

Multiple transcriptome mining coupled with tissue specific molecular cloning and mass spectrometry provide insights into agatoxin-like peptide conservation in decapod crustaceans



Andrew E. Christie^a, Cindy D. Rivera^b, Catherine M. Call^b, Patsy S. Dickinson^c, Elizabeth A. Stemmler^b, J. Joe Hull^{d,*}

^a Békésy Laboratory of Neurobiology, Pacific Biosciences Research Center, School of Ocean and Earth Science and Technology, University of Hawaii at Manoa, 1993 East-West Road, Honolulu, HI 96822, USA

^b Department of Chemistry, Bowdoin College, 6600 College Station, Brunswick, ME 04011, USA

^c Department of Biology, Bowdoin College, 6500 College Station, Brunswick, ME 04011, USA

^d Pest Management and Biocontrol Research Unit, US Arid Land Agricultural Research Center, USDA Agricultural Research Services, Maricopa, AZ 85138, USA

ARTICLE INFO

Keywords:

In silico transcriptome mining
RT-PCR
Mass spectrometry
Nervous system
Testis
Hepatopancreas

ABSTRACT

Over the past decade, *in silico* genome and transcriptome mining has led to the identification of many new crustacean peptide families, including the agatoxin-like peptides (ALPs), a group named for their structural similarity to agatoxin, a spider venom component. Here, analysis of publicly accessible transcriptomes was used to expand our understanding of crustacean ALPs. Specifically, transcriptome mining was used to investigate the phylogenetic/structural conservation, tissue localization, and putative functions of ALPs in decapod species. Transcripts encoding putative ALP precursors were identified from one or more members of the Penaeoidea (penaeid shrimp), Sergestoidea (sergestid shrimps), Caridea (caridean shrimp), Astacidea (clawed lobsters and freshwater crayfish), Achelata (spiny/slipper lobsters), and Brachyura (true crabs), suggesting a broad, and perhaps ubiquitous, conservation of ALPs in decapods. Comparison of the predicted mature structures of decapod ALPs revealed high levels of amino acid conservation, including eight identically conserved cysteine residues that presumably allow for the formation of four identically positioned disulfide bridges. All decapod ALPs are predicted to have amidated carboxyl-terminals. Two isoforms of ALP appear to be present in most decapod species, one 44 amino acids long and the other 42 amino acids in length, both likely generated by alternative splicing of a single gene. In carideans, a gene or terminal exon duplication appears to have occurred, with alternative splicing producing four ALPs, two 44 and two 42 amino acid isoforms. The identification of ALP precursor-encoding transcripts in nervous system-specific transcriptomes (e.g., *Homarus americanus* brain, eye-stalk ganglia, and cardiac ganglion assemblies, finding confirmed using RT-PCR) suggests that members of this peptide family may serve as locally-released and/or hormonally-delivered neuromodulators in decapods. Their detection in testis- and hepatopancreas-specific transcriptomes suggests that members of the ALP family may also play roles in male reproduction and innate immunity/detoxification.

1. Introduction

Over the past decade, considerable effort has been put into the development of molecular resources (genomes and deep transcriptomes) for arthropod species, including many crustaceans (e.g., Armstrong et al., 2019; Christie et al., 2017, 2018a, 2018b; Havird and Santos, 2016; He et al., 2012; Huerlimann et al., 2018; Li et al., 2015; Lv et al., 2017; Manfrin et al., 2015; Northcutt et al., 2016; Oliphant

et al., 2018; Rahi et al., 2019; Santos et al., 2018; Souza et al., 2018; Sun et al., 2014; Tom et al., 2014; Ventura et al., 2014; Verbruggen et al., 2015; Wang et al., 2019; Xu et al., 2015). Although these datasets were initially developed to serve a variety of functions, they have proven to be powerful resources for a wide array of gene and, by proxy, protein discoveries. Crustacean genomes and transcriptomes have been extensively exploited to facilitate the identification of the molecular components (genes and proteins) of peptidergic signaling systems,

* Corresponding author at: Pest Management and Biocontrol Research Unit, US Arid Land Agricultural Research Center, USDA Agricultural Research Services, 21881 North Cardon Lane, Maricopa, AZ 85138, USA.

E-mail address: joe.hull@ars.usda.gov (J.J. Hull).

<https://doi.org/10.1016/j.ygcen.2020.113609>

Received 24 July 2020; Received in revised form 25 August 2020; Accepted 29 August 2020

Available online 09 September 2020

0016-6480/ Published by Elsevier Inc.

Table 1
Agatoxin-like peptide (ALP) precursor transcript/protein discovery in decapod crustaceans.

Species	Transcript		Deduced protein			
Taxonomic group ^a	Species	Accession No.	Tissue source	Name	Length	Type
Penaeoidea	<i>Litopenaeus vannamei</i>	GFRP01012298	Mixed	Litva-prepro-ALP-v1	112	F
		GGQV01059635	Eyestalk	Litva-prepro-ALP-v2	108	F
		GGK001019727	Mixed	Litva-prepro-ALP-v3	99	F
	<i>Penaeus monodon</i>	GFRP01012297	Mixed	Litva-prepro-ALP-v3	99	F
		GGLH01104228	Mixed	Penmo-prepro-ALP-v1	108	F
		GGLH01049277	Mixed	Penmo-prepro-ALP-v2	99	F
	<i>Metapenaeus bennettiae</i>	GHDJ01038348	Hepatopancreas	Metbe-prepro-ALP-v1	112	F
		GHDJ01038347	Hepatopancreas	Metbe-prepro-ALP-v2	99	F
	<i>Fenneropenaeus penicillatus</i>	GFRT01015650	Mixed	Fenpe-prepro-ALP	46	C
	<i>Acetes chinensis</i>	GGVZ01058550	Mixed	Acech-prepro-ALP-v1	106	F
Sergestoidea	<i>Macrobrachium australiense</i>	GGVZ01005325	Mixed	Acech-prepro-ALP-v1	106	F
		GGVZ01058552	Mixed	Acech-prepro-ALP-v2	104	F
		GGVZ01005326	Mixed	Acech-prepro-ALP-v2	104	F
		GGVZ01058551	Mixed	Acech-prepro-ALP-v3	95	F
		GGVZ01005324	Mixed	Acech-prepro-ALP-v3	95	F
		GHD01074459	Mixed	Macau-prepro-ALP-I-v1	117	F
	<i>Macrobrachium koombooloomba</i>	GHD01074458	Mixed	Macau-prepro-ALP-I-v2	111	F
		GHD01074460	Mixed	Macau-prepro-ALP-I-v3	106	F
		GHD01074462	Mixed	Macau-prepro-ALP-I-v4	100	F
		GHD01074464	Mixed	Macau-prepro-ALP-II-v1	117	F
		GHD01074461	Mixed	Macau-prepro-ALP-II-v2	111	F
		GHD01074463	Mixed	Macau-prepro-ALP-II-v3	100	F
Caridea	<i>Macrobrachium novaehollandiae</i>	GHDU01066920	Mixed	Macko-prepro-ALP-I-v1	111	F
		GHDU01066922	Mixed	Macko-prepro-ALP-I-v2	100	F
		GHDU01066919	Mixed	Macko-prepro-ALP-II-v1	111	F
		GHDU01066918	Mixed	Macko-prepro-ALP-II-v2	100	F
		GHDW01093390	Mixed	Macno-prepro-ALP-I	111	F
		GHDW01093388	Mixed	Macno-prepro-ALP-II-v1	117	F
	<i>Macrobrachium tolmerum</i>	GHDW01093389	Mixed	Macno-prepro-ALP-II-v2	111	F
		GHDQ01089244	Mixed	Macto-prepro-ALP-I-v1	111	F
		GHDQ01089242	Mixed	Macto-prepro-ALP-I-v2	100	F
		GHDQ01089241	Mixed	Macto-prepro-ALP-II	111	F
Astacidea	<i>Metabetaeus lohena</i>	GHDQ01089243	Mixed	Macto-prepro-ALP-II	111	F
		GHAP01062116	Mixed	Metlo-prepro-ALP-I-v1	111	F
		GHAP01062115	Mixed	Metlo-prepro-ALP-I-v2	100	F
		GHAP01062117	Mixed	Metlo-prepro-ALP-II-v1	111	F
		GHAP01062114	Mixed	Metlo-prepro-ALP-II-v2	100	F
		GHBK01056446	Mixed	Halru-prepro-ALP-I-v1	117	F
	<i>Halocaridina rubra</i>	GHBK01056445	Mixed	Halru-prepro-ALP-I-v2	106	F
		GHBK01108229	Mixed	Halru-prepro-ALP-II	53	C
	<i>Halocaridinides trigonophthalma</i>	GHB01088024	Mixed	Haltr-prepro-ALP-I	108	C
	<i>Neocaridina denticulata</i>	GGXN01030525	Mixed	Neode-prepro-ALP-I	102	F
Astacidea	<i>Antecaridina lauensis</i>	GHB01037781	Mixed	Antla-prepro-ALP-I	103	F
		GHB01037782	Mixed	Antla-prepro-ALP-I	103	F
	<i>Homarus americanus</i>	GFDA01040082	Eyestalk ganglia	Homam-prepro-ALP-v1	113	F
		GGPK01031489	Cardiac ganglion	Homam-prepro-ALP-v1	113	F
		GFUC01090608	Brain	Homam-prepro-ALP-v1	113	F
		GGPK01031491	Cardiac ganglion	Homam-prepro-ALP-v2	109	F
		GFUC01090606	Brain	Homam-prepro-ALP-v2	109	F
		GFDA01040083	Eyestalk ganglia	Homam-prepro-ALP-v3	100	F
Astacidea	<i>Cherax quadricarinatus</i>	GGPK01031490	Cardiac ganglion	Homam-prepro-ALP-v3	100	F
		GEBG01002733	Nervous system	Homam-prepro-ALP-v3	100	F
		HACK01031533	Mixed	Chequ-prepro-ALP-v1	113	F
		HACK01031534	Mixed	Chequ-prepro-ALP-v2	109	F
		HACB02001107	Mixed	Chequ-prepro-ALP-v3	100	F
		HACK01031532	Mixed	Chequ-prepro-ALP-v3	100	F
	<i>Pontastacus leptodactylus</i>	GBEI01092752	Mixed	Ponle-prepro-ALP	113	F
	<i>Procambarus clarkii</i>	GARH01001590	Eyestalk	Procl-prepro-ALP	129	F
	<i>Jasus edwardsii</i>	GGHM01043346	Mixed	Jased-prepro-ALP-v1	113	F
		GGHM01043347	Mixed	Jased-prepro-ALP-v2	100	F
Brachyura	<i>Carcinus maenas</i>	GFYV01055236	Mixed	Carma-prepro-ALP-v1	113	F
		GFYV01063051	Mixed	Carma-prepro-ALP-v1	61	C
		GFYV01063050	Mixed	Carma-prepro-ALP-v1	61	C
		GFYV01063049	Mixed	Carma-prepro-ALP-v1	61	C
		GFXF01149802	Mixed	Carma-prepro-ALP-v2	109	F
		GBXE01084435	Mixed	Carma-prepro-ALP-v2	109	F
	<i>Eriocheir sinensis</i>	GFYV01055237	Mixed	Carma-prepro-ALP-v3	100	F
		GBUF01003221	Mixed	Erisi-prepro-ALP-v1	113	F
		GFB01086394	Mixed	Erisi-prepro-ALP-v1	113	F
		GFB01086391	Mixed	Erisi-prepro-ALP-v1	113	F
Brachyura	<i>Eriocheir sinensis</i>	GBZW01010384	Eyestalk	Erisi-prepro-ALP-v2	109	F
		JR775346	Testis	Erisi-prepro-ALP-v2	52	C

(continued on next page)

Table 1 (continued)

Species		Transcript		Deduced protein		
Taxonomic group ^a	Species	Accession No.	Tissue source	Name	Length	Type
	<i>Portunus trituberculatus</i>	HAAX01020856	Mixed	Erisi-prepro-ALP-v3	100	F
		GGQO01017109	Mixed	Erisi-prepro-ALP-v3	100	F
		HAAX01020856	Mixed	Erisi-prepro-ALP-v3	100	F
		GBUF01003220	Mixed	Erisi-prepro-ALP-v3	100	F
		GFBL01086390	Mixed	Erisi-prepro-ALP-v3	100	F
		GFBL01086392	Mixed	Erisi-prepro-ALP-v3	100	F
		GFFJ01041078	Eyestalk	Portr-prepro-ALP	109	F

^a Taxonomic groups: Penaeoidea, a superfamily within the suborder Dendrobranchiata; Sergestoidea, a superfamily within the suborder Dendrobranchiata; Caridea, an infraorder within the suborder Pleocyemata; Astacidea, an infraorder within the suborder Pleocyemata; Achelata, an infraorder within the suborder Pleocyemata; Brachyura, an infraorder within the suborder Pleocyemata. Peptide type abbreviations: F, full-length protein; C, carboxyl-terminal partial protein.

including peptide precursors (e.g., Bao et al., 2015; Christie and Hull, 2019; Christie et al., 2015; Nguyen et al., 2016; Oliphant et al., 2018; Veenstra, 2015, 2016), peptide processing enzymes (Christie et al., 2018c), and peptide receptors (e.g., Bao et al., 2018; Buckley et al., 2016; Christie and Hull, 2019; Christie et al., 2015; Dickinson et al., 2019; Oliphant et al., 2018; Tran et al., 2019; Veenstra, 2016).

