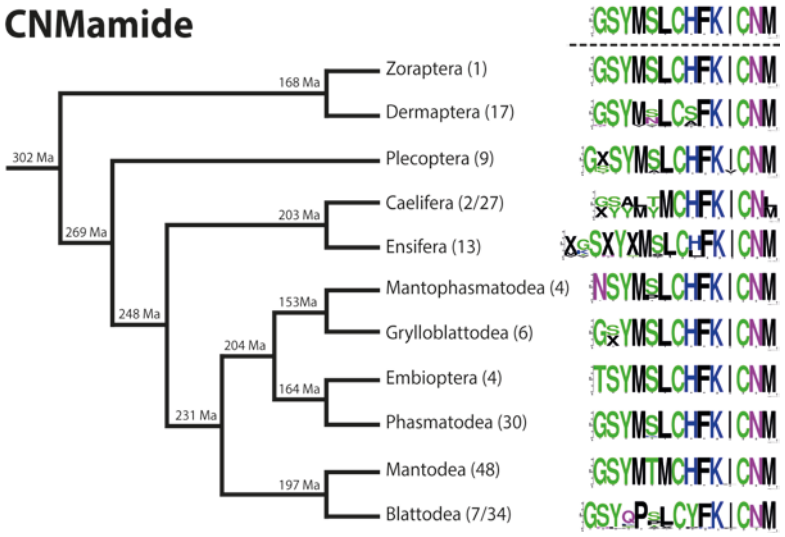
| CCHamide-2 | presence | transcripts | position neuropeptide | length |
|---|---|---|---|---|
| Blattodea | + | 1 | N-terminal | 124-142 |
| Dermaptera | + | 1 | N-terminal | 156-174 |
| Embioptera | + | 2 | N-terminal | 108-118/ n/a |
| Grylloblattodea | + | 1 | N-terminal | 125-132 |
| Mantodea | + | 1 | N-terminal | 97-104 |
| Mantophasmatodea | + | 1 | N-terminal | 103 |
| Orthoptera: Caelifera | + | 2 | N-terminal | 97-122/128-131 |
| Orthoptera: Ensifera | + | 1 | N-terminal | n/a |
| Phasmatodea | + | 1 | N-terminal | 105-113 |
| Plecoptera | + | 1 | N-terminal | 111-120 |
| Zoraptera | + | 1 | N-terminal | 120 |

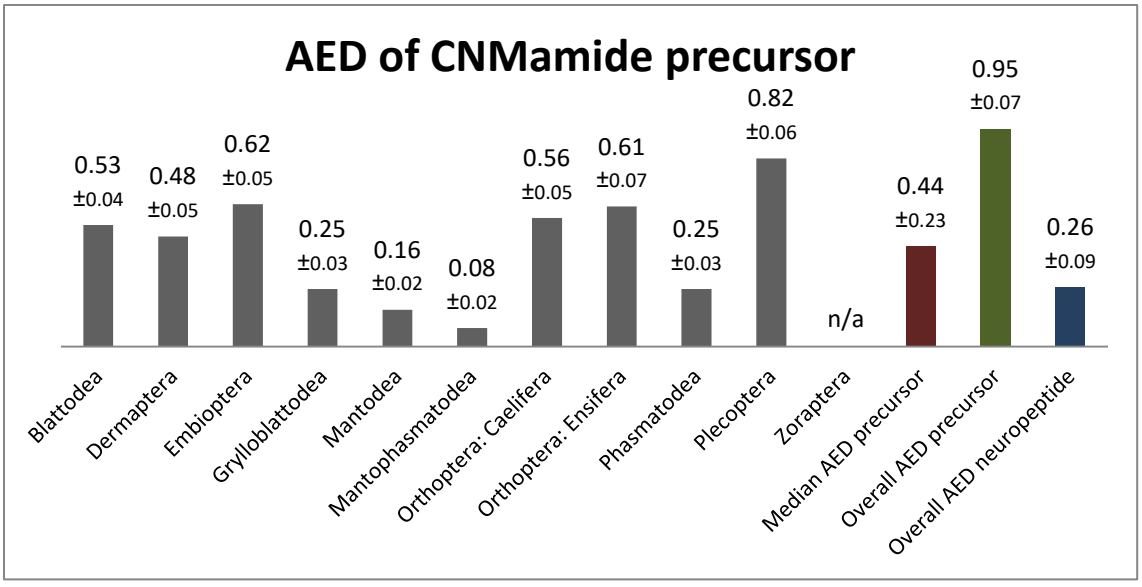


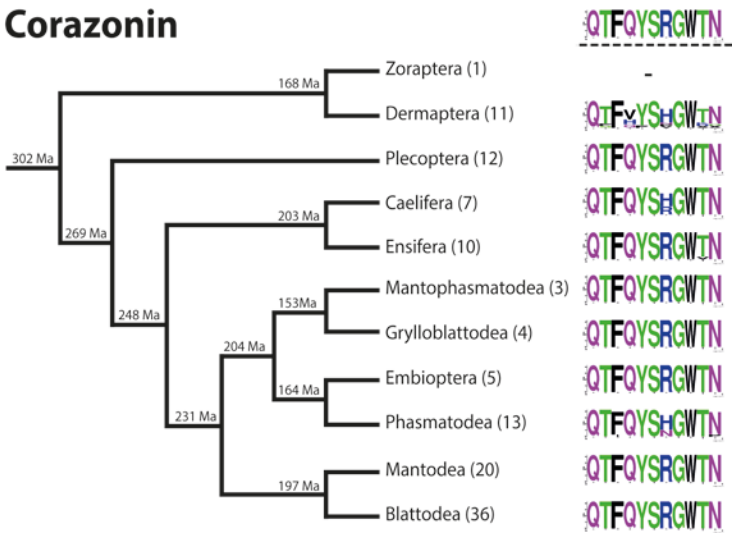
# H:

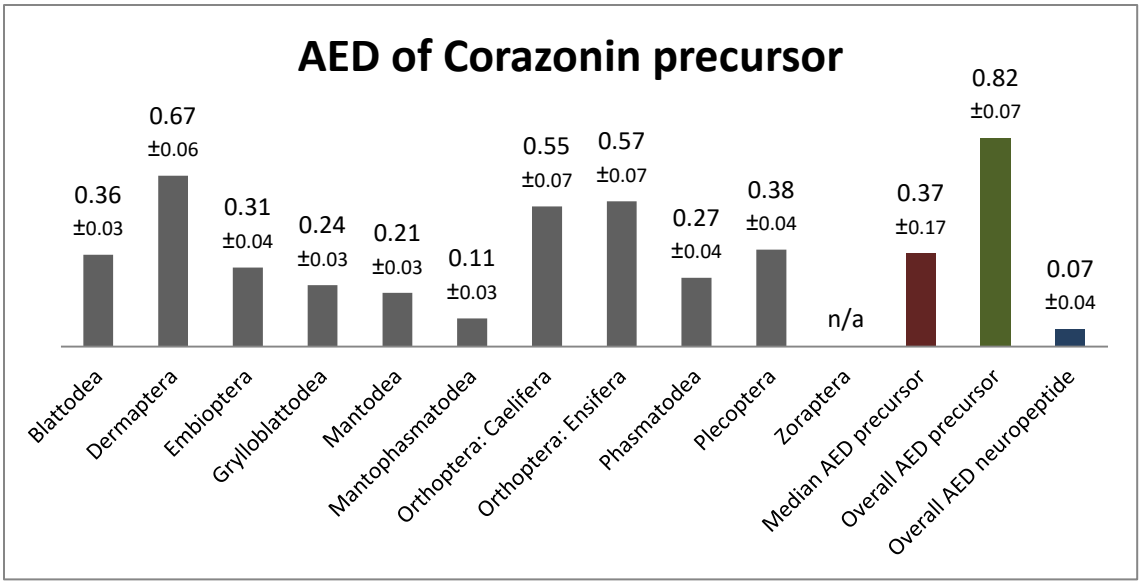
| CNMamide | presence | transcripts | position neuropeptide | length |
|---|---|---|---|---|
| Blattodea | + | 2 | C-terminal | 158/147-165 |
| Dermaptera | + | 1 | C-terminal | 127-158 |
| Embioptera | + | 1 | C-terminal | 152-155 |
| Grylloblattodea | + | 1 | C-terminal | 167-177 |
| Mantodea | + | 1 | C-terminal | 143-165 |
| Mantophasmatodea | + | 1 | C-terminal | 131-132 |
| Orthoptera: Caelifera | + | 2 | C-terminal | n/a /165-193 |
| Orthoptera: Ensifera | + | 1 | C-terminal | 127-157 |
| Phasmatodea | + | 1 | C-terminal | 133-150 |
| Plecoptera | + | 1 | C-terminal | 147-179 |
| Zoraptera | + | 1 | C-terminal | 157 |



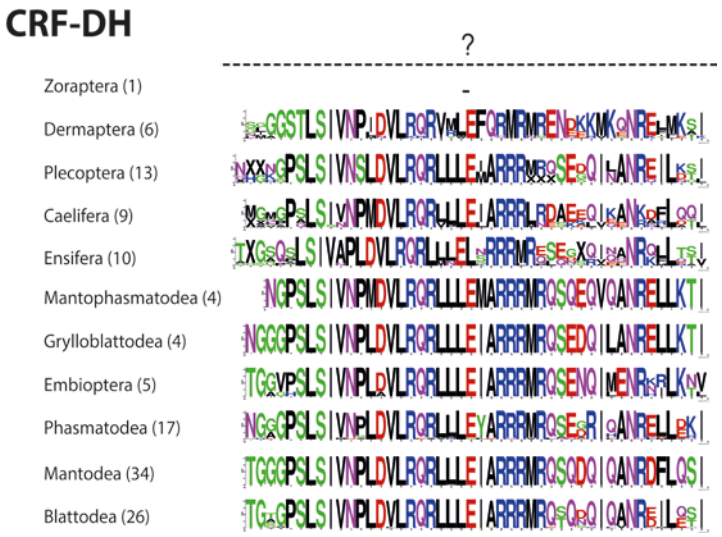



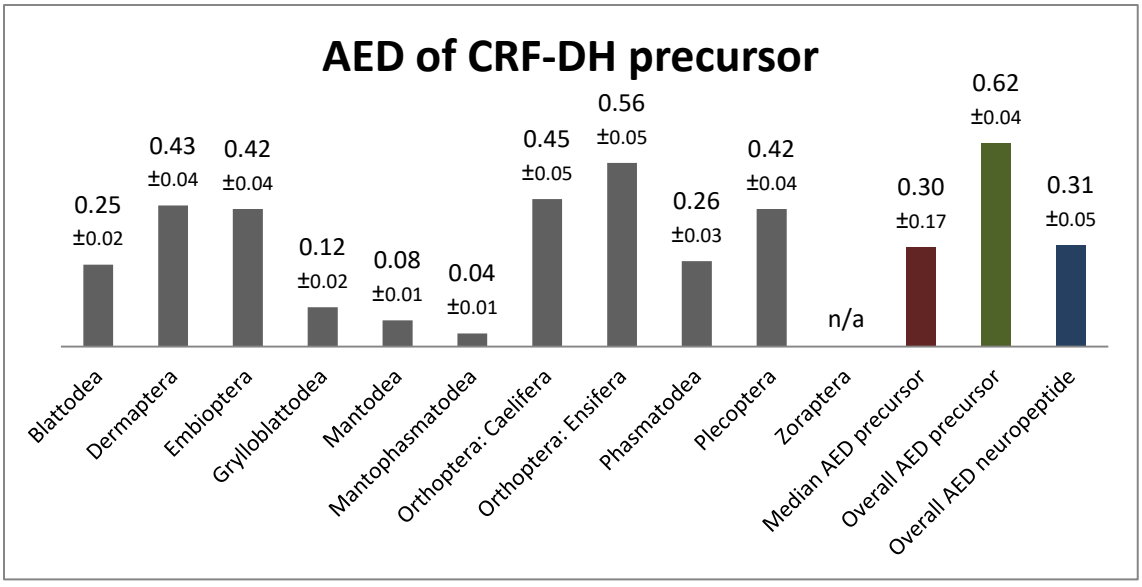
| Corazonin | presence | transcripts | position neuropeptide | length |
|---|---|---|---|---|
| Blattodea | + | 1 | N-terminal | 123-140 |
| Dermaptera | + | 1 | N-terminal | 102-109 |
| Embioptera | + | 1 | N-terminal | 118 |
| Grylloblattodea | + | 1 | N-terminal | 140 |
| Mantodea | + | 1 | N-terminal | 97 |
| Mantophasmatodea | + | 1 | N-terminal | 116-117 |
| Orthoptera: Caelifera | + | 1 | N-terminal | 94-133 |
| Orthoptera: Ensifera | + | 2 | N-terminal | 93/99-125 |
| Phasmatodea | + | 2 | N-terminal | 88/85-92 |
| Plecoptera | + | 1 | N-terminal | 108-128 |
| Zoraptera | - | n/a | n/a | n/a |