In silico mining of publicly accessible crustacean genomes and transcriptomes for peptide precursors has led to the identification of new isoforms for known crustacean peptide families as well as peptide groups either previously unknown or poorly represented in this taxon. Among the latter is the agatoxin-like peptide (ALP) family, which was first identified in the honey bee (*Apis mellifera*) corpora cardiaca (Sturm et al., 2016). Peptides in this family share structural similarities with a class of peptide toxins (agatoxins) initially isolated from the American funnel-web spider *Agelenopsis aperta* (Skinner et al., 1989; Bindokas and Adams, 1989; Adams et al., 1990). ALPs have since been identified in a number of arthropod genomic/transcriptomic datasets (Sturm et al., 2016; Veenstra, 2016). The ALP prepropeptide typically consists of a signal peptide, two to three precursor peptides, and an ~40 amino acid ALP region that contains a C-terminal amidation signal and eight highly conserved Cys residues that comprise two Cys motifs commonly found in spider toxins. The N-terminal principal structural motif is characterized by a Cys spacing pattern of CX₆CX₆CC (X = any amino acid) with the dual Cys residues (CC) four amino acid residues upstream of the C-terminal second extra structural motif, which follows a CXCX₆CXC pattern. Although alternatively spliced variants of the ALP prepropeptide have been identified (Sturm et al., 2016; Veenstra, 2016; von Reumont et al., 2014), the splice variants typically affect the precursor peptides rather than the ALP region itself, suggesting evolutionary pressure to conserve functionality. While the biological role of ALPs has yet to be described, the structurally related spider agatoxins have been reported to affect neurotransmitter release via modification of various receptor-activated and/or voltage-gated channels at insect neuromuscular junctions (Adams, 2004). These structural similarities and ALP expression in diverse venom glands (Bouazid et al., 2014; Liu et al., 2015; Torres et al., 2014; von Reumont et al., 2014) could indicate a similar toxin-like function; however, ALPs have been identified in transcriptomic and genomic datasets from a number of non-venomous arthropods (Bao et al., 2020; Christie, 2020; Liessem et al., 2018; Oliphant et al., 2020; Sturm et al., 2016; Veenstra, 2016). Furthermore, direct detection of ALPs in the neuroendocrine systems (corpora cardiaca, brain, and stomatogastric nervous system) of the honey bee (*A. mellifera*), the American cockroach (*Periplaneta americana*), and the firebrat (*Thermobia domestica*) suggest non-toxic functional roles (Sturm et al., 2016).

Although ALPs have been reported in members of the Crustacea (Sturm et al., 2016; Veenstra, 2016), numerous questions remain concerning these peptides in crustaceans. For example, to what extent are the ALPs phylogenetically and structurally conserved in the taxon? Similarly, what tissues produce ALPs in crustaceans, and what physiological roles do they play in

members of this arthropod subphylum? Here, publicly accessible transcriptomes were used to address these questions in members of the Decapoda, a crustacean order that is of high economic importance due to its role in commercial fisheries and aquaculture, and whose members have been widely used as models for investigating peptidergic control of physiology and behavior.

2. Materials and methods

2.1. Animals

Lobsters, *H. americanus*, were purchased from local (Brunswick, Maine, USA) seafood distributors. All animals used were adults of ~500 g and included both males and females. Lobsters were housed in recirculating natural seawater aquaria at 10–12 °C and were fed chopped squid weekly.

2.2. *In silico* transcriptome mining and peptide structural prediction

2.2.1. *In silico* transcriptome mining

Searches of publicly accessible decapod transcriptomic datasets for putative ALP precursor-encoding transcripts were conducted on or before July 19, 2019 using a well-established protocol. In brief, the database of the online program tblastn (National Center for Biotechnology Information, Bethesda, MD; <http://blast.ncbi.nlm.nih.gov/Blast.cgi>) was set to Transcriptome Shotgun Assembly (TSA) and restricted to data from the Penaeoidea (taxid:111520), Sergestoidea (taxid:111521), Caridea (taxid:6694), Astacidea (taxid:6712), Achelata (taxid:6730), or Brachyura (taxid:6752). An *A. mellifera* ALP precursor (Accession No. **XP_003249809**; unpublished direct GenBank submission) was used as the query sequence for all BLAST searches.

2.2.2. Peptide structural prediction

The putative mature structures of decapod ALPs (as well as ALP precursor-related peptides [ALP-PRPs]) were predicted using a well-vetted workflow. Specifically, all hits returned by a given BLAST search were translated using the ExPASy Translate tool (<http://web.expasy.org/translate/>) and assessed for completeness. Proteins listed as full-length exhibit a functional signal sequence (including a start methionine) and are flanked on their carboxyl (C)-terminus by a stop codon. Proteins listed as C-terminal fragments lack a start methionine. Each full-length precursor was assessed for the presence of a signal peptide using the online program SignalP 3.0 (<http://www.cbs.dtu.dk/services/SignalP/>; Bendtsen et al., 2004). Prohormone cleavage sites were identified based on information presented in Veenstra (2000). The sulfation state of tyrosine residues was predicted using the online program Sulfinator (<http://www.expasy.org/tools/sulfinator/>; Monigatti et al., 2002). Disulfide bonding between cysteine residues was assessed using the online programs DiANNA (<http://clavus.bc.edu/~clotelab/DiANNA/>; Ferre and Clote, 2005) and DISULF-IND (<http://disulfind.dsi.unifi.it/>; Ceroni et al., 2006). Other post-

A. *Litopenaeus vannamei* (Dendrobranchiata; Penaeoidea) ALP precursors

Prepro-ALP-v1 MGSKVVMVLLALTLVLSVMA**QPLLEEGREDEGQVQAEVDYTDLLERLLSRTQKQDDLA**E
 Prepro-ALP-v2 MGSKVVMVLLALTLVLSVMA**QPLLEEGREDEGQVQAEVDYTDLLERLLSRTQKRSSYIY**
 Prepro-ALP-v3 MGSKVVMVLLALTLVLSVMA**QPLLEEGREDEGQVQAEVDYTDLLERLLSRTQK-----**

 Prepro-ALP-v1 **TA**PV**KR****W**R**S**C**I**R**R**G**A**C**D**H**R**P**N**D**C**C**N**S**S**C**R**C**N**L**G**T**N**C**R**C**Q**R**M**G**L**F**Q**Q**W**G**K**
 Prepro-ALP-v2 **L**F**R**-----**R**S**C**I**R**R**G**A**C**D**H**R**P**N**D**C**C**N**S**S**C**R**C**N**L**G**T**N**C**R**C**Q**R**M**G**L**F**Q**Q**W**G**K
 Prepro-ALP-v3 -----**R**S**C**I**R**R**G**A**C**D**H**R**P**N**D**C**C**N**S**S**C**R**C**N**L**G**T**N**C**R**C**Q**R**M**G**L**F**Q**Q**W**G**K

B. *Acetes chinensis* (Dendrobranchiata; Sergestoidea) ALP precursors

Prepro-ALP-v1 MKLLALVLA**L**SL**T**VALA**QPLIDEGEPSAAAESGQPDVYLLERLLAQ**T**Q**K**Q**D**M**VE**A**P**A**K**R**
 Prepro-ALP-v2 MKLLALVLA**L**SL**T**VALA**QPLIDEGEPSAAAESGQPDVYLLERLLAQ**T**Q**K**R**SS**S**H**I**Y**L**F**R**-----
 Prepro-ALP-v3 MKLLALVLA**L**SL**T**VALA**QPLIDEGEPSAAAESGQPDVYLLERLLAQ**T**Q**K-----

 Prepro-ALP-v1 **W**R**S**C**I**R**R**NG**V**C**D**H**R**P**N**D**C**C**N**S**S**C**R**C**N**L**G**T**N**C**R**C**Q**R**M**G**L**F**Q**Q**W**G**K**
 Prepro-ALP-v2 -----**R**S**C**I**R**R**NG**V**C**D**H**R**P**N**D**C**C**N**S**S**C**R**C**N**L**G**T**N**C**R**C**Q**R**M**G**L**F**Q**Q**W**G**K
 Prepro-ALP-v3 -----**R**S**C**I**R**R**NG**V**C**D**H**R**P**N**D**C**C**N**S**S**C**R**C**N**L**G**T**N**C**R**C**Q**R**M**G**L**F**Q**Q**W**G**K

C. *Macrobrachium australiense* (Pleocyemata; Caridea) ALP precursors

Prepro-ALP-I-v1 MGSKVLMVLA**F**SLVLSIAMA**QPLVEDEPFYILDHEEDAPPEADYTTDLERLLARTAQ**
 Prepro-ALP-I-v2 MGSKVLMVLA**F**SLVLSIAMA**QPLVED**-----**D**HEEDAPPEADYTTDLERLLARTAQ
 Prepro-ALP-I-v3 MGSKVLMVLA**F**SLVLSIAMA**QPLVEDEPFYILDHEEDAPPEADYTTDLERLLARTAQ**
 Prepro-ALP-I-v4 MGSKVLMVLA**F**SLVLSIAMA**QPLVED**-----**D**HEEDAPPEADYTTDLERLLARTAQ
 Prepro-ALP-II-v1 MGSKVLMVLA**F**SLVLSIAMA**QPLVEDEPFYILDHEEDAPPEADYTTDLERLLARTAQ**
 Prepro-ALP-II-v2 MGSKVLMVLA**F**SLVLSIAMA**QPLVED**-----**D**HEEDAPPEADYTTDLERLLARTAQ
 Prepro-ALP-II-v3 MGSKVLMVLA**F**SLVLSIAMA**QPLVED**-----**D**HEEDAPPEADYTTDLERLLARTAQ

 Prepro-ALP-I-v1 **K**R**D**D**M**A**G**P**I**K**R****W**R**S**C**I**R**R**SG**S**C**D**H**R**P**H**D**C**C**N**S**S**C**R**C**N**L**G**T**N**C**R**C**Q**R**M**G**L**F**Q**Q**W**G**K**
 Prepro-ALP-I-v2 **K**R**D**D**M**A**G**P**I**K**R****W**R**S**C**I**R**R**SG**S**C**D**H**R**P**H**D**C**C**N**S**S**C**R**C**N**L**G**T**N**C**R**C**Q**R**M**G**L**F**Q**Q**W**G**K**
 Prepro-ALP-I-v3 **K**-----**R**S**C**I**R**R**SG**S**C**D**H**R**P**H**D**C**C**N**S**S**C**R**C**N**L**G**T**N**C**R**C**Q**R**M**G**L**F**Q**Q**W**G**K
 Prepro-ALP-I-v4 **K**-----**R**S**C**I**R**R**SG**S**C**D**H**R**P**H**D**C**C**N**S**S**C**R**C**N**L**G**T**N**C**R**C**Q**R**M**G**L**F**Q**Q**W**G**K
 Prepro-ALP-II-v1 **K**R**D**D**M**A**G**P**I**K**R****W**R**S**C**I**P**R**G**G**A**C**T**H**R**P**Q**A**C**N**S**S**S**C**R**C**N**L**G**T**N**C**R**C**Q**R**M**G**L**F**Q**Q**L**G**K
 Prepro-ALP-II-v2 **K**R**D**D**M**A**G**P**I**K**R****W**R**S**C**I**P**R**G**G**A**C**T**H**R**P**Q**A**C**N**S**S**S**C**R**C**N**L**G**T**N**C**R**C**Q**R**M**G**L**F**Q**Q**L**G**K
 Prepro-ALP-II-v3 **K**-----**R**S**C**I**P**R**G**G**A**C**T**H**R**P**Q**A**C**N**S**S**S**C**R**C**N**L**G**T**N**C**R**C**Q**R**M**G**L**F**Q**Q**L**G**K**

 * **** *.*: * ***: ** .*****

D. *Homarus americanus* (Pleocyemata; Astacidea) ALP precursors

Prepro-ALP-v1 MGSKVLMVLA**L**ALVLSVMA**QPLLEEGREDEGQVQAEVDYAADLLERLLARTQ**K**R**D**D**V**A**
 Prepro-ALP-v2 MGSKVLMVLA**L**ALVLSVMA**QPLLEEGREDEGQVQAEVDYAADLLERLLARTQ**K**R**S-----
 Prepro-ALP-v3 MGSKVLMVLA**L**ALVLSVMA**QPLLEEGREDEGQVQAEVDYAADLLERLLARTQ**K-----

 Prepro-ALP-v1 **G**S**D**P**I**K**R****W**R**S**C**I**R**R**GG**M**C**D**H**R**P**N**D**C**C**N**S**S**C**R**C**N**L**G**T**N**C**R**C**Q**R**M**G**I**F**Q**Q**W**G**K**
 Prepro-ALP-v2 -----**S**Y**I**Y**L**F**R****S**C**I**R**R**GG**M**C**D**H**R**P**N**D**C**C**N**S**S**C**R**C**N**L**G**T**N**C**R**C**Q**R**M**G**I**F**Q**Q**W**G**K**
 Prepro-ALP-v3 -----**R**S**C**I**R**R**GG**M**C**D**H**R**P**N**D**C**C**N**S**S**C**R**C**N**L**G**T**N**C**R**C**Q**R**M**G**I**F**Q**Q**W**G**K

E. *Jasus edwardsii* (Pleocyemata; Achelata) ALP precursors

Prepro-ALP-v1 MGSKVLMVLA**L**SLVLT**L**VTG**QPLIEEGREDEAIQTPEMDYSADLMERLLASQ**K**R**D**D**L**A**
 Prepro-ALP-v2 MGSKVLMVLA**L**SLVLT**L**VTG**QPLIEEGREDEAIQTPEMDYSADLMERLLASQ**K-----

 Prepro-ALP-v1 **G**V**P**P**V**K**R****W**R**S**C**V**R**R**GG**P**C**D**H**R**P**N**D**C**C**N**S**S**C**R**C**N**L**G**T**N**C**R**C**Q**R**M**G**L**F**Q**Q**W**G**K**
 Prepro-ALP-v2 -----**R**S**C**V**R**GG**P**C**D**H**R**P**N**D**C**C**N**S**S**C**R**C**N**L**G**T**N**C**R**C**Q**R**M**G**L**F**Q**Q**W**G**K**

F. *Carcinus maenas* (Pleocyemata; Brachyura) ALP precursors

Prepro-ALP-v1 MGSKTAVMVA**L**SLVSVVLA**QPLLEEGREPDMQQA**EVDY**TADLLDHL**L**G**R**A**Q**N**R**E**D**T**A
 Prepro-ALP-v2 MGSKTAVMVA**L**SLVSVVLA**QPLLEEGREPDMQQA**EVDY**TADLLDHL**L**G**R**A**Q**K**R**S**P**D**I
 Prepro-ALP-v3 MGSKTAVMVA**L**SLVSVVLA**QPLLEEGREPDMQQA**EVDY**TADLLDHL**L**G**R**A**Q**K**-----

 Prepro-ALP-v1 **A**I**A**R**D**K**K****W**R**S**C**I**R**R**GG**A**C**D**H**R**P**N**D**C**C**N**S**S**C**R**C**N**L**G**T**N**C**R**C**Q**R**M**G**L**F**Q**Q**W**G**K**
 Prepro-ALP-v2 **Y**L**F**R-----**R**S**C**I**R**R**GG**A**C**D**H**R**P**N**D**C**C**N**S**S**C**R**C**N**L**G**T**N**C**R**C**Q**R**M**G**L**F**Q**Q**W**G**K
 Prepro-ALP-v3 -----**R**S**C**I**R**R**GG**A**C**D**H**R**P**N**D**C**C**N**S**S**C**R**C**N**L**G**T**N**C**R**C**Q**R**M**G**L**F**Q**Q**W**G**K

Fig. 1. MAFFT alignments of selected decapod agatotoxin-like peptide (ALP) precursor proteins. (A) Alignment of *Litopenaeus vannamei* ALP precursor variants. (B) Alignment of *Acetes chinensis* ALP precursor variants. (C) Alignment of *Macrobrachium australiense* ALP-I and II precursor variants. (D) Alignment of *Homarus americanus* ALP precursor variants. (E) Alignment of *Jasus edwardsii* ALP precursor variants. (F) Alignment of *Carcinus maenas* ALP precursor variants. In each protein, the signal peptide is shown in gray, the ALP isoform is shown in red, and the linker/precursor-related peptides are shown in blue. In the line immediately below each sequence grouping, the symbol “*” indicates amino acids that are identical in all proteins, while “.” and “:” denote amino acids that are similar in structure among all sequences.

translational modifications, i.e., cyclization of N-terminal glutamine/glutamic acid residues and C-terminal amidation at glycine residues, were predicted by homology to known arthropod peptides. All protein and peptide alignments were done using either the online program MAFFT version 7 (<http://mafft.cbrc.jp/alignment/software/>; Katoh and Standley, 2013) or default MUSCLE (Edgar, 2004) settings in Geneious v10.1.3 (Biomatters Ltd., Auckland, New Zealand; Kearse et al., 2012). Percent identities were calculated from either pairwise alignments or multiple sequence alignments using MUSCLE.

2.3. Assessment of phylogenetic relationships among agatotoxin-like precursor proteins

The phylogenetic relationships of selected decapod ALP precursor proteins were inferred from a multiple protein sequence alignment constructed with MUSCLE. Evolutionary analyses were conducted in MEGA 7 (Kumar et al., 2016) using the maximum likelihood method based on the Jones-Taylor-Thorton matrix model (Jones et al., 1992). Initial trees for the heuristic search were obtained automatically by

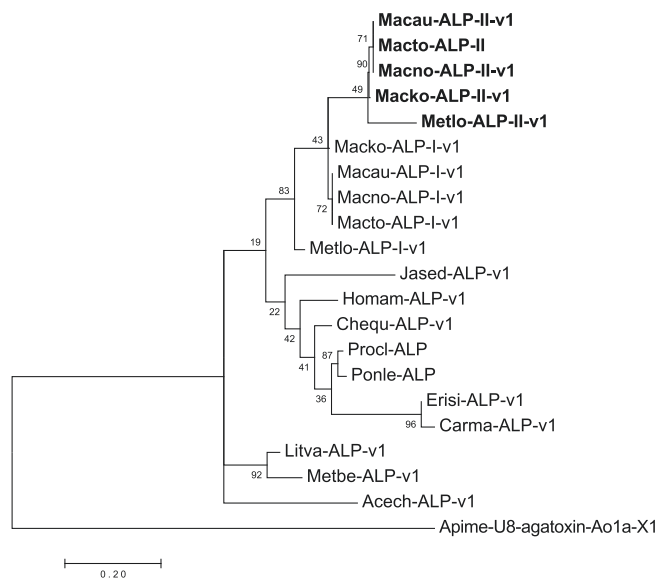


Fig. 2. Phylogenetic relationship of putative decapod agatotoxin-like peptide (ALP) precursors with those of other arthropod species. The evolutionary history was inferred using the maximum likelihood method with the highest log likelihood (-1307.53) tree shown. Bootstrap support (1000 iterations) is shown next to the branches. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. Type II ALP sequences are indicated in bold. Species abbreviations are: Macau, *Macrobrachium australiense*; Macno, *Macrobrachium novaehollandiae*; Macko, *Macrobrachium koombooloomba*; Macto, *Macrobrachium tolmerum*; Metlo, *Metabetaeus lohena*; Jased, *Jasus edwardsii*; Homam, *Homarus americanus*; Chequ, *Cherax quadricarinatus*; Procl, *Procambarus clarkii*; Ponle, *Pontastacus leptodactylus*; Erisi, *Eriocheir sinensis*; Carma, *Carcinus maenas*; Litva, *Litopenaeus vannamei*; Metbe, *Metapenaeus bennettiae*; Acech, *Acetes chinensis*; and Apime, *Apis mellifera*. The tree was rooted with *Apis mellifera* U8-agatotoxin-Ao1a isoform X1 (Accession No. XP_003249809). Accession numbers for sequences used in the phylogenetic tree are listed in [Supplementary Data 1](#).

applying Neighbor-Joining and BioNJ algorithms to a matrix of pairwise distances and then selecting the topology with the highest log likelihood value. A discrete gamma distribution was used to model evolutionary rate differences among sites (five categories [+G, parameter = 0.8139]). The analysis involved 21 amino acid sequences. All positions with less than 95% site coverage were eliminated; consequently, fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position. The final dataset consisted of a total of 110 positions. Analyses included bootstrap support for 1000 iterations. Phylogenetic inferences made using neighbor joining (Saitou and Nei, 1987) and minimum evolution (Rzhetsky and Nei, 1992) approaches generated trees with similar topologies. Accession numbers for sequences used in the phylogenetic analyses are provided in [Supplementary Data 1](#).