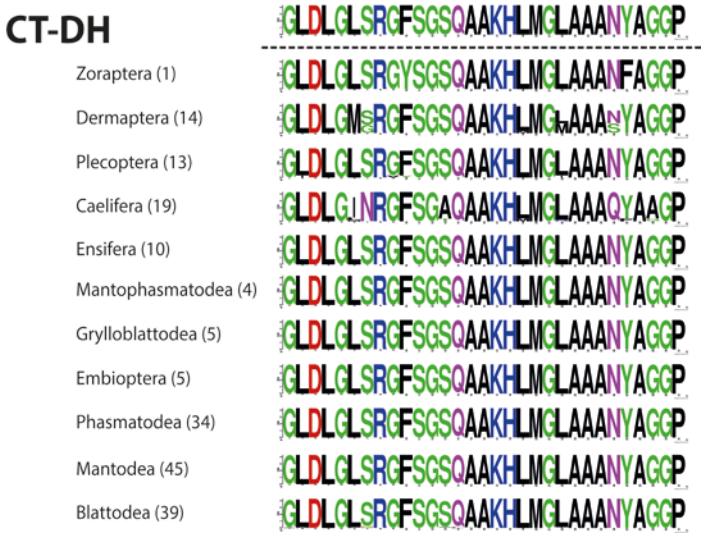



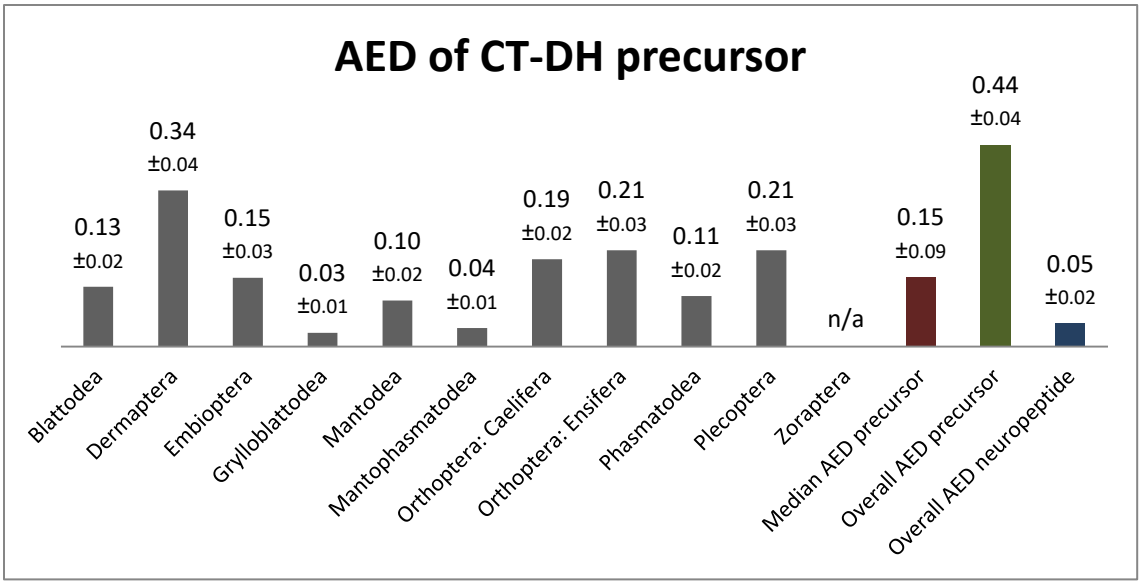
| CRF-DH | presence | transcripts | position neuropeptide | length |
|---|---|---|---|---|
| Blattodea | + | 1 | in the middle | 189-203 |
| Dermaptera | + | 1 | in the middle | 189-284 |
| Embioptera | + | 1 | in the middle | 195-201 |
| Grylloblattodea | + | 1 | in the middle | 192-195 |
| Mantodea | + | 1 | in the middle | 194-195 |
| Mantophasmatodea | + | 1 | in the middle | 156-157 |
| Orthoptera: Caelifera | + | 1 | in the middle | 139-190 |
| Orthoptera: Ensifera | + | 1 | in the middle | 138-172 |
| Phasmatodea | + | 1 | in the middle | 164-178 |
| Plecoptera | + | 1 | in the middle | 174-185 |
| Zoraptera | + | 1 | n/a | n/a |



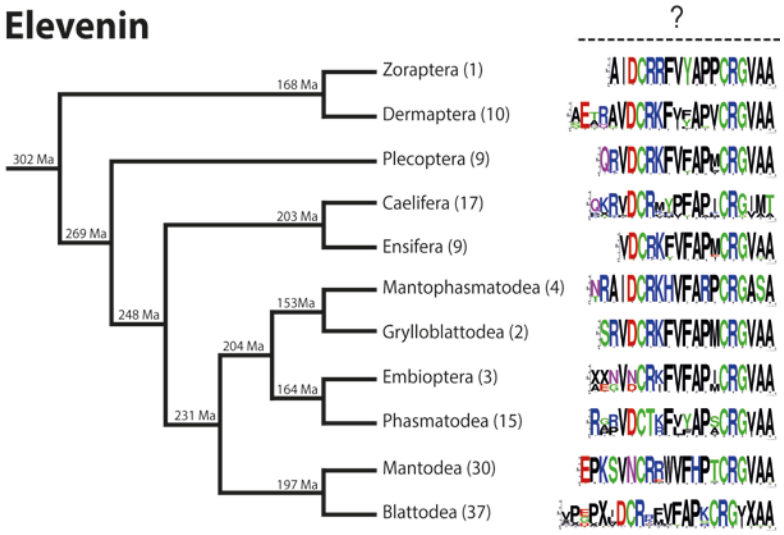
| CT-DH | presence | transcripts | position neuropeptide | length |
|---|---|---|---|---|
| Blattodea | + | 1 | C-terminal | 114-117 |
| Dermaptera | + | 1 | C-terminal | 124-178 |
| Embioptera | + | 1 | C-terminal | 110-114 |
| Grylloblattodea | + | 1 | C-terminal | 113 |
| Mantodea | + | 1 | C-terminal | 115-119 |
| Mantophasmatodea | + | 1 | C-terminal | 119 |
| Orthoptera: Caelifera | + | 1 | C-terminal | 107-128 |
| Orthoptera: Ensifera | + | 1 | C-terminal | 114-116 |
| Phasmatodea | + | 1 | C-terminal | 103-109 |
| Plecoptera | + | 1 | C-terminal | 112-117 |
| Zoraptera | + | 1 | C-terminal | 113 |

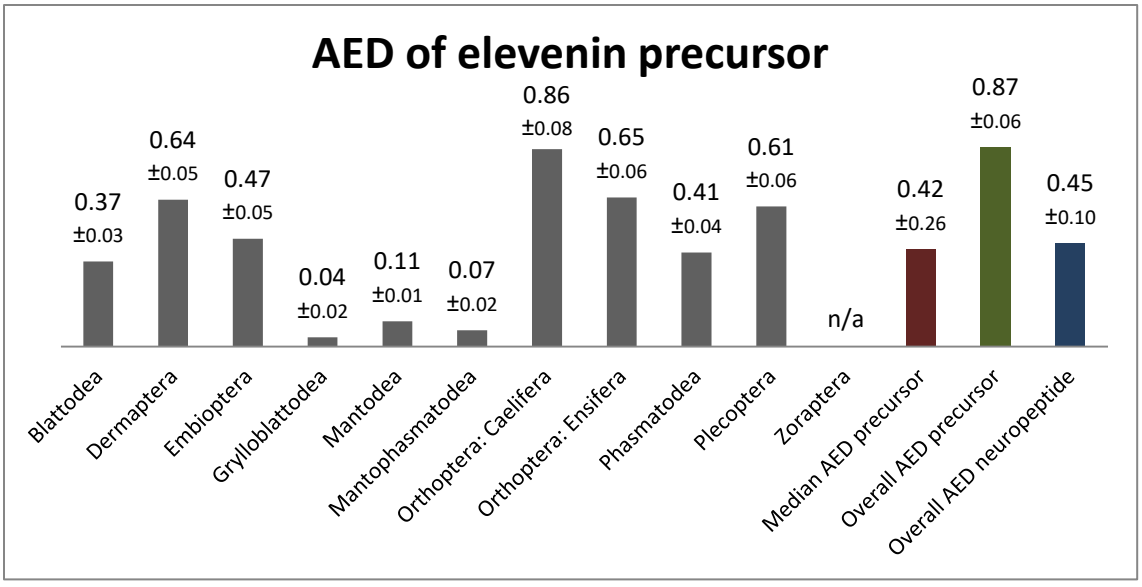


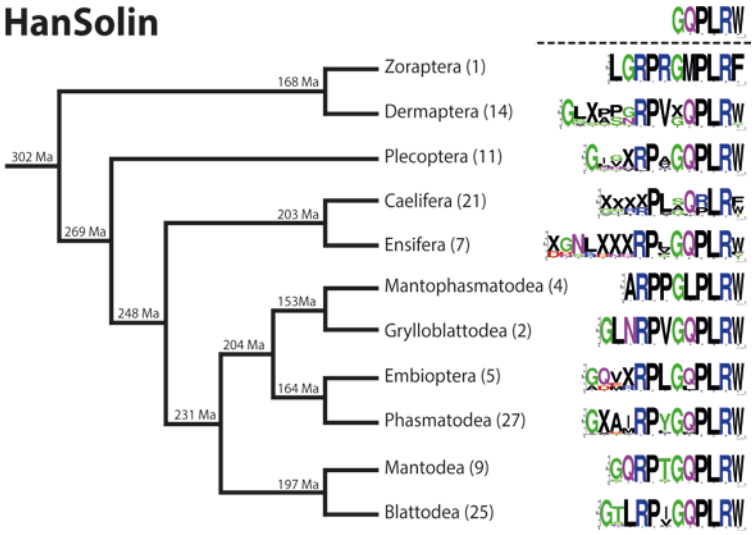
*L:*

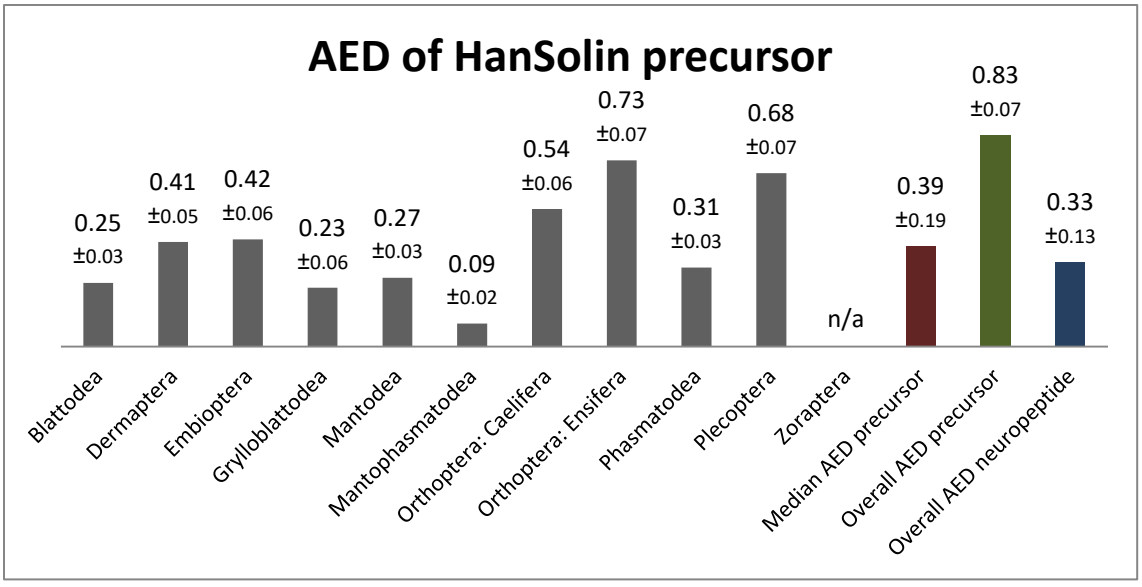
| Elevenin | presence | transcripts | position neuropeptide | length |
|---|---|---|---|---|
| Blattodea | + | 2 | N-terminal | 137-145 / 168 |
| Dermaptera | + | 1 | N-terminal | 166-170 |
| Embioptera | + | 1 | N-terminal | 147 |
| Grylloblattodea | + | 1 | N-terminal | 130/132 |
| Mantodea | + | 1 | N-terminal | 138-140 |
| Mantophasmatodea | + | 1 | N-terminal | 115 |
| Orthoptera: Caelifera | + | 1 | N-terminal | 99-113 |
| Orthoptera: Ensifera | + | 1 | N-terminal | 125-132 |
| Phasmatodea | + | 1 | N-terminal | 120-129 |
| Plecoptera | + | 1 | N-terminal | 121-144 |
| Zoraptera | + | 1 | N-terminal | 151 |



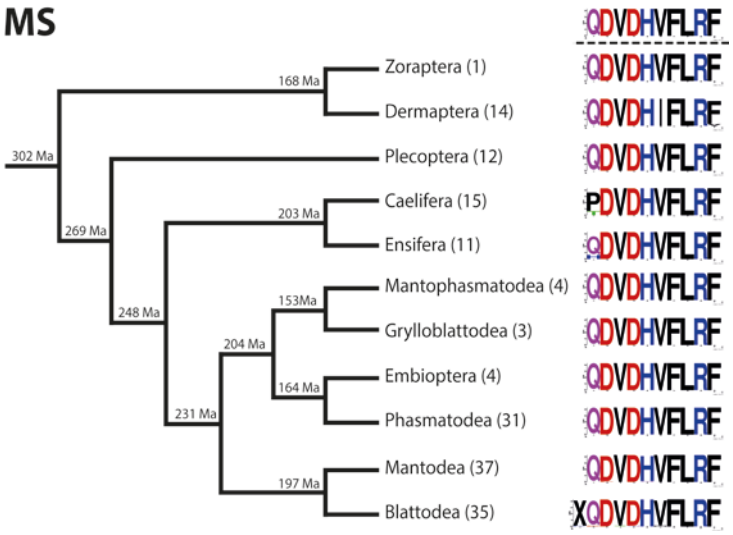
| HanSolin | presence | transcripts | position neuropeptide | length |
|---|---|---|---|---|
| Blattodea | + | 1 | C-terminal | 108-130 |
| Dermaptera | + | 1 | C-terminal | 113-139 |
| Embioptera | + | 1 | C-terminal | 129-139 |
| Grylloblattodea | + | 1 | C-terminal | n/a |
| Mantodea | + | 1 | C-terminal | 122-125 |
| Mantophasmatodea | + | 1 | C-terminal | 88-89 |
| Orthoptera: Caelifera | + | 1 | C-terminal | 101-117 |
| Orthoptera: Ensifera | + | 1 | C-terminal | 121-138 |
| Phasmatodea | + | 1 | C-terminal | 118-126 |
| Plecoptera | + | 1 | C-terminal | 125-133 |
| Zoraptera | + | 1 | C-terminal | 127 |

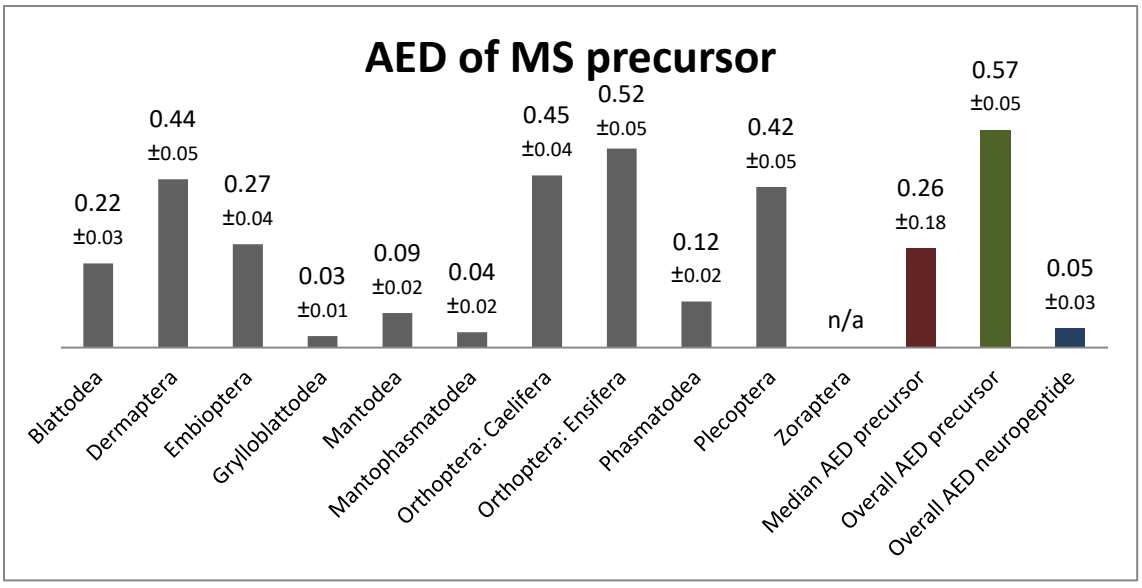


# N:

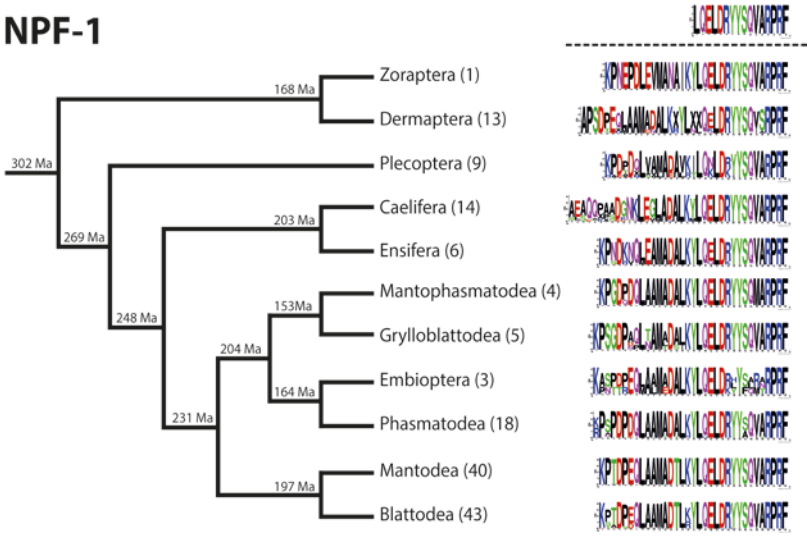
| MS | presence | transcripts | position neuropeptide | length |
|---|---|---|---|---|
| Blattodea | + | 1 | C-terminal | 89-101 |
| Dermaptera | + | 2 | C-terminal | 109/107-110 |
| Embioptera | + | 1 | C-terminal | 101-104 |
| Grylloblattodea | + | 1 | C-terminal | 100 |
| Mantodea | + | 2 | C-terminal | 90-97/133-138 |
| Mantophasmatodea | + | 1 | C-terminal | 96 |
| Orthoptera: Caelifera | + | 2 | C-terminal | 96/84-99 |
| Orthoptera: Ensifera | + | 1 | C-terminal | 84-100 |
| Phasmatodea | + | 1 | C-terminal | 95-101 |
| Plecoptera | + | 1 | C-terminal | 98-103 |
| Zoraptera | + | 1 | C-terminal | 95 |

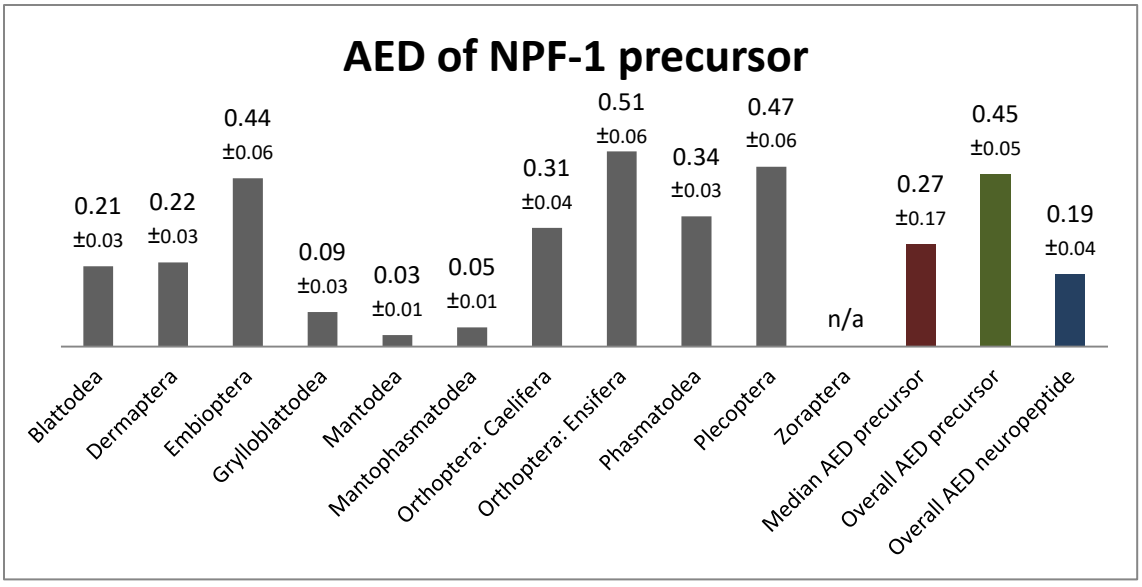


*O:*

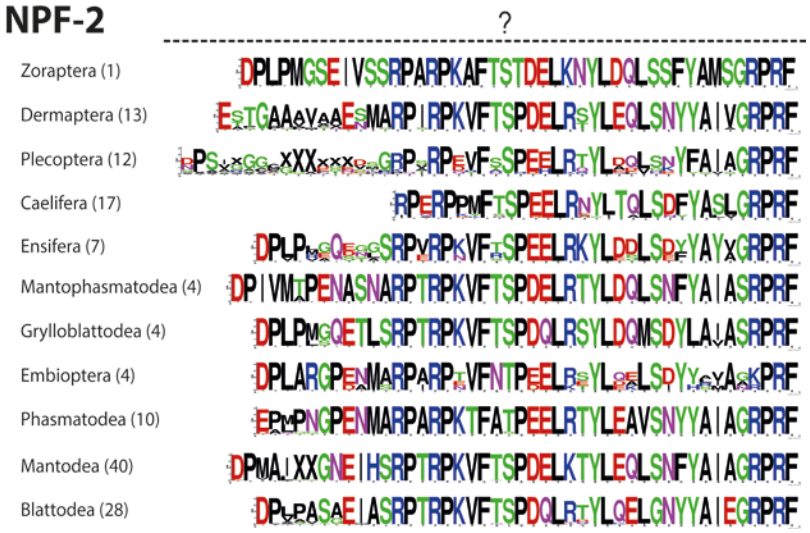
| NPF-1 | presence | transcripts | position neuropeptide | length |
|---|---|---|---|---|
| Blattodea | + | 1 | N-terminal | 85-91 |
| Dermaptera | + | 1 | N-terminal | 81-85 |
| Embioptera | + | 2 | N-terminal | 95/131 |
| Grylloblattodea | + | 2 | N-terminal | 95/ n/a |
| Mantodea | + | 2 | N-terminal | 88/125 |
| Mantophasmatodea | + | 1 | N-terminal | 83 |
| Orthoptera: Caelifera | + | 1 | N-terminal | 86-95 |
| Orthoptera: Ensifera | + | 1 | N-terminal | 90-94 |
| Phasmatodea | + | 1 | N-terminal | 87-97 |
| Plecoptera | + | 1 | N-terminal | 85-87 |
| Zoraptera | + | 1 | N-terminal | 81 |