2.4. Cloning of agatotoxin-like peptide precursors from the lobster *Homarus americanus*

Total RNAs were purified from freshly dissected individual *H. americanus* supraoesophageal ganglia (brains) using TRI Reagent (Life Technologies, Carlsbad, CA) and Direct-zol RNA MiniPrep spin columns (Zymo Research, Irvine, CA, USA). First-strand complementary DNAs (cDNAs) were generated from 500 ng of DNase I-treated total RNA using Superscript III Reverse Transcriptase (Life Technologies) and custom made random pentadecamers (IDT, San Diego, CA, USA). The complete open reading frame for the putative *H. americanus* agatotoxin-like peptide precursor was PCR amplified using primers (sense –

A. Agatotoxin-like peptides from decapod Type-I precursors

	PSM	ESM
Penaeoidea-1	WRSCIRRMGACDHRPNDCCYNSSRCNLWGTNCRQRMGLFQQW	
Penaeoidea-2	--SCIRRMGACDHRPNDCCYNSSRCNLWGTNCRQRMGLFQQW	
Sergestoidea-1	WRSCIRRNQVCDHRPNDCCYNSSRCNFLTGTNCRQRMGLFQQW	
Sergestoidea-2	--SCIRRNQVCDHRPNDCCYNSSRCNFLTGTNCRQRMGLFQQW	
Caridea-1a	WRSCIRRSQSCDHRPNDCCYNSSRCNLWGTNCRQRMGLFQQW	
Caridea-1b	WRSCIRRSQSCDHRPNDCCYNSSRCNLWGTNCRQRMGLFQQW	
Caridea-1c	WRSCIRRSQSCDHRPNDCCYNSSRCNLWGTNCRQRMGLFQQW	
Caridea-2a	--SCIRRSQSCDHRPNDCCYNSSRCNLWGTNCRQRMGLFQQW	
Caridea-2b	--SCIRRSQSCDHRPNDCCYNSSRCNLWGTNCRQRMGLFQQW	
Caridea-2c	--SCIRRSQSCDHRPNDCCYNSSRCNLWGTNCRQRMGLFQQW	
Caridea-2d	--SCIRRSQSCDHRPNDCCYNSSRCNLWGTNCRQRMGLFQQW	
Astacidea-1a	WRSCIRRGTCGCDHRPNDCCYNSSRCNLWGTNCRQRMGLFQQW	
Astacidea-1b	WRSCIRRGTCGCDHRPNDCCYNSSRCNLWGTNCRQRMGLFQQW	
Astacidea-2a	--SCIRRGTCGCDHRPNDCCYNSSRCNLWGTNCRQRMGLFQQW	
Astacidea-2b	--SCIRRGTCGCDHRPNDCCYNSSRCNLWGTNCRQRMGLFQQW	
Achelata-1	WRSCVRRGGPCDHRPNDCCYNSSRCNLWGTNCRQRMGLFQQW	
Achelata-2	--SCVRRGGPCDHRPNDCCYNSSRCNLWGTNCRQRMGLFQQW	
Brachyura-1	WRSCIRRGACDHRPNDCCYNSSRCNLWGTNCRQRMGLFQQW	
Brachyura-2a	--SCIRRGACDHRPNDCCYNSSRCNLWGTNCRQRMGLFQQW	
Brachyura-2b	--SCIRRGACDHRPNDCCYNSSRCNLWGTNCRQRMGLFQQW	

B. Agatotoxin-like peptides from caridean Type-II precursors

	PSM	ESM
Macxx-1	WRSCIPRGACTHRPQACNSSSRCNLWGTNCRQRMGLFQQW	
Macxx-2	--SCIPRGACTHRPQACNSSSRCNLWGTNCRQRMGLFQQW	
Metlo-1	WRACIPRGGSCTHRPQACNSSSRCNLWGTNCRQRMGLFQQW	
Metlo-2	--ACIPRGGSCTHRPQACNSSSRCNLWGTNCRQRMGLFQQW	
Halru-2	--ACIPRSGGCTHRPQACNSSSRCNLWGTNCRQRMGLFQQW	

Fig. 3. MAFFT alignments of all predicted decapod agatotoxin-like peptide (ALP) isoforms. (A) Alignment of ALPs derived from decapod type-I precursors. (B) Alignment of ALPs derived from caridean type-II precursors. In the line immediately below each sequence grouping, the symbol “*” indicates amino acids that are identical in all proteins, while “.” and “:” denote amino acids that are similar in structure among all sequences. All peptides are predicted to possess amidated carboxyl-termini (not shown in the alignments). All peptides are also predicted to possess four identically positioned disulfide bridges (also not shown in the alignments). The conserved Cys motifs are indicated: PSM - principal structural motif; ESM - extra structural motif. Species identifications for A: Penaeoidea-1, *Litopenaeus vannamei* and *Metapenaeus bennettiae*; Penaeoidea-2, *L. vannamei*, *M. bennettiae*, *Penaeus monodon*, and *Fenneropenaeus penicillatus*; Sergestoidea-1, *Acetes chinensis*; Sergestoidea-2, *A. chinensis*; Caridea-1a, *Macrobrachium australiense*, *Macrobrachium koombooloomba*, *Macrobrachium novaehollandiae*, and *Macrobrachium tolmerum*; Caridea-1b, *Metabetaeus lohena*; Caridea-1c, *Halocaridina rubra*; Caridea-2a, *M. australiense*, *M. koombooloomba*, and *M. tolmerum*; Caridea-2b, *M. lohena*; Caridea-2c, *H. rubra*, *Halocaridinides trigonophthalma*, and *Antecaridina laevis*; Caridea-2d, *Neocaridina denticulata*; Astacidea-1a, *Cherax quadricarinatus*, *Pontastacus leptodactylus*, and *Procambarus clarkii*; Astacidea-1b, *Homarus americanus*; Astacidea-2a, *C. quadricarinatus*; Astacidea-2b, *H. americanus*; Achelata-1, *Jasus edwardsii*; Achelata-2, *J. edwardsii*; Brachyura-1, *Carcinus maenas* and *Eriocheir sinensis*; Brachyura-2a, *C. maenas* and *E. sinensis*; Brachyura-2b, *Portunus trituberculatus*. Species identifications for B: Macxx-1, *M. australiense*, *M. koombooloomba*, *M. novaehollandiae*, and *M. tolmerum*; Macxx-2, *M. australiense* and *M. koombooloomba*; Metlo-1, *M. lohena*; Metlo-2, *M. lohena*; Halru-2, *H. rubra*.

5'-ATGGGTAGCAAGGTGTTGG; antisense – 5'-TTATTCCCCCACTGCTGG) constructed to encompass the predicted start and stop sites in conjunction with Sapphire Amp Fast PCR Master Mix (Takara Bio USA Inc., Mountain View, CA). The primers were designed based on alignment of the *H. americanus* transcriptomic sequences GFDA01040082.1, GFUC01090606.1, and GFUC01090606.1. PCR was performed on a Biometra TRIO multiblock thermocycler (Biometra, Göttingen, Germany) in a 20-μL reaction volume containing 0.5 μL cDNA template and 0.2 μM of each primer. The thermocycler conditions consisted of: 95 °C for 2 min followed by 37 cycles at 95 °C for 20 s, 56 °C for 20 s, 72 °C for 30 s, and a final extension at 72 °C for 5 min. The resulting products were separated on a 1.5% agarose gel using a tris/acetate/EDTA buffer system with SYBR Safe (Life Technologies). The reaction products were subcloned into pCR2.1-TOPO TA (Life Technologies) and sequenced at the Arizona State University DNA Core Laboratory (Tempe, AZ).

A. *Homarus americanus* prepro-agatoxin-like peptide variant 1

MGSKVLVLLALALVLSVVM^QPLLEEGREEDGVQQAEPDYAADLLERLLARTQKRRDDVAGSDPIK^RWRSCIRRGGMCDHRPNDCYNSSSRCNLWG
TNCRCQRMGIFQ^QWGK

↓ Signal peptide peptidase (cleavage locus highlighted above)

QPLLEEGREEDGVQQAEPDYAADLLERLLARTQKRRDDVAGSDPIK^RWRSCIRRGGMCDHRPNDCYNSSSRCNLWG^TTNCRCQRMGIFQ^QWGK

↓ Prohormone processing protease (cleavage loci highlighted above)

QPLLEEGREEDGVQQAEPDYAADLLERLLARTQKRRDDVAGSDPIK^RWRSCIRRGGMCDHRPNDCYNSSSRCNLWG^TTNCRCQRMGIFQ^QWGK

↓ Carboxypeptidase (cleavage loci highlighted above)

QPLLEEGREEDGVQQAEPDYAADLLERLLARTQRDDVAGSDPIK^RWRSCIRRGGMCDHRPNDCYNSSSRCNLWG^TTNCRCQRMGIFQ^QWG^G

↓ Peptidylglycine α-hydroxylating monooxygenase/peptidyl-α-hydroxyglycine-α-amidating lyase (amidation locus highlighted above)

QPLLEEGREEDGVQQAEPDYAADLLERLLARTQRDWRSCIRRGGMCDHRPNDCYNSSSRCNLWG^TTNCRCQRMGIFQ^QWamide

↓ Tyrosylprotein sulfotransferase (tyrosine target highlighted above)

QPLLEEGREEDGVQQAEPDY^(SO3H)AADLLERLLARTQRDWRSCIRRGGMCDHRPNDCYNSSSRCNLWG^TTNCRCQRMGIFQ^QWamide

↓ Glutamyl cyclase (cyclization locus highlighted above)

PQPLLEEGREEDGVQQAEPDY^(SO3H)AADLLERLLARTQRDWRSCIRRGGMCDHRPNDCYNSSSRCNLWG^TTNCRCQRMGIFQ^QWamide

↓ Disulfide isomerase (bridged cysteines highlighted above)

WRS^cIRRGGMCDHRPNDCYNSSSRCNLWG^TTNCRCQRMGIFQ^QWamide

B. *Homarus americanus* prepro-agatoxin-like peptide variant 2

MGSKVLVLLALALVLSVVM^QPLLEEGREEDGVQQAEPDYAADLLERLLARTQKR^{SS}SIYLFRR^SSCIRRGGMCDHRPNDCYNSSSRCNLWG^TTNCRCQRMGIFQ^QWGK

↓ Signal peptide peptidase (cleavage locus highlighted above)

QPLLEEGREEDGVQQAEPDYAADLLERLLARTQKR^{SS}SIYLFRR^SSCIRRGGMCDHRPNDCYNSSSRCNLWG^TTNCRCQRMGIFQ^QWGK

↓ Prohormone processing protease (cleavage loci highlighted above)

QPLLEEGREEDGVQQAEPDYAADLLERLLARTQKR^{SS}SIYLFRR^SSCIRRGGMCDHRPNDCYNSSSRCNLWG^TTNCRCQRMGIFQ^QWGK

↓ Carboxypeptidase (cleavage loci highlighted above)

QPLLEEGREEDGVQQAEPDYAADLLERLLARTQ^{SS}SIYLFRR^SSCIRRGGMCDHRPNDCYNSSSRCNLWG^TTNCRCQRMGIFQ^QWG^G

↓ Peptidylglycine α-hydroxylating monooxygenase/peptidyl-α-hydroxyglycine-α-amidating lyase (amidation locus highlighted above)

QPLLEEGREEDGVQQAEPDYAADLLERLLARTQ^{SS}SCIRRGGMCDHRPNDCYNSSSRCNLWG^TTNCRCQRMGIFQ^QWamide

↓ Tyrosylprotein sulfotransferase (tyrosine target highlighted above)

QPLLEEGREEDGVQQAEPDY^(SO3H)AADLLERLLARTQ^{SS}SCIRRGGMCDHRPNDCYNSSSRCNLWG^TTNCRCQRMGIFQ^QWamide

↓ Glutamyl cyclase (cyclization locus highlighted above)

PQPLLEEGREEDGVQQAEPDY^(SO3H)AADLLERLLARTQ^{SS}SCIRRGGMCDHRPNDCYNSSSRCNLWG^TTNCRCQRMGIFQ^QWamide

↓ Disulfide isomerase (bridged cysteines highlighted above)

SCIRRGGMCDHRPNDCYNSSSRCNLWG^TTNCRCQRMGIFQ^QWamide

C. *Homarus americanus* prepro-agatoxin-like peptide variant 3

MGSKVLVLLALALVLSVVM^QPLLEEGREEDGVQQAEPDYAADLLERLLARTQKRSCIRRGGMCDHRPNDCYNSSSRCNLWG^TTNCRCQRMGIFQ^QWGK

↓ Signal peptide peptidase (cleavage locus highlighted above)

QPLLEEGREEDGVQQAEPDYAADLLERLLARTQKR^SSCIRRGGMCDHRPNDCYNSSSRCNLWG^TTNCRCQRMGIFQ^QWGK

↓ Prohormone processing protease (cleavage loci highlighted above)

QPLLEEGREEDGVQQAEPDYAADLLERLLARTQKR^SSCIRRGGMCDHRPNDCYNSSSRCNLWG^TTNCRCQRMGIFQ^QWGK

↓ Carboxypeptidase (cleavage loci highlighted above)

QPLLEEGREEDGVQQAEPDYAADLLERLLARTQ^SSCIRRGGMCDHRPNDCYNSSSRCNLWG^TTNCRCQRMGIFQ^QWG^G

↓ Peptidylglycine α-hydroxylating monooxygenase/peptidyl-α-hydroxyglycine-α-amidating lyase (amidation locus highlighted above)

QPLLEEGREEDGVQQAEPDYAADLLERLLARTQ^SSCIRRGGMCDHRPNDCYNSSSRCNLWG^TTNCRCQRMGIFQ^QWamide

↓ Tyrosylprotein sulfotransferase (tyrosine target highlighted above)

QPLLEEGREEDGVQQAEPDY^(SO3H)AADLLERLLARTQ^SSCIRRGGMCDHRPNDCYNSSSRCNLWG^TTNCRCQRMGIFQ^QWamide

↓ Glutamyl cyclase (cyclization locus highlighted above)

PQPLLEEGREEDGVQQAEPDY^(SO3H)AADLLERLLARTQ^SSCIRRGGMCDHRPNDCYNSSSRCNLWG^TTNCRCQRMGIFQ^QWamide

↓ Disulfide isomerase (bridged cysteines highlighted above)

SCIRRGGMCDHRPNDCYNSSSRCNLWG^TTNCRCQRMGIFQ^QWamide

Fig. 4. Theoretical processing schemes for the three lobster, *Homarus americanus*, agatoxin-like peptide (ALP) precursors. In this figure, the predicted structures of mature ALP isoforms are shown in red, with those of mature precursor-related peptides (PRPs) shown in blue. In the mature ALPs, the presence of cysteine residues that putatively participate in the formation of disulfide bridges are indicated by “c”. In the mature structure of the first PRP present in each precursor (the sole PRP in present in the precursor shown in panel C), a sulfated tyrosine residue is indicated by “Y(SO_{3H})”, while the presence of an amino-terminal pyroglutamic acid residue is indicated by “pQ”.