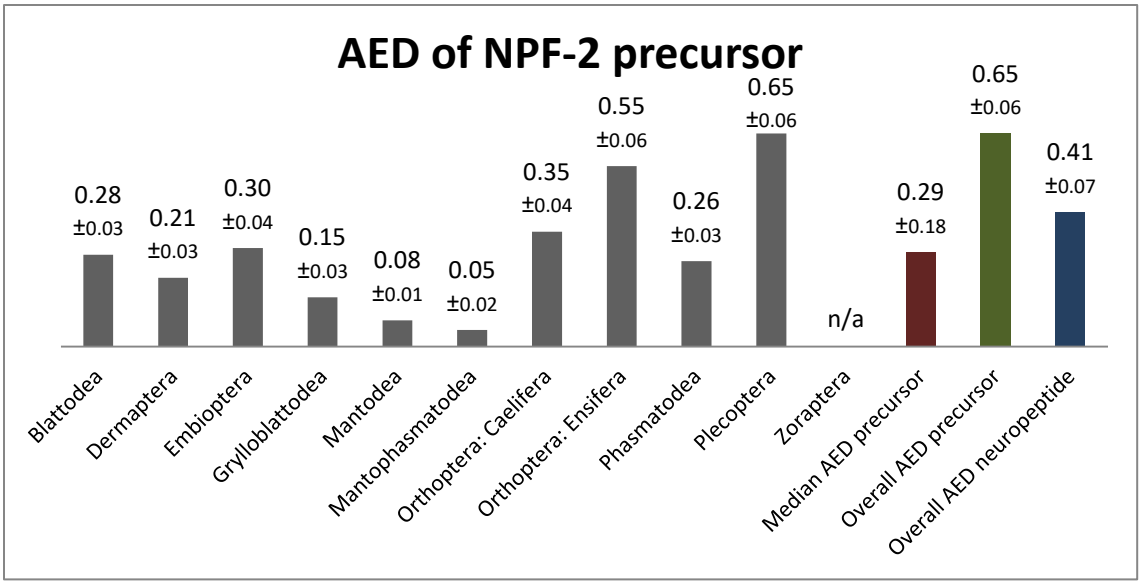


*P:*

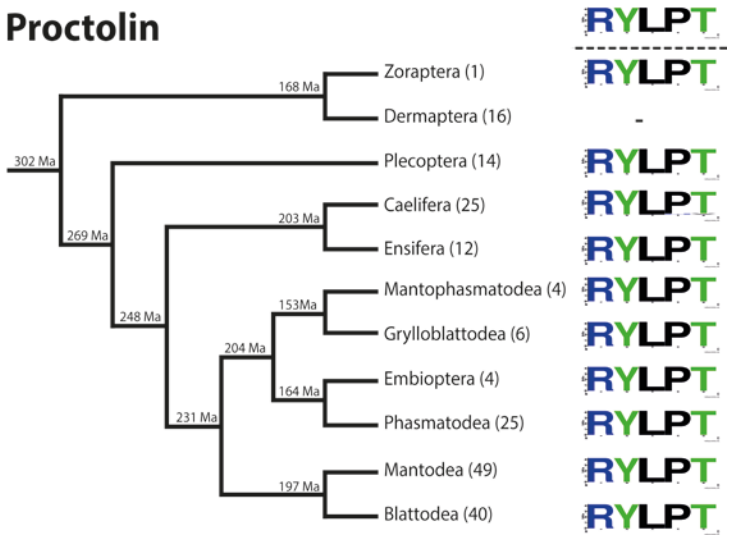
| NPF-2 | presence | transcripts | position neuropeptide | length |
|---|---|---|---|---|
| Blattodea | + | 1 | N-terminal | 116-128 |
| Dermaptera | + | 1 | N-terminal | 121-134 |
| Embioptera | + | 1 | N-terminal | 113-117 |
| Grylloblattodea | + | 1 | N-terminal | 114 |
| Mantodea | + | 1 | N-terminal | 107-123 |
| Mantophasmatodea | + | 1 | N-terminal | 116 |
| Orthoptera: Caelifera | + | 1 | N-terminal | 85-107 |
| Orthoptera: Ensifera | + | 1 | N-terminal | 112-120 |
| Phasmatodea | + | 1 | N-terminal | 114-121 |
| Plecoptera | + | 1 | N-terminal | 124-130 |
| Zoraptera | + | 1 | N-terminal | 118 |

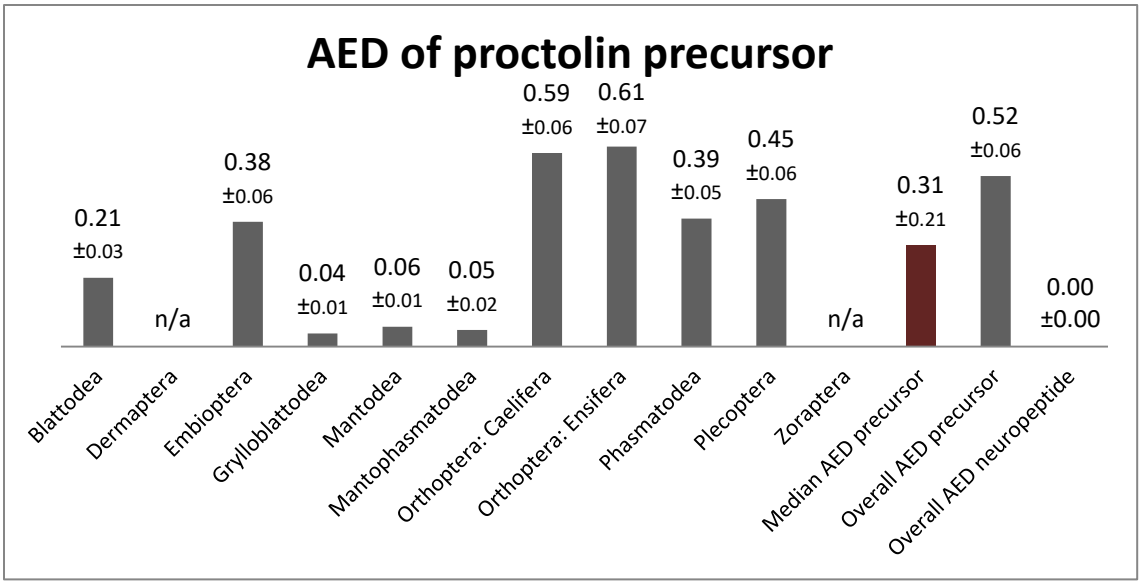


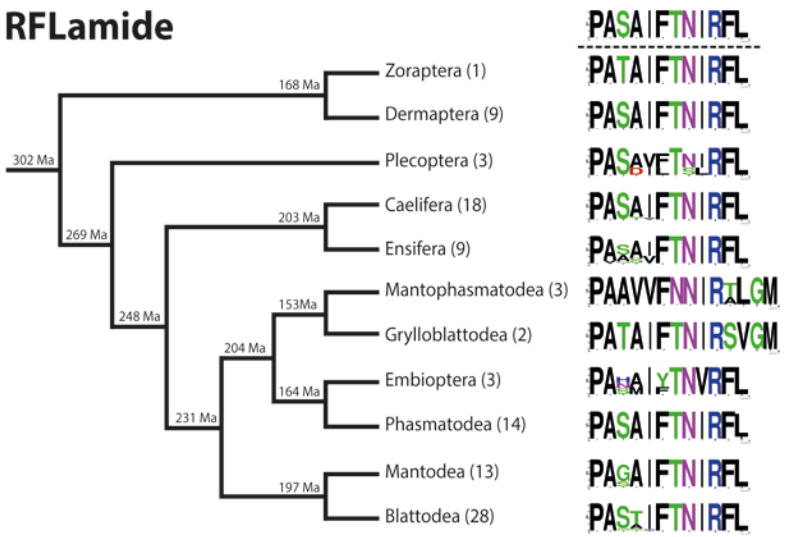
Q:

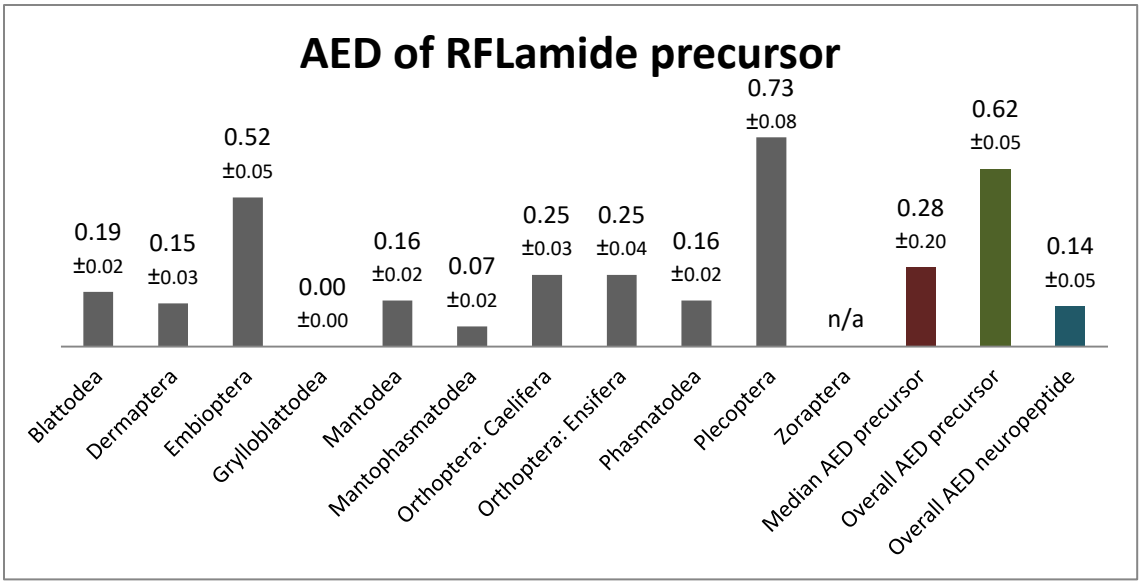
| Proctolin | presence | transcripts | position neuropeptide | length |
|---|---|---|---|---|
| Blattodea | + | 1 | N-terminal | 80-93 |
| Dermaptera | - | n/a | n/a | n/a |
| Embioptera | + | 1 | N-terminal | 83-84 |
| Grylloblattodea | + | 1 | N-terminal | 83-84 |
| Mantodea | + | 1 | N-terminal | 79-83 |
| Mantophasmatodea | + | 1 | N-terminal | 86-87 |
| Orthoptera: Caelifera | + | 1 | N-terminal | 74-86 |
| Orthoptera: Ensifera | + | 1 | N-terminal | 76-104 |
| Phasmatodea | + | 1 | N-terminal | 80-87 |
| Plecoptera | + | 1 | N-terminal | 80-90 |
| Zoraptera | + | 2 | N-terminal | 89/ n/a |



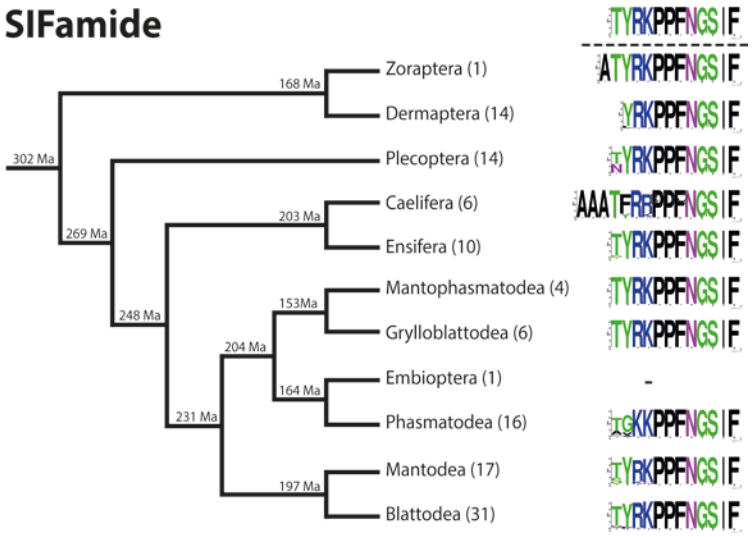



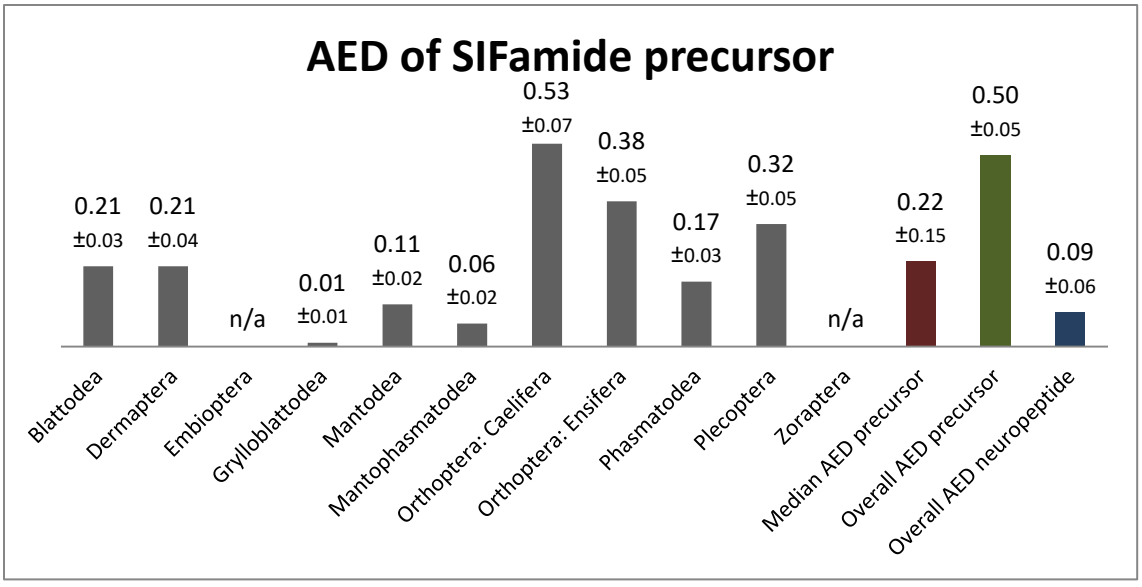
| RFLamide | presence | transcripts | position neuropeptide | length |
|---|---|---|---|---|
| Blattodea | + | 1 | C-terminal | 165-192 |
| Dermaptera | + | 1 | C-terminal | 211 |
| Embioptera | + | 1 | C-terminal | 178-180 |
| Grylloblattodea | + | 1 | C-terminal | n/a |
| Mantodea | + | 1 | C-terminal | 169-173 |
| Mantophasmatodea | + | 1 | C-terminal | 152 |
| Orthoptera: Caelifera | + | 1 | C-terminal | 146-171 |
| Orthoptera: Ensifera | + | 1 | C-terminal | 122-165 |
| Phasmatodea | + | 1 | C-terminal | 180-208 |
| Plecoptera | + | 1 | C-terminal | 172 |
| Zoraptera | + | 1 | C-terminal | n/a |



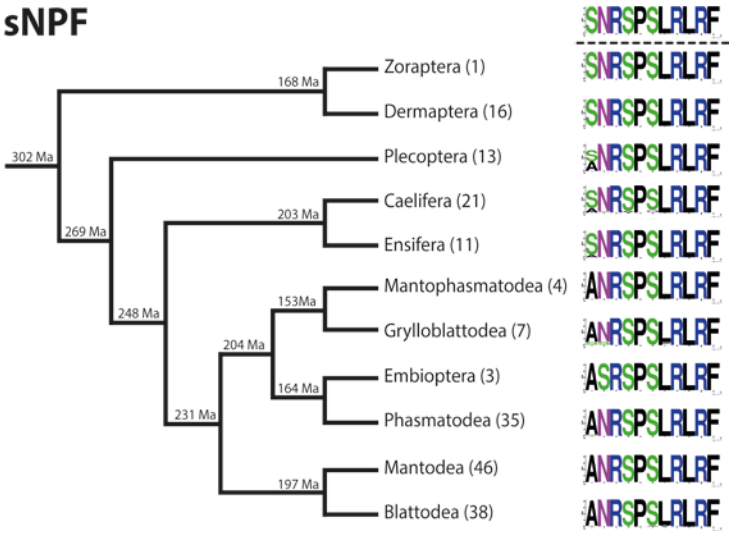
| SIFamide | presence | transcripts | position neuropeptide | length |
|---|---|---|---|---|
| Blattodea | + | 2 | in the middle | 72-74/73-74 |
| Dermaptera | + | 1 | in the middle | 74-76 |
| Embioptera | + | 1 | multiple copy | 100 |
| Grylloblattodea | + | 1 | in the middle | 72-73 |
| Mantodea | + | 1 | in the middle | 73 |
| Mantophasmatodea | + | 1 | in the middle | 73-75 |
| Orthoptera: Caelifera | + | 1 | in the middle | 73-75 |
| Orthoptera: Ensifera | + | 1 | in the middle | 73-103 |
| Phasmatodea | + | 1 | in the middle | 71-79 |
| Plecoptera | + | 1 | in the middle | 75-78 |
| Zoraptera | + | 1 | in the middle | 74 |

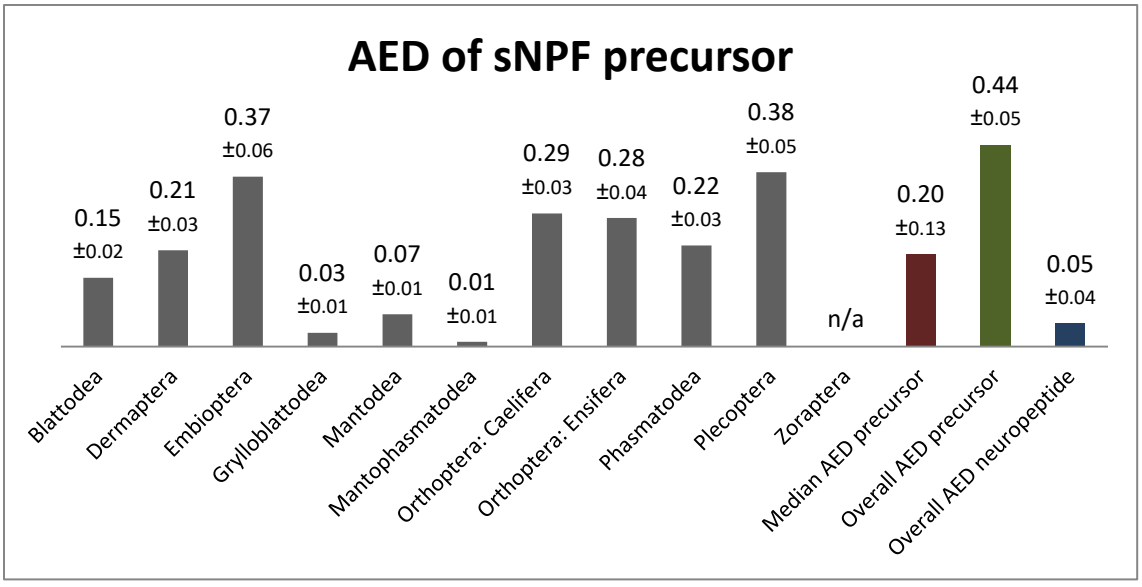


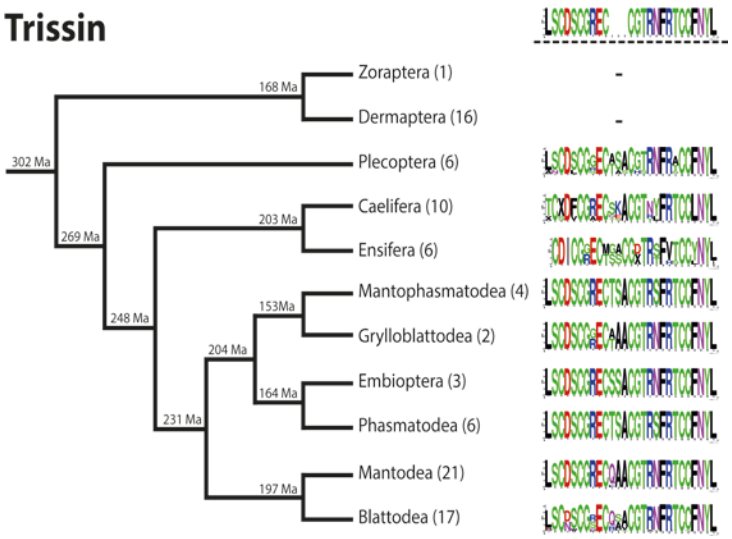
*T:*

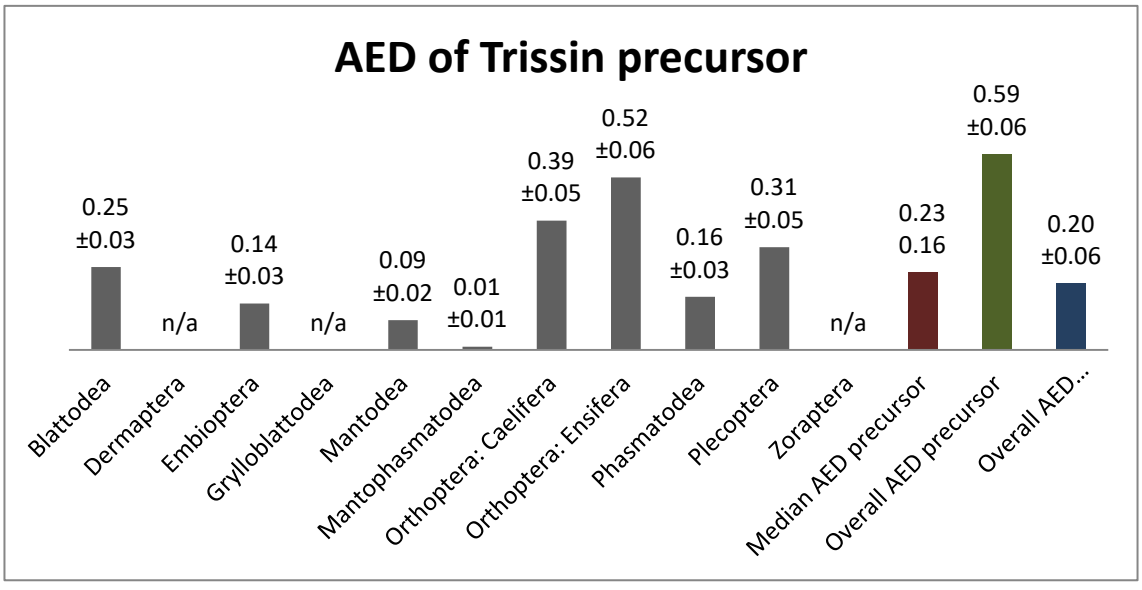
| sNPF | presence | transcripts | position neuropeptide | length |
|---|---|---|---|---|
| Blattodea | + | 1 | in the middle | 100-105 |
| Dermaptera | + | 1 | in the middle | 91-96 |
| Embioptera | + | 1 | in the middle | 97-98 |
| Grylloblattodea | + | 1 | in the middle | 96 |
| Mantodea | + | 1 | in the middle | 103-109 |
| Mantophasmatodea | + | 1 | in the middle | 97 |
| Orthoptera: Caelifera | + | 1 | in the middle | 93-134 |
| Orthoptera: Ensifera | + | 1 | in the middle | 97-100 |
| Phasmatodea | + | 1 | in the middle | 95-97 |
| Plecoptera | + | 1 | in the middle | 100-109 |
| Zoraptera | + | 1 | in the middle | 86 |



| Trissin | presence | transcripts | position neuropeptide | length |
|---|---|---|---|---|
| Blattodea | + | 1 | N-terminal | 112-117 |
| Dermaptera | - | n/a | n/a | n/a |
| Embioptera | + | 1 | N-terminal | 97-102 |
| Grylloblattodea | + | 1 | N-terminal | n/a |
| Mantodea | + | 1 | N-terminal | 110-111 |
| Mantophasmatodea | + | 1 | N-terminal | 90 |
| Orthoptera: Caelifera | + | 2 | N-terminal | 95-111/77 |
| Orthoptera: Ensifera | + | 1 | N-terminal | 88-92 |
| Phasmatodea | + | 1 | N-terminal | 95-96 |
| Plecoptera | + | 1 | N-terminal | 102-104 |
| Zoraptera | - | n/a | n/a | n/a |



## SM 3 The power of Neuropeptide Precursor Sequences to Reveal Phylogenetic Relationships in Insects: a Case Study on Blattodea.