2.5. RT-PCR profiling of agatoxin-like peptide precursor expression in specific *Homarus americanus* nervous system regions

To examine neural expression of the agatoxin-like peptide precursor-encoding transcripts, RNAs were purified from three biological replicates of *H. americanus* brain (one brain per replicate), eyestalk ganglia (one eyestalk pair per replicate), and cardiac ganglion (ten cardiac ganglia per replicate). cDNAs were generated as described above from ~100 ng (cardiac ganglia) or 500 ng (brain and eyestalk ganglia) of DNase I-treated total RNAs. RT-PCR was performed in 20-μL reaction volumes containing 0.5 μL of each cDNA template and primers designed to amplify the complete agatoxin-like peptide precursor open reading frame (see above) as well as a 503-bp fragment of *H. americanus* actin (sense – 5'-GGTCGTACCACCGGTATT; antisense – 5'-CATCCTG

TCGGCAATTCC). Thermocycler conditions were as described above. RT-PCR was performed in triplicate. The resulting products were visualized on 3% agarose gels with images obtained using an AlphaImager Gel Documentation System (ProteinSimple, San Jose, CA) and processed in Photoshop CS6 v13.0 (Adobe Systems Inc., San Jose, CA).

2.6. Mass spectrometry

2.6.1. Tissue isolation and sample preparation

Lobsters were anaesthetized on ice for ~30 min, after which brains from individual lobsters (n = 8) were isolated via manual microdissection in chilled (8–10 °C) physiological saline (composition in mM: NaCl, 479.12; KCl, 12.74; CaCl₂, 13.67; MgSO₄, 20.00; Na₂SO₄, 3.91;

Table 2

Summary of agatoxin-like peptide (ALP) and ALP precursor-related peptide (ALP-PRP) isoforms detected in *H. americanus* supraoesophageal ganglion (brain) extracts.

Peptide	Precursor variant contained within	Structure	Predicted mass (Da) ^a	Error (ppm) ^b	Detection Summary ^c
ALP isoforms					
ALP-1	1	WRSclRRGGMcDHRPNDccYNSScRcNLWGTNcRcQRMGIFQQWamide	5290.19060	NA	0/8
ALP-2	2, 3	ScIRRGGMcDHRPNDccYNSScRcNLWGTNcRcQRMGIFQQWamide	4948.0102	NA	0/8
ALP-PRP isoforms					
ALP-PRP-1a	1, 2, 3	pQPLLEEGREEDGVQQAEPDY _(SO3H) AADLLERLLARTQ	3814.79542	3.9	7/8
ALP-PRP-1b	1, 2, 3	pQPLLEEGREEDGVQQAEPDYAADLLERLLARTQ	3734.83862	2.0	4/8
ALP-PRP-1c	1, 2, 3	QPLLEEGREEDGVQQAEPDY _(SO3H) AADLLERLLARTQ	3831.82197	0.1	4/8
ALP-PRP-1d	1, 2, 3	QPLLEEGREEDGVQQAEPDYAADLLERLLARTQ	3751.86517	0.3	6/8
ALP-PRP-2	1	DDVAGSDPI	887.38715	-1.2	7/8
ALP-PRP-3	2	SSYIYLF	891.43775	-2.4	8/8

^a Monoisotopic mass for [M + H]⁺.

^b Error (ppm) = ((Mmeasured - Mpredicted) * 106/Mpredicted).

^c Number of brain tissues where peptide was detected/total number of brain tissues analysed.

Trizma base, 11.45; maleic acid, 4.82 [pH, 7.45]) and placed in 58 µL of LCMS water (Fisherbrand; Life Technologies Corp.) in a 1.5 mL low retention tube. Samples were heated at 100 °C for 5 min to deactivate proteolytic enzymes (Stemmler et al., 2013), then cooled to room temperature for immediate analysis or stored at -20 °C.

To extract peptides, 142 µL of extraction solvent A (90.1% methanol; Fisher Scientific; LCMS-grade and 9.9% glacial acetic acid; SigmaAldrich, St. Louis, MO, USA; reagent grade; > 99%) was added to the tube. The brain/solvent mixture was homogenized using a motor-driven tissue grinder (Argos Technologies; Vernon Hills, IL) equipped with a polypropylene pestle (SigmaAldrich) for 2–3 min, sonicated for 5 min and centrifuged at 14.5 krpm (Mini-Spin Plus, Eppendorf; Hauppauge, NY, USA) for 5 min. The supernatant was transferred to a clean 1.5 mL tube and the remaining tissue pellet was re-suspended and re-homogenized in 50 µL of extraction solvent B (30% water; 65% methanol; 5% glacial acetic acid), sonicated for 5 min, and centrifuged at 14.5 krpm for 5 min; the supernatant was again removed. The combined supernatants were filtered through a 30 kDa molecular weight cut-off (MWCO) filter (Amicon Ultra-0.5 mL; Millipore, Burlington, MA, USA), which had been prewashed with two-200 µL volumes of extraction solvent B. MWCO filters were centrifuged at 14.5 krpm for up to 20 min, and the flow-through (approximately 200 µL) was collected for further processing. Samples were then subjected to delipidation with chloroform (n = 4) or high pH fractionation (n = 4). For delipidation, chloroform (NMR-grade 13CHCl₃; Cambridge Isotope Laboratories, Tewksbury, MA, USA; 200 µL) was added to the MWCO filter flow-through solution. The two-phase mixture was vigorously shaken or sonicated for ~ 1 min. The bottom organic layer was removed and discarded. Chloroform (200 µL) was added and the extraction was repeated. A portion of the top aqueous layer (30 µL) was mixed with 0.1% formic acid in water (LCMS-grade; Fisherbrand; 70 µL) for analysis by LCMS. With the goal of fractionating peptides and removing salts and phospholipids, brain extracts (n = 4) were fractionated using Pierce™ High pH Reversed-Phase Peptide Fractionation Kits (Thermo Fisher Scientific). In preparation for fractionation, the MWCO filter flow-through was dried and reconstituted in 300 µL of 0.1% TFA (Sigma Aldrich; ≥99.7%) in water (Fisher Scientific; LCMS-grade). The SPE columns were prepared following the manufacturer instructions; after sample loading and washing, peptides were eluted into 10 fractions using 300 µL steps of: 5%, 7%, 10%, 12.5%, 15%, 17.5%, 20%, 50%, 70%, and 80% acetonitrile, all in 0.1% trimethylamine in water (Thermo Fisher Scientific). The eluted peptides were dried and reconstituted in 40 µL of 4.5% formic acid (Thermo Fisher Scientific; 99+%) and 13% acetonitrile in water.

2.6.2. Liquid chromatography/mass spectrometry

Liquid chromatographic/mass spectrometric (LC/MS) analyses of individual *H. americanus* brain extracts (delipidated or fractionated) were performed using a 6530 quadrupole time-of-flight (Q-TOF) mass

analyzer (Agilent Technologies, Santa Clara, CA, USA). Mass spectra (MS and MS/MS) were collected in positive ion mode; the ionization voltage ranged from 1850 to 1975 V and the ion source temperature was held at 350 °C. Spectra were internally calibrated using dibutyl phthalate (C₁₆H₂₂O₄) and hexakis(1H, 1H, 4H-hexafluorobutyloxy) phosphazine (HP-1221; C₂₄H₁₈O₆N₃P₃F₃₆) continuously infused and detected as [M + H]⁺. Collision induced dissociation (CID)-MS/MS experiments were executed with precursor ions subjected to CID using nitrogen as the target gas. Chromatographic separation and nano-electrospray ionization (nanoESI) were performed using a 1260 Chip Cube system (Agilent Technologies) using a ProtID-chip with a 40 nL enrichment column and a 150 mm × 75 µm analytical column (Agilent Technologies). The enrichment and analytical columns were packed with 300 Å, 5 µm particles with Zorbax 300SB-C18 stationary phase. The mobile phases were 0.1% formic acid/water (A) and 0.1% formic acid/2% water/acetonitrile (B). Samples (1–16 µL) were loaded on the enrichment column using 99:1 (A:B) at 4 µL/min. Samples were analyzed with the 150-mm analytical column using either a linear gradient of 99:1 (A:B) for 1 min to 65:35 (A:B) at 40 min, to 30:70 (A:B) at 45 min and 0:100 (A:B) at 50.0 min or a longer linear gradient of 98:2 (A:B) for 1 min to 65:35 (A:B) at 130 min, to 40:60 (A:B) at 133 min and 0:100 (A:B) at 140.0 min using a flow rate of 0.3 µL/min. LC/MS figures were generated by exporting Mass Hunter (Agilent Technologies) chromatograms or mass spectral data as text files, graphing the data using Prism 7 (GraphPad Software, San Diego, CA), and importing these graphics into CorelDRAW X4 (Corel Corporation, Austin, TX, USA) for final figure production.

3. Results and discussion

3.1. Transcriptomic data suggest agatoxin-like peptides are broadly conserved in the Decapoda

To assess the extent to which the presence of ALPs is conserved in members of the Decapoda, the publicly accessible TSA datasets for the Penaeoidea (penaeid shrimp), Sergestoidea (sergestid shrimp), Caridea (caridean shrimp), Astacidea (clawed lobsters and freshwater crayfish), Achelata (spiny/slipper lobsters), and Brachyura (true crabs) were searched for putative ALP precursor-encoding transcripts using an ALP precursor from the honey bee, *A. mellifera*, as the query protein. These searches resulted in the identification of putative ALP-encoding sequences from 22 species, including one or more members of each of the abovementioned taxa (Table 1). For the Penaeoidea, putative ALP precursor-encoding transcripts were found in *Litopenaeus vannamei*, *Penaeus monodon*, *Metapenaeus bennettiae*, and *Fenneropenaeus penicillatus* transcriptomes, while for the Sergestoidea, putative ALP-encoding sequences were identified from an *Acetes chinensis* assembly. For