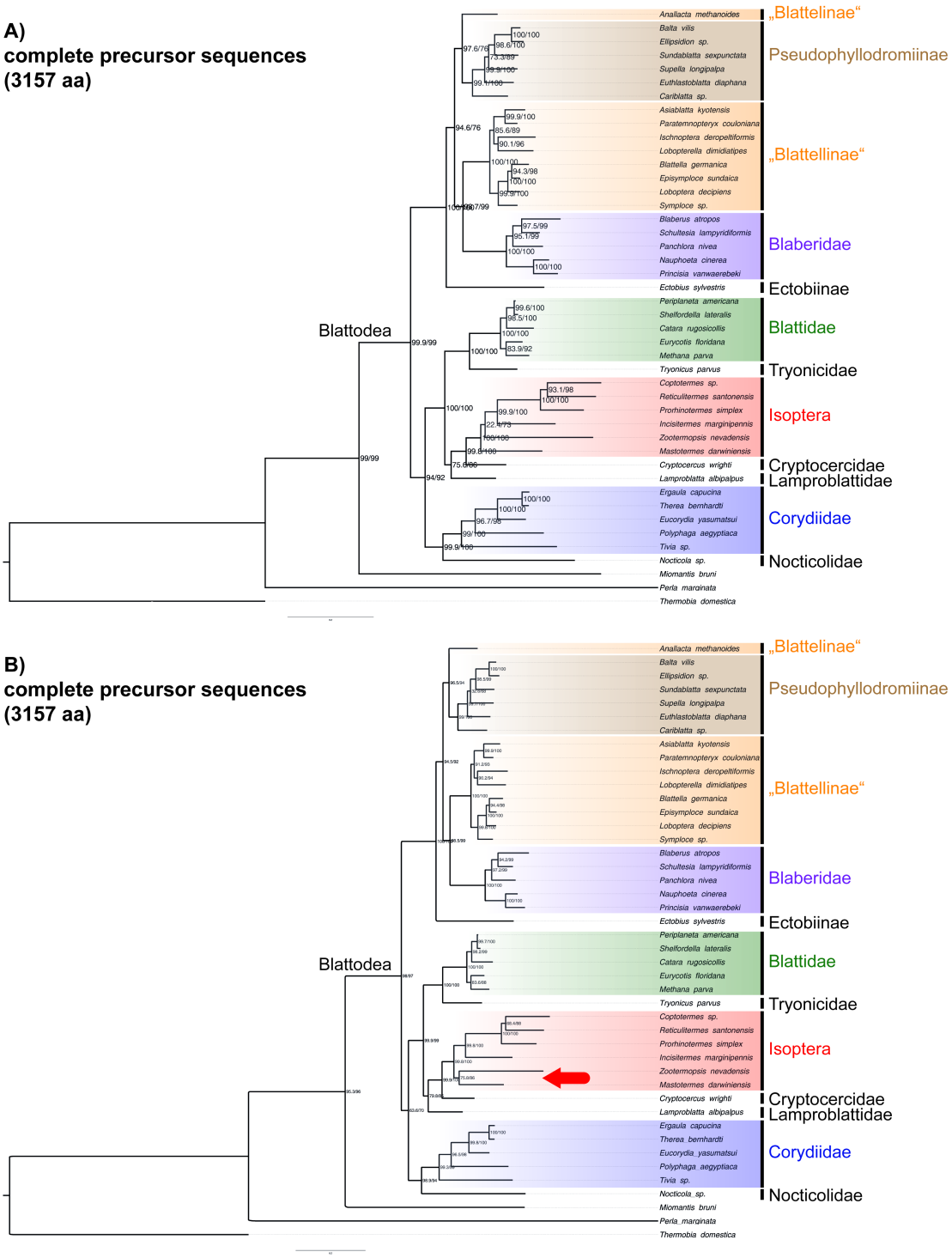



**Figure S1**. Molecular phylogenetic hypothesis using Maximum likelihood of a concatenated dataset of 17 neuropeptide precursors for 41 species of Blattodea and three hexapod outgroups (A) mtMet+F+I+G4 chosen according to BIC, B) site-heterogeneous mixture model CAT20). Numbers in parentheses represent the SH-aLRT support (%) / ultrafast bootstrap support (%).Traditionally accepted families are indicated by colors. Instead of the paraphyletic family Ectobiidae, subfamilies are given. The paraphyletic subfamily Blattelinae is in quotation marks. Note the position of Z. nevadensis as sister group to M. darwiniensis in B).

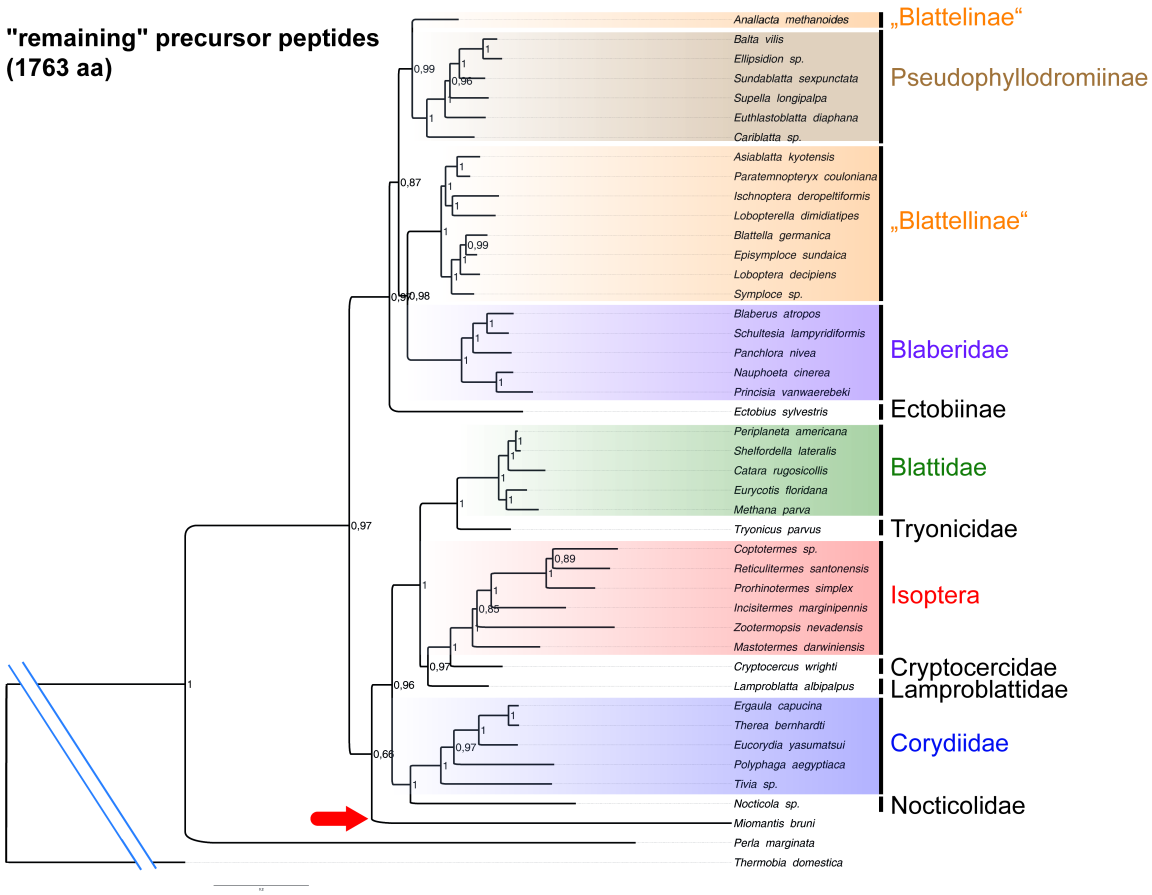



**Figure S2.** Bayesian Posterior Probability phylogenetic analysis of a concatenated dataset of 17 precursor peptide sequences for 41 species of Blattodea and three hexapod outgroups (model = GTR+G+F). This dataset does not contain the signal peptides and predicted neuropeptide sequences. The topology is very similar to that shown in Fig. 2, but with lower support of some nodes and the placement of Miomantis brunni inside of Blattodea. Traditionally accepted families are indicated by colors. Instead of the paraphyletic family Ectobiidae, subfamilies are given. The paraphyletic subfamily Blattelinae is in quotation marks.