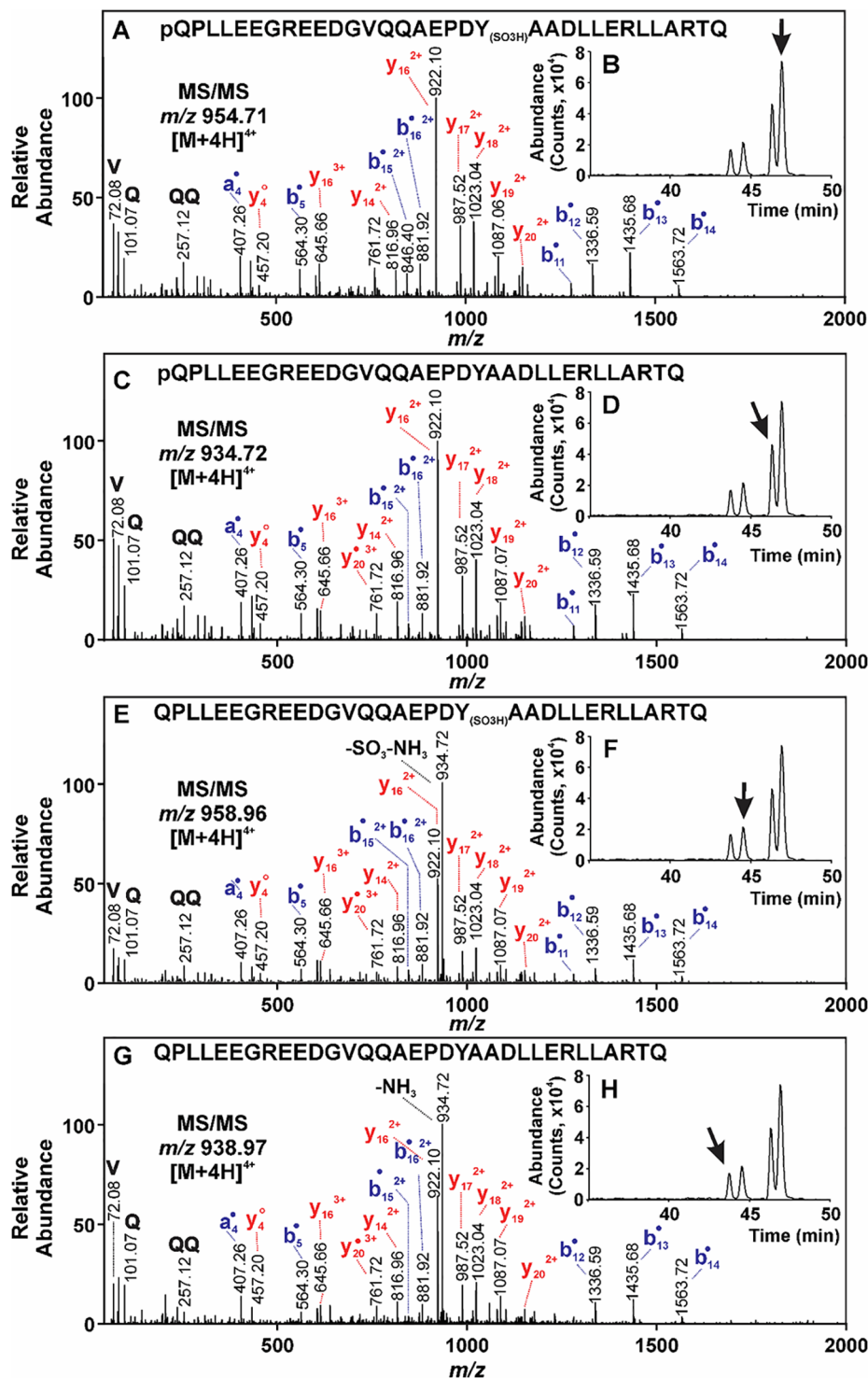


Fig. 5. Mass spectrometric identification of mature and immature forms of agatoxin-like peptide precursor-related peptide 1 (ALP-PRP-1) from *Homarus americanus* brain extract. (A) MS/MS spectrum for the m/z 954.71, $[M+4H]^{4+}$ ion from pQPLLEEGREEDGVQQAEPDY_(SO3H)AADLLERLLARTQ (ALP-PRP-1a) at a collision energy of 30.6 eV; (B) Summed extracted ion chromatograms (EICs) for the $[M+4H]^{4+}$ and $[M+3H]^{3+}$ peaks from ALP-PRP-1a, b, c, d; ALP-PRP-1a eluted last and was detected with the highest signal intensity; (C) MS/MS spectrum for the m/z 934.72, $[M+4H]^{4+}$ ion from pQPLLEEGREEDGVQQAEPDYAADLLERLLARTQ (ALP-PRP-1b) at a collision energy of 30.0 eV; (D) Summed extracted ion chromatograms (EICs) for the $[M+4H]^{4+}$ and $[M+3H]^{3+}$ peaks from ALP-PRP-1a, b, c, d; ALP-PRP-1b eluted third and was detected with the second highest signal intensity; (E) MS/MS spectrum for the m/z 958.96, $[M+4H]^{4+}$ ion from QPLLEEGREEDGVQQAEPDY_(SO3H)AADLLERLLARTQ (ALP-PRP-1c) at a collision energy of 30.8 eV; (F) Summed extracted ion chromatograms (EICs) for the $[M+4H]^{4+}$ and $[M+3H]^{3+}$ peaks from ALP-PRP-1a, b, c, d; ALP-PRP-1c eluted second and was detected with the third highest signal intensity; (G) MS/MS spectrum for the m/z 938.97, $[M+4H]^{4+}$ ion from QPLLEEGREEDGVQQAEPDYAADLLERLLARTQ (ALP-PRP-1d) at a collision energy of 30.2 eV; (H) Summed extracted ion chromatograms (EICs) for the $[M+4H]^{4+}$ and $[M+3H]^{3+}$ peaks from ALP-PRP-1a, b, c, d; ALP-PRP-1d eluted first and was detected with the lowest signal intensity. The assigned sequences were supported by N-terminus containing b-type product ions, many of which lost CO to produce a-type ions. C-terminus containing y-type ions provided additional sequence support, as did the detection of internal product ions, including QQ, and immonium ions, including peaks for V, L, and Q. Ions that have lost NH₃ are shown with a filled circle; ions that have lost H₂O are shown with an open circle. Monoisotopic masses are displayed.

the Caridea, putative ALP precursor-encoding transcripts were identified from *Macrobrachium australiense*, *Macrobrachium koombooloomba*, *Macrobrachium novaeollandiae*, *Macrobrachium tolmerum*, *Metabetaeus lohena*, *Halocaridina rubra*, *Halocaridinides trigonophthalma*, *Neocaridina denticulate*, and *Antecaridina lauensis* datasets. For the Astacidea, transcripts encoding putative ALP precursors were identified in *Homarus americanus* (sequences confirmed via RT-PCR; Accession Nos. [MN304943-MN304945](#)), *Cherax quadricarinatus*, *Pontastacus leptodactylus*, and *Procambarus clarkii* assemblies, while for the Achelata, transcripts encoding putative ALPs were found in a *Jasus edwardsii* transcriptome. For the Brachyura, ALP precursor-encoding sequences

were found in *Carcinus maenas*, *Eriocheir sinensis*, and *Portunus trituberculatus* datasets. These transcriptomic data strongly suggest that the presence of the ALP family is highly, and likely ubiquitously, conserved in the major taxa of the Decapoda.

3.2. Most decapods appear to have two isoforms of agatoxin-like peptide, both derived from a single gene by alternative splicing

Translation of the putative decapod ALP precursor-encoding sequences suggests the presence of a single ALP gene that can be alternatively spliced to generate (typically) three variants in nearly all

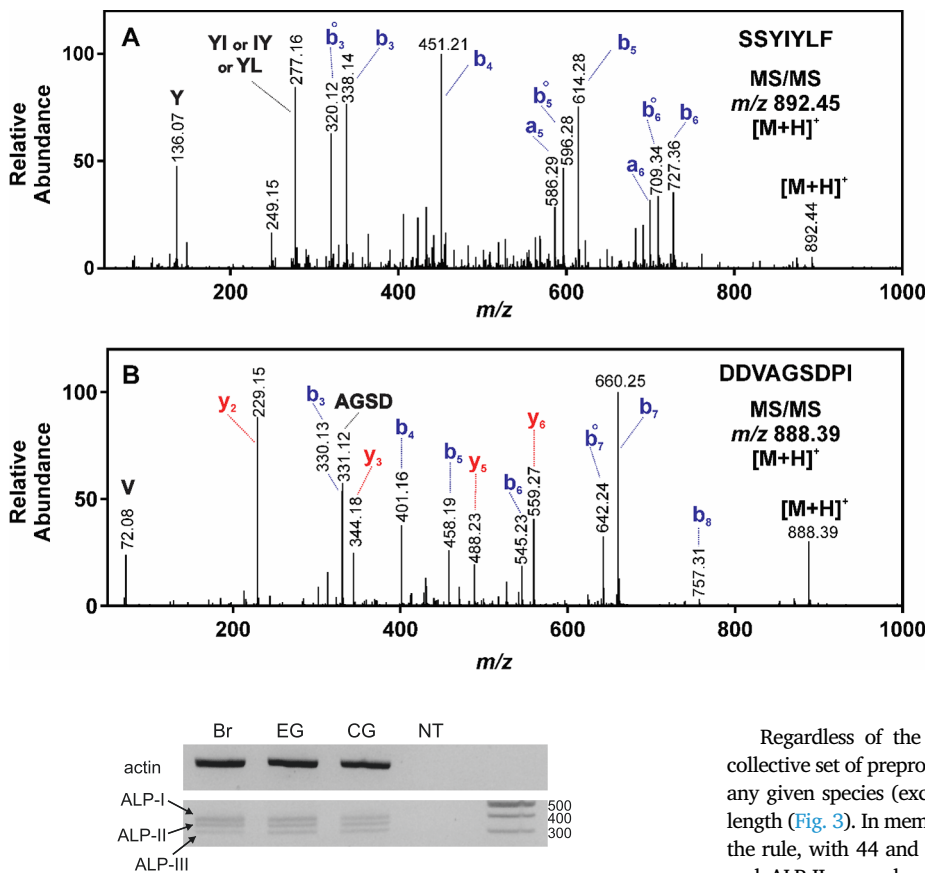


Fig. 7. RT-PCR profiling of agatoxin-like peptide (ALP) precursor-encoding transcripts in *Homarus americanus* neural tissues. The three ALP variants (342, 330, and 303-bp) were amplified using a single primer set from all three tissues examined. As a positive control, a 503-bp fragment of *H. americanus* actin was likewise amplified. The 3% agarose gel image shown is representative of three independent reactions using different cDNA replicates. Abbreviations are: Br, brain; EG, eyestalk ganglia; CG, cardiac ganglia; NT, RT-PCR reaction without template.

species (Fig. 1 and Supplementary Data 2). The one possible exception is in the Caridea, in which a duplication of either the ancestral ALP gene (named ALP-I in members of this infraorder) or one of the terminal exons appears to have resulted in a paralog (caridean ALP-II) that also undergoes alternative splicing (Fig. 1B and Supplementary Data 2). Comparison of the caridean ALP-I and -II open reading frames across the first 198 nucleotides revealed ~88% sequence identity that was independent of paralog type, whereas percent identity across the remaining 138 nucleotides segregated based on paralog type with identity > 95% (Supplementary Data 3A). When compared with the same region of the *H. americanus* and *P. clarkii* sequences, sequence identity was highest with the type I paralogs. Alignment of this region of the caridean ALP-I and -II sequences revealed clear areas of conservation (Supplementary Data 3B) that are consistent with a caridean-specific exon duplication event. In this scenario, the respective variants could be generated from homologous equivalent exon substitution (Supplementary Data 3C), a prevalent alternative splice mechanism that substitutes one homologous, possibly duplicated, exon for another (Abascal et al., 2015; Ezkurdia et al., 2012). Comparison of selected decapod ALP precursor sequences, including caridean ALP-I and ALP-II proteins, shows that the caridean ALP-II precursors cluster in a clade distinct from the ALP-I sequences, providing additional support for their paralog status (Fig. 2, Supplementary Data 4). The functional relevance of the caridean-specific ALP-I and -II sequences remains to be determined.

Fig. 6. Mass spectrometric identification of agatoxin-like peptide precursor-related peptide (ALP-PRP)-2 and -3 from *Homarus americanus* brain extract. (A) MS/MS spectrum for the m/z 888.39, $[M + H]^+$ ion from DDVAGSDPI (ALP-PRP-2) at a collision energy of 28.7 eV. (B) MS/MS spectrum for the m/z 892.45, $[M + H]^+$ ion from SSIYILF (ALP-PRP-3) at a collision energy of 28.8 eV. The assigned sequences were supported by N-terminus containing b-type product ions, many of which lost CO to produce a-type ions. C-terminus containing y-type ions provided additional sequence support, as did the detection of internal product ions and immonium ions. Ions that have lost H_2O are shown with an open circle. Monoisotopic masses are displayed.