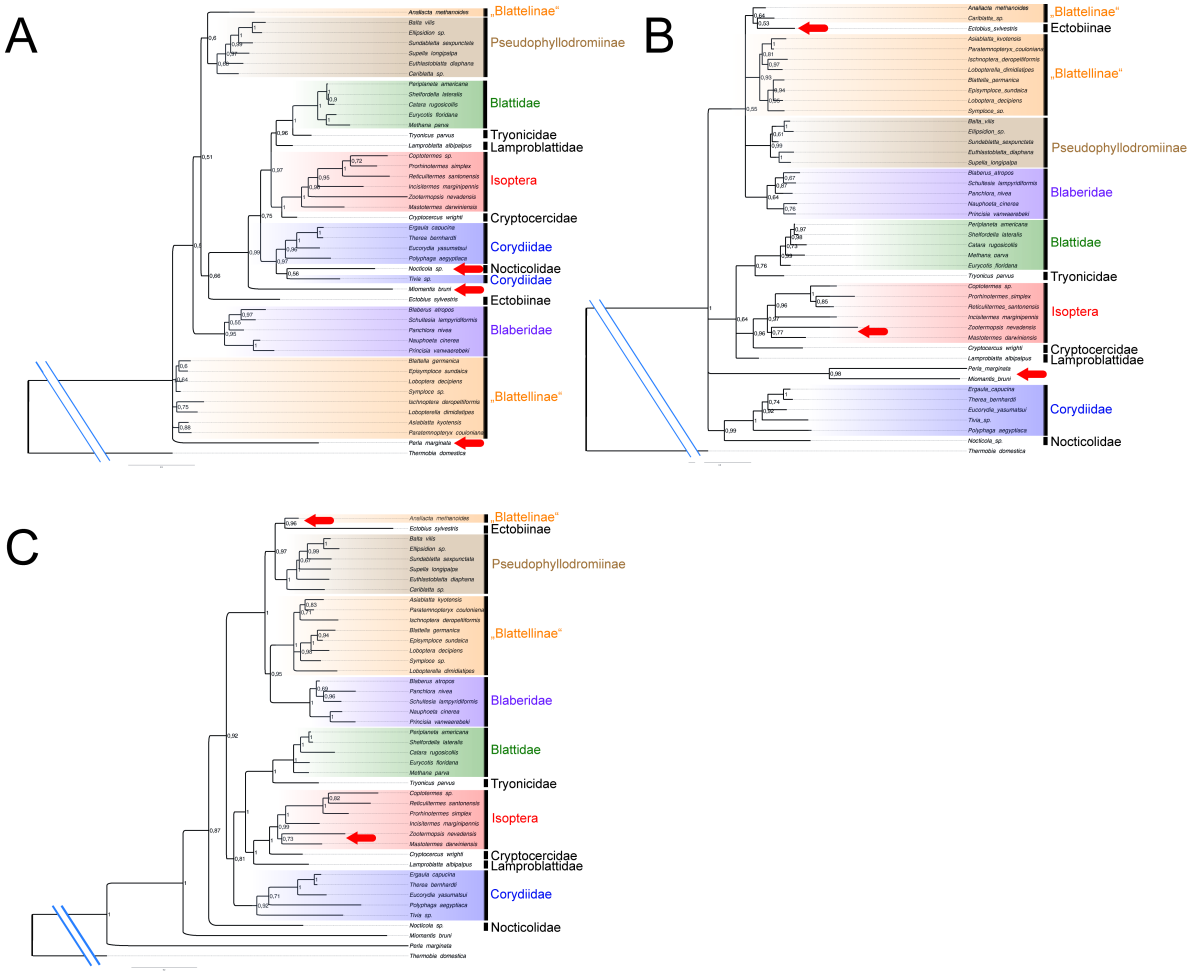



**Figure S3.** The Bayesian Posterior Probability phylogenetic analyses of the same dataset as shown in Figure S2, but divided in three sequence portions (A: 576 aa, B: 576 aa, C: 936 aa, model = GTR+G+F). Sequence length in A) and B) is identical to that for the neuropeptide sequences whose phylogenetic analysis is shown in Figure 3. However, the topologies in A) and B) show a low resolution. The topology in C) is based on more amino acids (936 aa), shows a better resolution and, with the exception of the placement of Ectobiinae and Z. nevadensis, provides the same lineages as the analysis of the complete precursor sequence (see Fig. 2). Traditionally accepted families are indicated by colors. Instead of the paraphyletic family Ectobiidae, subfamilies are given. The paraphyletic subfamily Blattelinae is in quotation marks.

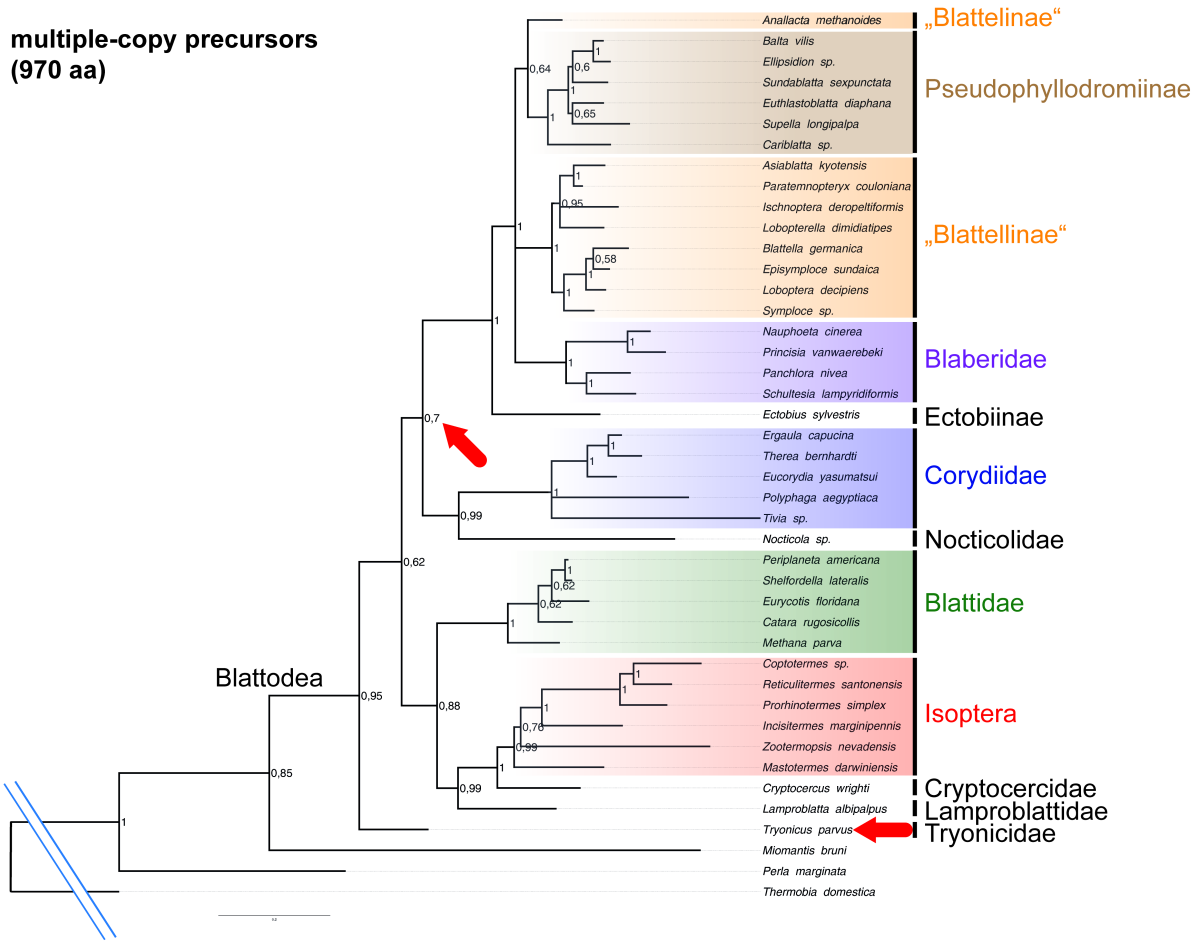



**Figure S4**. Bayesian Posterior Probability (BPP) phylogenetic analysis of the concatenated dataset of three multiple-copy neuropeptide precursors for 40 species of Blattodea (Blaberus atropos excluded due to missing precursor sequences) and three hexapod outgroups (model = GTR+G+F) with BPP values. Note the placement of Tryonicidae as sister to all remaining Blattodea. The position of Corydiidae and Nocticolidae is not solved. Traditionally accepted families are indicated by colors. Instead of the paraphyletic family Ectobiidae, subfamilies are given. The paraphyletic subfamily Blattelinae is in quotation marks.

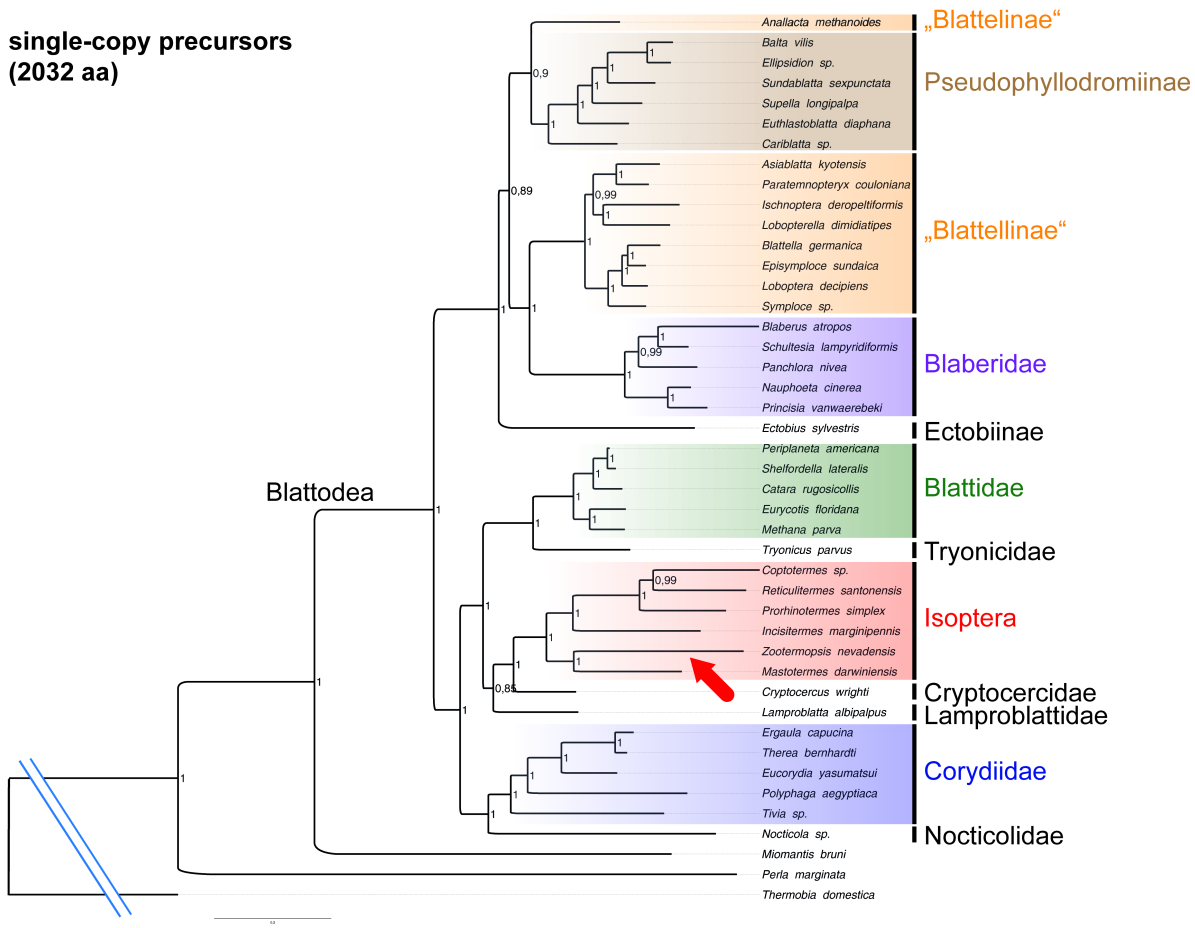



**Figure S5.** Bayesian Posterior Probabilities phylogenetic analysis of a concatenated dataset of 14 single-copy neuropeptide precursors for 41 species of Blattodea and three hexapod outgroups (model = GTR+I+G+F) BPP values. The only significant difference to the analysis of the complete precursor sequence (see Fig. 2) is the placement of Z. nevadensis as sister to Mastotermes darwiniensis. Traditionally accepted families are indicated by colors. Instead of the paraphyletic family Ectobiidae, subfamilies are given. The paraphyletic subfamily Blattelinae is in quotation marks.