Regardless of the number of precursor variants, it appears that the collective set of preprohormones gives rise to two distinct isoforms of ALP in any given species (except carideans), which are 44 and 42 amino acids in length (Fig. 3). In members of the Caridea, four isoforms of ALP appear to be the rule, with 44 and 42 amino acid peptides derived from both the ALP-I and ALP-II preprohormones (Fig. 3). All decapod ALPs appear to be C-terminally amidated. All also possess eight identically positioned cysteine residues (positions 4, 11, 18, 19, 24, 26, 33, and 35 in the 44 amino acid peptides, including those derived from the caridean ALP-II precursors; Fig. 3), which are theorized to give rise to four identically-positioned disulfide bridges; the positioning of these bonds remains unclear as predictions of their locations by DiANNA and DISULFIND showed considerable variability both between programs and among ALP isoforms (data not shown). Comparison of the amino acids at other positions in decapod ALP isoforms (those derived from ALP-I for the carideans) reveals extremely high levels of residue conservation in these peptides, with substitutions/deletions seen at just nine positions (Fig. 3A), positions 1 and 2 in the 44 amino acid peptides (tryptophan-arginine in all isoforms), which are deleted in the 42 amino acid peptides, and substitutions at positions 5, 8, 10, 16, 28, 29, and 40 (numbering based on the 44 amino acid isoforms). Four of the substitutions are conservative (isoleucine or valine at position 5, asparagine or histidine at position 16, leucine or phenylalanine at position 28, and isoleucine or leucine at position 40), with the remaining three being non-conservative (methionine, asparagine, serine, or glycine at position 8, alanine, valine, serine, threonine, methionine, or proline at position 10, and tryptophan or leucine at position 29). The ALPs derived from the caridean ALP-II precursors also show extremely high levels of amino acid identity among species, varying at just eight positions, i.e., insertions/deletions at positions 1 and 2 and variable residues at positions 3, 8, 10, 16, 21, and 36, all of which are conservative substitutions (Fig. 3B). Interestingly, there is more variation between the ALP-I- and ALP-II-derived peptides from a given caridean species (e.g., Fig. 1B) than there is among the caridean ALP-I-derived isoforms and those from members of other decapod superfamilies/infraorders (Fig. 3A).

In addition to ALP, the decapod prepro-ALPs also contain one or more linker/precursor-related peptide (e.g., Fig. 1). In their predicted mature forms, these linker/precursor-related sequences show greater variability within and between species than do the ALP isoforms, though there is considerable amino acid sequence conservation across species, and even across members of different superfamilies/infraorders (e.g., Fig. 1). The

putative mature structures of all predicted PRPs derived from decapod ALP precursors are provided in [Supplementary Data 5](#).

To determine if the predicted mature forms of the two ALP isoforms were present in *H. americanus* (Fig. 4), we extracted brains from individual lobsters ($n = 8$) and analyzed the extracts using a chip-based nanoLC-QTOF-MS/MS instrument. Given the methods used, we expected the mature ALP isoforms to be challenging peptides to detect and characterize because of their size (~ 5 kDa; Table 2), which results in a reduction in signal-to-noise associated with increasing dispersal of signal across a broader distribution of isotopic combinations and charge states (Compton et al., 2011). Furthermore, the use of MS/MS for structural confirmation of putative ALPs would be impeded by the presence of disulfide bonds, which restrict formation of fragment ions under our low energy CID conditions and makes structural confirmation more challenging. In the eight brain samples examined, we were unable to unambiguously confirm that native forms of the peptides were present. To circumvent the challenges associated with analysis of the native peptides, we reduced, alkylated and trypsinized three brain extracts to apply a proteomic strategy to the peptide analysis (data not shown). This approach also failed to produce MS support for the direct detection of the two ALP isoforms.

In contrast, we were able to detect all three of the predicted precursor/linker peptides (Table 2 and Figs. 4–6). Homam-ALP-PRP-1, common to all three variants of the ALP-preprohormone and having a mature structure predicted to be pQPLLEEGREEDGVQQAEPDY_(SO₃H)AADLLERLLARTQ (ALP-PRP-1a), ionized to produce abundant $[M + 4H]^{4+}$ and $[M + 3H]^{3+}$ ions. The MS/MS spectrum of the $[M + 3H]^{3+}$ ion yielded an abundant product ion resulting from the neutral loss of SO_3 , which supported the presence of sulfation in the sequence. While the facile loss of SO_3 precluded our ability to verify the localization of the sulfate group to the single tyrosine (Y) residue, the MS/MS spectrum of the $[M + 4H]^{4+}$ ion (Fig. 5A) supported the amino acid sequence assignment through the detection of b-type ions, which included the N-terminus, and y-type ions, which included the C-terminus. We also detected and used MS/MS to characterize three additional immature forms of ALP-PRP-1, pQPLLEEGREEDGVQQAEPDYAADLLERLLARTQ (ALP-PRP-1b; Fig. 5C), QPLLEEGREEDGVQQAEPDY_(SO₃H)AADLLERLLARTQ (ALP-PRP-1c; Fig. 5E), and QPLLEEGREEDGVQQAEPDYAADLLERLLARTQ (ALP-PRP-1d; Fig. 5G). ALP-PRP-1a, the mature form of the peptide, was most strongly retained by the column and was detected with the highest signal intensity (Fig. 5B); ALP-PRP-1d, the most immature form of the peptide (no N-terminal pyroglutamate and no sulfation) was least strongly retained by the column and was detected with the lowest signal intensity (Fig. 5H). We also detected and characterized DDVAGSDPI (Homam-ALP-PRP-2; Table 2 and Fig. 6A), and SSYIYLF (Homam-ALP-PRP-3; Table 2 and Fig. 6B). These PRPs were specific to variant 1 and 2, respectively, of the *Homarus* ALP preprohormone. Whether or not any of the *H. americanus* ALP-PRPs represent functionally important or bioactive peptides remains unknown. However, the mass spectral detection of them in their predicted mature forms in the lobster brain suggests that they are not subject to rapid degradation and is a finding that may indicate that these peptides play a role as bioactive modulators or as peptides playing a functional role in preprohormone or peptide processing.

3.3. Possible functions of agatoxin-like peptides in members of the Decapoda

While the majority of the transcripts identified here as encoding putative ALP precursors come from transcriptomes generated using mixed tissues (either whole organism or selected multiple tissues) as the source of RNAs (e.g., Havird and Santos, 2016; Huerlimann et al., 2018; Li et al., 2015; Oliphant et al., 2018; Rahi et al., 2019; Santos et al., 2018; Souza et al., 2018; Sun et al., 2014; Tom et al., 2014; Verbruggen et al., 2015), some are from tissue or tissue region-specific assemblies (Armstrong et al., 2019; Christie et al., 2017, 2018a, 2018b; He et al., 2012; Northcutt et al., 2016). The transcripts identified from the latter datasets provide insight into the putative source(s) of ALP production in

members of the Decapoda. Of particular note are the ALP-encoding sequences identified from lobster, *H. americanus*, where all were identified from nervous system-specific assemblies (Table 1), including several region-specific ones, i.e., those for the brain (Christie et al., 2018a), eyestalk ganglia (Christie et al., 2017), and cardiac ganglion (Christie et al., 2018b). Their presence in these portions of the lobster nervous system was confirmed by RT-PCR (Fig. 7). ALP-encoding transcripts were also identified from multiple eyestalk transcriptomes (Table 1), i.e., those for the shrimp, *L. vannamei* (Wang et al., 2019), the crayfish, *P. clarkii* (Manfrin et al., 2015), and the crabs, *E. sinensis* (Xu et al., 2015) and *P. trituberculatus* (Lv et al., 2017), for which the neural eyestalk ganglia undoubtedly represent the largest single tissue source. Detection of ALP precursor-encoding transcripts in neural-specific assemblies suggests that ALPs may serve as neuropeptides, though precisely how they function within the nervous system remains to be determined. It is possible that they may act as locally-released and/or circulating neuromodulators, as many peptides produced by the nervous systems of decapods have been shown to do (e.g., Christie, 2011; Christie et al., 2010), exerting their effects by binding to a G-protein coupled receptor or some other type of cell surface receptor, thereby modifying the properties of their neuronal targets. However, no ALP receptor has been identified in any arthropod species. Alternatively, ALPs might exert their action by directly interacting with ion channels, possibly neuronal calcium channels, leading to a reduction in calcium influx, and hence decreased neurotransmitter release, which is the mode of action of spider venom-derived agatoxins (e.g., Bindokas and Adams, 1989; Adams et al., 1990). Indeed, Sturm et al. (2016) postulated that the spider agatoxins and the venom gland ALPs may have been co-opted from the ancestral ALPs, which persist in most arthropods, to act as toxic ligands of previously targeted signaling proteins. Functional genomics and/or pharmacological-based studies, however, will be needed to clarify this issue.

In addition to their localization in neural-specific transcriptomes, ALP precursor-encoding transcripts were identified from a testis-specific assembly for the crab, *E. sinensis* (He et al., 2012), and a hepatopancreas-specific transcriptome for the shrimp, *M. bennettiae* (Armstrong et al., 2019). Previous investigations have identified many “neuropeptides” in non-neuronal tissues, including the reproductive organs and hepatopancreas. For example, transcripts encoding adipokinetic hormone-corazonin-like peptide, allatostatin A, allatostatin B, allatostatin C, bursicon, CCHamide, corazonin, crustacean cardioactive peptide, crustacean hyperglycemic hormone (CHH), diuretic hormone 31, eclosion hormone, FMRFamide-like peptide, HIGSLYRamide, insulin-like peptide, inotocin, leucokinin, myosuppressin, neuroparsin, neuropeptide F, orcokinin, pigment dispersing hormone, pyrokinin, red pigment concentrating hormone, RYamide, short neuropeptide F, SIFamide, and tachykinin-related peptide precursors were recently identified from a testis-specific transcriptome for the crab, *Scylla olivacea* (Christie, 2016), with members of the CNMamide family recently identified from multiple decapod testis- and/or ovary-specific assemblies (Christie and Hull, 2019). The presence of ALPs in the testis suggests that members of this peptide family may be involved in male reproductive control and/or may signal male reproductive status to other regulatory systems.

A number of “neuropeptides” have likewise been predicted from transcriptomes derived from the hepatopancreas. These include transcripts encoding CHH family members identified in a shrimp, *M. bennettiae*, hepatopancreas-specific transcriptome (Armstrong et al., 2019), as well as several “neuropeptides”, again including CHH family members, in a hepatopancreas-specific transcriptome derived from the shrimp, *L. vannamei* (A.E. Christie, unpublished observations). Given the various functions of the hepatopancreas, the identification of ALP-encoding sequences from it suggests that ALP family members may be involved in the regulation of stress response, detoxification, and/or innate immunity in members of the Decapoda.

4. Summary and conclusions

In the study presented here, publicly accessible transcriptomes were used to investigate the phylogenetic and structural conservation, tissue localization, and putative functions of ALPs in decapod crustaceans. Transcripts encoding members of this peptide family were identified from one or more members of the Penaeoidea and Sergestoidea, two superfamilies of the suborder Dendrobranchiata, and one or more species from the Caridea, Astacidea, Achelata, and Brachyura, four infraorders of the suborder Pleocyemata. This diversity suggests that ALPs are broadly, and perhaps ubiquitously, conserved in decapods. Comparison of the mature structures of the ALPs predicted from the identified transcripts shows high levels of amino acid conservation, suggesting conserved function(s). The identification of ALP precursor-encoding transcripts in decapod nervous system-specific transcriptomes suggests that ALPs (and/or ALP-PRPs) may have a neuromodulatory role, while the presence of transcripts in the testis- and hepatopancreas-specific assemblies suggests that ALPs (and/or ALP-PRPs) may help regulate reproduction, stress response, detoxification, and/or innate immunity.