**Table S1.** List of species used in this study, including the corresponding authors and years of description, and the source of the transcriptome. In cases of unidentified species, the first description of the genus is given. Species marked with an asterisk contain partial sequences of multiple-copy precursors (CAPA, FXPRL or TKRP) extracted from raw data.

| species name | authority | origin |
|---|---|---|
| *Anallacta methanoides* | Shelford, 1908 | Genbank Umbrella Bioproject ID PRJNA183205 |
| *Asiablatta kyotensis* | (Asahina, 1976) | Genbank Umbrella Bioproject ID PRJNA183205 |
| *Balta vilis* | (Brunner von Wattenwyl, 1865) | Genbank Umbrella Bioproject ID PRJNA183205 |
| *Blaberus atropos* | Lopes & Oliveira, 2013 | Genbank Umbrella Bioproject ID PRJNA183205 |
| *Blattella germanica* | (Linnaeus, 1767) | Genbank Umbrella Bioproject ID PRJNA183205 |
| *Cariblatta* sp.* | Hebard, 1916 | Genbank Umbrella Bioproject ID PRJNA183205 |
| *Catara rugosicollis* | (Brunner von Wattenwyl, 1865) | Genbank Umbrella Bioproject ID PRJNA183205 |
| *Cryptocercus wrighti* | Burnside, Smith & Kambhampati, 1999 | Genbank Umbrella Bioproject ID PRJNA183205 |
| *Ectobius sylvestris** | (Poda, 1761) | Genbank Umbrella Bioproject ID PRJNA183205 |
| *Ellipsidion* sp. | Saussure, 1863 | Genbank Umbrella Bioproject ID PRJNA183205 |
| *Episymploce sundaica** | (Hebard, 1929) | Genbank Umbrella Bioproject ID PRJNA183205 |
| *Ergaula capucina* | (Brunner von Wattenwyl, 1893) | Genbank Umbrella Bioproject ID PRJNA183205 |
| *Eucorydia yasumatsui** | Asahina, 1971 | Genbank Umbrella Bioproject ID PRJNA183205 |
| *Eurycotis floridana** | (F.Walker, 1868) | Genbank Umbrella Bioproject ID PRJNA183205 |
| *Euthlastoblatta diaphana* | (Fabricius, 1793) | Genbank Umbrella Bioproject ID PRJNA183205 |
| *Ischnoptera deropeltiformis** | (Brunner von Wattenwyl, 1865) | Genbank Umbrella Bioproject ID PRJNA183205 |
| *Lamproblatta albipalpus* | Hebard, 1919 | Genbank Umbrella Bioproject ID PRJNA183205 |
| *Loboptera decipiens** | (Germar, 1817) | Genbank Umbrella Bioproject ID PRJNA183205 |
| *Lobopterella dimidiatipes** | (Bolívar, 1890) | Genbank Umbrella Bioproject ID PRJNA183205 |
| *Methana parva** | Shaw, 1925 | Genbank Umbrella Bioproject ID PRJNA183205 |
| *Nauphoeta cinerea* | (Olivier, 1789) | Genbank Umbrella Bioproject ID PRJNA183205 |
| *Nocticola* sp. | Bolívar, 1892 | Genbank Umbrella Bioproject ID PRJNA183205 |
| *Panchlora nivea** | (Linnaeus, 1758) | Genbank Umbrella Bioproject ID PRJNA183205 |
| *Paratemnopteryx couloniana* | (Saussure, 1863) | Genbank Umbrella Bioproject ID PRJNA183205 |
| *Periplaneta americana* | (Linnaeus, 1758) | Genbank Umbrella Bioproject ID PRJNA183205 |
| *Polyphaga aegyptiaca* | (Linnaeus, 1758) | Genbank Umbrella Bioproject ID PRJNA183205 |
| *Princisia vanwaerebeki* | van Herrewege, 1973 | Genbank Umbrella Bioproject ID PRJNA183205 |
| *Schultesia lampyridiformis** | Roth & L.M., 1973 | Genbank Umbrella Bioproject ID PRJNA183205 |
| *Shelfordella lateralis* | (F.Walker, 1868) | Genbank Umbrella Bioproject ID PRJNA183205 |
| *Sundablatta sexpunctata* | (Hanitsch, 1923) | Genbank Umbrella Bioproject ID PRJNA183205 |
| *Supella longipalpa* | (Fabricius, 1798) | Genbank Umbrella Bioproject ID PRJNA183205 |
| *Symploce* sp. | Hebard, 1916 | Genbank Umbrella Bioproject ID PRJNA183205 |
| *Therea bernhardti* | Fritzsche, 2009 | Genbank Umbrella Bioproject ID PRJNA183205 |
| *Tivia* sp. | Walker, 1869 | Genbank Umbrella Bioproject ID PRJNA183205 |
| *Tryonicus parvus* | (Tepper, 1895) | Genbank Umbrella Bioproject ID PRJNA183205 |
| *Coptotermes* sp. | Adams, 1990 | Genbank Umbrella Bioproject ID PRJNA183205 |
| *Incisitermes marginipennis** | (Latreille, 1817) | Genbank Umbrella Bioproject ID PRJNA183205 |
| *Mastotermes darwiniensis* | Froggatt, 1897 | Genbank Umbrella Bioproject ID PRJNA183205 |
| *Zootermopsis nevadensis* | (Hagen, 1874) | GenBank nucleotide core database ID AUST00000000 |
| *Prorhinotermes simplex* | (Hagen, 1858) | Genbank Umbrella Bioproject ID PRJNA183205 |
| *Reticulitermes santonensis* | Feytaud, 1924 | Genbank Umbrella Bioproject ID PRJNA183205 |
| *Miomantis brunni* | Giglio-Tos, 1911 | Genbank Umbrella Bioproject ID PRJNA183205 |
| *Perla marginata* | (Panzer, 1799) | This study |
| *Thermobia domestica* | (Packard, 1873) | This study |

# Marcel Bläser

Methweg 4 • 50823 Cologne, Germany

Mblaese1@uni-koeln.de • Tel.: +49 163- 16 07 93 8

## Personal Information

| | |
|---|---|
| Date of Birth: | 23.08.1989 |
| Place of Birth: | Bonn-Duisdorf, Germany |
| Nationality: | German |

## University Educational History

| | |
|---|---|
| 2016 – 2020 | PhD thesis: “Evolution of Neuropeptide Precursors in Polyneoptera (Insecta)”<br>Lab of R. Predel, University of Cologne |
| 2013 – 2016 | Rheinische Friedrich Wilhelm University Bonn, Germany<br>Degree: Master of Science: Organismic, Evolutionary and Palaeobiology<br>Master thesis: “Dating the Tree of Chalidoidea (Hymenoptera)”<br>Lab of R. S. Peters, ZFMK Bonn |
| 2009 – 2013 | University of Cologne, Germany<br>Degree: Bachelor of Science: Biology<br>Bachelor thesis: “Identification of the Neuropeptides of Zygentoma (Hexapoda)”<br>Lab of R. Predel, University of Cologne |

## Work and Research History

| | |
|---|---|
| 07.2014 – 03.2015 | Scientific assistant, Lab of R. Predel, University of Cologne<br>Area: Functional Peptidomics |
| 09.2014 – 12.2014 | Scientific assistant in the „CO.BRA - Computational Bioacoustics Research Unit” project of the Universidad Federal de Mato Grosso, Brazil<br>Area: Wildlife ecology/ Camera Trapping |
| 04.2013 – 09.2013 | Student assistant, Lab of R. Predel, University of Cologne<br>Area: Functional Peptidomics |
| 04.2011 – 12.2012 | Student assistant, Lab of M. Bucher, University of Cologne<br>Area: Plant Physiology |
| 05.2012 – 07.2012 | Internship, Otter-research-center Hankensbüttel, Germany<br>Area: Wildlife Ecology |

## Peer Reviewed Publication

**Bläser, M.**, Krogmann, L., Peters, R. S. (2015). Two new fossil genera and species of Cerocephalinae (Hymenoptera, Chalcidoidea, Pteromalidae), including the first record from the Eocene. *ZooKeys*, (545), 89.

Redeker, J., **Bläser, M.**, Neupert, S., & Predel, R. (2017). Identification and distribution of products from novel tryptopyrokinin genes in the locust, *Locusta migratoria*. *BBRC*, 486(1), 70-75.

Peters, R.S., Niehuis, O., Gunkel, S., **Bläser, M.**, et al. (2018).Transcriptome sequence-based phylogeny of chalcidoid wasps (Hymenoptera: Chalcidoidea) reveals a history of rapid radiations, convergence, and evolutionary success. *Mol. Phylogenet. Evol.,* 120 (2018), pp. 286-296

**Bläser, M.**, Misof, B., Predel, R. (2020). The power of neuropeptide precursor sequences to reveal phylogenetic relationships in insects: a case study on Blattodea*. Mol. Phylogenet. Evol*. 106686

**Bläser, M.**, Predel, R. (2020). Evolution of neuropeptide precursors in Polyneoptera (Insecta). *Front. Endocrinol*. 11, 197

## Scientific Meetings and Conferences

| | |
|---|---|
| 09.2019 | 112th Annual Meeting of the German Zoological Society (DZG)<br>Poster Presentation |
| 09.2018 | 111th Annual Meeting of the German Zoological Society (DZG)<br>Oral Presentation |
| 05.2018 | Arthropod Neuroscience Network Spring Meeting 2018<br>Poster Presentation |
| 09.2017 | 8th international Dresden Meeting on Insect Phylogeny<br>Oral Presentation |
| 09.2016 | 109th Annual Meeting of the German Zoological Society (DZG)<br>Poster Presentation |
| 05.2016 | 14th Rauischholzhausen Seminar for Insect Neurology<br>Oral Presentation |
| 05.2013 | 11th Rauischholzhausen Seminar for Insect Neurology<br>Oral Presentation |

## Teaching History

| | |
|---|---|
| 04.2016 – ongoing | Bachelor/ Master thesis supervision, Lab of R. Predel, University of Cologne. Area: Functional Peptidomics/ Transcriptom Analysis |

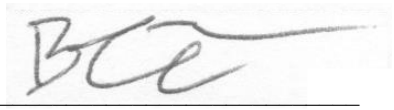
| | |
|---|---|
| 10.2015 – 02.2016 | Scientific assistant: Student course supervision, University of Cologne<br>Area: Zoology |
| 06.2013 – 08.2013 | Student assistant: Student course supervision, University of Cologne<br>Area: Ecology |
| WS 10/11 & WS 11/12 | Tutor in the Freshman-Program of the student association Biology & Biochemistry, University of Cologne |

## Work-related Responsibilities

| | |
|---|---|
| 03.2015 – 03.2015 | Organization of the 16th Annual Meeting of the Society for Biological Systematics (GfBS) |
| 04.2013 – 09.2013 | Acclaimed member of the 58th Student parliament of the University of Cologne |
| 10.2010 – 09.2013 | Acclaimed member of the student association Biology & Biochemistry, University of Cologne. Responsibilities included e.g.:<br>Member of Bachelor examination board (WS 10/11 - SS 2013)<br>Head of student association (WS 11/12 – WS12/13) |

## Courses and Career Development Workshops

| | |
|---|---|
| Project Management | - Accelerate your career! Entrepreneurship, strategy and leadership for MINT students, University of Cologne<br>- PhD and Project Management, GSfBS, University of Cologne<br>- Conflict Management, GSfBS, University of Cologne<br>- Skills for Project Leaders/Managers, GSfBS, University of Cologne |
| Language courses | - English Coaching - Conversational English for Doctoral Candidates, GSfBS, University of Cologne<br>- How to write clearly in science and research, University of Cologne<br>- Präsentieren 2.0 in der Wissenschaft – Alternativen zu Powerpoint & Co, University of Cologne |
| Summer School | - ERASMUS Education Program: Origin, Evolution and Future of the Biosphere (2 weeks) |
| Scientific Practice | - International Tropical Ecology Course: Edge effect and biodiversity of Northern Pantanal forests, University Bonn, Germany & Universidad Federal de Mato Grosso, Brazil (2 weeks)<br>- Seminar in Good Scientific Practice – Ethics in Science and Good Scientific Practice, University of Cologne |

Köln, 20.04.2020 

Marcel Bläser

# 10. Erklärung

*Hiermit versichere ich an Eides statt, dass ich die vorliegende Dissertation selbstständig und ohne die Benutzung anderer als der angegebenen Hilfsmittel und Literatur angefertigt habe. Alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten und nicht veröffentlichten Werken dem Wortlaut oder dem Sinn nach entnommen wurden, sind als solche kenntlich gemacht. Ich versichere an Eides statt, dass diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; dass sie - abgesehen von unten angegebenen Teilpublikationen und eingebundenen Artikeln und Manuskripten - noch nicht veröffentlicht worden ist sowie, dass ich eine Veröffentlichung der Dissertation vor Abschluss der Promotion nicht ohne Genehmigung des Promotionsausschusses vornehmen werde. Die Bestimmungen dieser Ordnung sind mir bekannt. Darüber hinaus erkläre ich hiermit, dass ich die Ordnung zur Sicherung guter wissenschaftlicher Praxis und zum Umgang mit wissenschaftlichem Fehlverhalten der Universität zu Köln gelesen und sie bei der Durchführung der Dissertation zugrundeliegenden Arbeiten und der schriftlich verfassten Dissertation beachtet habe und verpflichte mich hiermit, die dort genannten Vorgaben bei allen wissenschaftlichen Tätigkeiten zu beachten und umzusetzen. Ich versichere, dass die eingereichte elektronische Fassung der eingereichten Druckfassung vollständig entspricht.*

Köln, den 13.07.2020

Marcel Bläser