CRedit authorship contribution statement

Andrew E. Christie: Conceptualization, Data curation, Formal analysis, Writing - original draft, Writing - review & editing. **Cindy D. Rivera:** Data curation. **Catherine M. Call:** Data curation. **Patsy S. Dickinson:** Conceptualization, Formal analysis, Writing - original draft, Writing - review & editing. **Elizabeth A. Stemmler:** Conceptualization, Formal analysis, Writing - original draft. **J. Joe Hull:** Data curation, Formal analysis, Writing - original draft, Writing - review & editing.

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

Lisa Baldwin and Colin Brent are thanked for reading and commenting on an earlier version of this article. Mention of trade names or commercial products in this article is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the US Department of Agriculture; the USDA is an equal opportunity provider and employer.

Funding

This work was supported by funds from the National Science Foundation (IOS-1353023 [to AEC]; IOS-1856307 [to AEC]; IOS-1354567 [to PSD]; IOS-1856433 [to PSD and EAS]; CHE-1126657 [to EAS]), the National Institutes of Health (INBRE grant 8P20GM103423-12), the Cades Foundation (to AEC), and base CRIS funding from the US Department of Agriculture (Project #2020-22620-022-00D; to JJH).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ygcen.2020.113609>.

References

Abascal, F., Ezkurdia, I., Rodriguez-Rivas, J., Rodriguez, J.M., del Pozo, A., Vázquez, J., Valencia, A., Tress, M.L., Käll, L., 2015. Alternatively spliced homologous exons have ancient origins and are highly expressed at the protein level. *PLoS Comput. Biol.* 11 (6), e1004325.

- Adams, M.E., 2004. Agatoxins: ion channel specific toxins from the american funnel web spider, *Agelenopsis aperta*. *Toxicon* 43 (5), 509–525.
- Adams, M.E., Bindokas, V.P., Hasegawa, L., Venema, V.J., 1990. Omega-agatoxins: novel calcium channel antagonists of two subtypes from funnel web spider (*Agelenopsis aperta*) venom. *J. Biol. Chem.* 265, 861–867.
- Armstrong, E.K., Miller, A.D., Mondon, J.A., Greenfield, P.A., Stephenson, S.A., Tan, M.H., Gan, H.M., Hook, S.E., 2019. De novo assembly and characterisation of the greentail prawn (*Metapenaeus bennettiae*) hepatopancreas transcriptome – identification of stress response and detoxification transcripts. *Mar. Genomics* 47, 100677. <https://doi.org/10.1016/j.margen.2019.04.002>.
- Bao, C., Liu, F., Yang, Y., Lin, Q.i., Ye, H., 2020. Identification of peptides and their GPCRs in the peppermint shrimp *Lyssmata vittata*, a protandric simultaneous hermaphrodite species. *Front. Endocrinol.* 11.
- Bao, C., Yang, Y., Huang, H., Ye, H., 2015. Neuropeptides in the cerebral ganglia of the mud crab, *Scylla paramamosain*: transcriptomic analysis and expression profiles during vitellogenesis. *Sci. Rep.* 5 (1). <https://doi.org/10.1038/srep17055>.
- Bao, C., Yang, Y., Zeng, C., Huang, H., Ye, H., 2018. Identifying neuropeptide GPCRs in the mud crab, *Scylla paramamosain*, by combinatorial bioinformatics analysis. *Gen. Comp. Endocrinol.* 269, 122–130.
- Bendtsen, J.D., Nielsen, H., von Heijne, G., Brunak, S., 2004. Improved prediction of signal peptides: SignalP 3.0. *J. Mol. Biol.* 340 (4), 783–795.
- Bindokas, V.P., Adams, M.E., 1989. Omega-Aga-I: A presynaptic calcium channel antagonist from venom of the funnel web spider, *Agelenopsis aperta*. *J. Neurobiol.* 20 (4), 171–188.
- Bouzid, W., Verdenaud, M., Klopp, C., Ducancel, F., Noirot, C., Vétillard, A., 2014. De novo sequencing and transcriptome analysis for *Tetramorium bicarinatum*: a comprehensive venom gland transcriptome analysis from an ant species. *BMC Genomics* 15 (1), 987. <https://doi.org/10.1186/1471-2164-15-987>.
- Buckley, S.J., Fitzgibbon, Q.P., Smith, G.G., Ventura, T., 2016. In silico prediction of the G-protein coupled receptors expressed during the metamorphic molt of *Sagmariasus verreauxi* (Crustacea: Decapoda) by mining transcriptomic data: RNA-seq to repertoire. *Gen. Comp. Endocrinol.* 228, 111–127.
- Christie, A.E., 2011. Crustacean neuroendocrine systems and their signaling agents. *Cell Tissue. Res.* 345 (1), 41–67.
- Christie, A.E., 2016. Prediction of *Scylla olivacea* (Crustacea; Brachyura) peptide hormones using publicly accessible transcriptome shotgun assembly (TSA) sequences. *Gen. Comp. Endocrinol.* 230–231, 1–16.
- Christie, A.E., 2020. Identification of putative neuropeptidergic signaling systems in the spiny lobster, *Panulirus argus*. *Invert. Neurosci.* 20 (1). <https://doi.org/10.1007/s10158-020-0235-9>.
- Christie, A.E., Chi, M., Lameyer, T.J., Pascual, M.G., Shea, D.N., Stanhope, M.E., Schulz, D.J., Dickinson, P.S., Johnson, E.C., 2015. Neuropeptidergic signaling in the American lobster *Homarus americanus*: new insights from high-throughput nucleotide sequencing. *PLoS ONE* 10 (12), e0145964.
- Christie, A.E., Hull, J.J., 2019. What can transcriptomics reveal about the phylogenetic/structural conservation, tissue localization, and possible functions of CNMamide peptides in decapod crustaceans? *Gen. Comp. Endocrinol.* 282, 113217. <https://doi.org/10.1016/j.ygcen.2019.113217>.
- Christie, A.E., Roncalli, V., Cieslak, M.C., Pascual, M.G., Yu, A., Lameyer, T.J., Stanhope, M.E., Dickinson, P.S., 2017. Prediction of a neuropeptidome for the eyestalk ganglia of the lobster *Homarus americanus* using a tissue-specific de novo assembled transcriptome. *Gen. Comp. Endocrinol.* 243, 96–119.
- Christie, A.E., Stanhope, M.E., Gandler, H.I., Lameyer, T.J., Pascual, M.G., Shea, D.N., Yu, A., Dickinson, P.S., Hull, J.J., 2018a. Molecular characterization of putative neuropeptide, amine, diffusible gas and small molecule transmitter biosynthetic enzymes in the eyestalk ganglia of the American lobster, *Homarus americanus*. *Invert. Neurosci.* 18 (4). <https://doi.org/10.1007/s10158-018-0216-4>.
- Christie, A.E., Stemmler, E.A., Dickinson, P.S., 2010. Crustacean neuropeptides. *Cell. Mol. Life Sci.* 67 (24), 4135–4169.
- Christie, A.E., Yu, A., Pascual, M.G., Roncalli, V., Cieslak, M.C., Warner, A.N., Lameyer, T.J., Stanhope, M.E., Dickinson, P.S., Hull, J.J., 2018b. Circadian signaling in *Homarus americanus*: Region-specific de novo assembled transcriptomes show that both the brain and eyestalk ganglia possess the molecular components of a putative clock system. *Mar. Genomics* 40, 25–44.
- Ceroni, A., Passerini, A., Vullo, A., Frasconi, P., 2006. DISULFIND: a disulfide bonding state and cysteine connectivity prediction server. *Nucleic Acids Res.* 34 (Web Server), W177–W181.
- Christie, A.E., Yu, A., Roncalli, V., Pascual, M.G., Cieslak, M.C., Warner, A.N., Lameyer, T.J., Stanhope, M.E., Dickinson, P.S., Hull, J.J., 2018c. Molecular evidence for an intrinsic circadian pacemaker in the cardiac ganglion of the American lobster, *Homarus americanus* - Is diel cycling of heartbeat frequency controlled by a peripheral clock system? *Mar. Genomics* 41, 19–30.
- Compton, P.D., Zamborg, L., Thomas, P.M., Kelleher, N.L., 2011. On the scalability and requirements of whole protein mass spectrometry. *Anal. Chem.* 83 (17), 6868–6874.
- Dickinson, P.S., Hull, J.J., Miller, A., Oleisky, E.R., Christie, A.E., 2019. To what extent may peptide receptor gene diversity/complement contribute to functional flexibility in a simple pattern-generating neural network? *Comp. Biochem. Physiol. D: Genomics Proteomics* 30, 262–282.
- Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32 (5), 1792–1797.
- Ezkurdia, I., del Pozo, A., Frankish, A., Rodriguez, J.M., Harrow, J., Ashman, K., Valencia, A., Tress, M.L., 2012. Comparative proteomics reveals a significant bias toward alternative protein isoforms with conserved structure and function. *Mol. Biol. Evol.* 29 (9), 2265–2283.
- Ferre, F., Clote, P., 2005. DiANNA: a web server for disulfide connectivity prediction. *Nucleic Acids Res.* 33 (Web Server), W230–W232.

- Havird, J.C., Santos, S.R., 2016. Developmental transcriptomics of the Hawaiian an-chialine shrimp *Halocaridina rubra* Holthuis, 1963 (Crustacea: Atyidae). *Integr. Comp. Biol.* 56 (6), 1170–1182.
- He, L., Wang, Q., Jin, X., Wang, Y., Chen, L., Liu, L., Wang, Y., Liu, Z., 2012. Transcriptome profiling of testis during sexual maturation stages in *Eriocheir sinensis* using Illumina sequencing. *PLoS ONE* 7 (3), e33735.
- Huerlimann, R., Wade, N.M., Gordon, L., Montenegro, J.D., Goodall, J., McWilliam, S., Tinning, M., Siemerling, K., Giardina, E., Donovan, D., Sellars, M.J., Cowley, J.A., Condon, K., Coman, G.J., Khatkar, M.S., Raadsma, H.W., Maes, G.E., Zenger, K.R., Jerry, D.R., 2018. *De novo* assembly, characterization, functional annotation and expression patterns of the black tiger shrimp (*Penaeus monodon*) transcriptome. *Sci. Rep.* 8 (1). <https://doi.org/10.1038/s41598-018-31148-4>.
- Jones, D.T., Taylor, W.R., Thornton, J.M., 1992. The rapid generation of mutation data matrices from protein sequences. *Bioinformatics* 8 (3), 275–282.
- Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30 (4), 772–780.
- Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., Drummond, A., 2012. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28 (12), 1647–1649.
- Kumar, S., Stecher, G., Tamura, K., 2016. MEGA 7: Molecular Evolutionary Genetics Analysis version 7 for bigger datasets. *Mol. Biol. Evol.* 33, 1870–1874.
- Li, Y., Hui, M., Cui, Z., Liu, Y., Song, C., Shi, G., 2015. Comparative transcriptomic analysis provides insights into the molecular basis of the metamorphosis and nutrition metabolism change from zoeae to megalopae in *Eriocheir sinensis*. *Comp. Biochem. Physiol. D: Genomics Proteomics* 13, 1–9.
- 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.
- Liu, Z., Chen, S., Zhou, Y., Xie, C., Zhu, B., Zhu, H., Liu, S., Wang, W., Chen, H., Ji, Y., 2015. Deciphering the venom transcriptome of killer-wasp *Vespa velutina*. *Sci Rep* 5 (1). <https://doi.org/10.1038/srep09454>.
- Lv, J., Zhang, L., Liu, P., Li, J., Wegener, C., 2017. Transcriptomic variation of eyestalk reveals the genes and biological processes associated with molting in *Portunus trituberculatus*. *PLoS ONE* 12 (4), e0175315.
- Manfrin, C., Tom, M., De Moro, G., Gerdol, M., Giulianini, P.G., Pallavicini, A., 2015. The eyestalk transcriptome of red swamp crayfish *Procambarus clarkii*. *Gene* 557 (1), 28–34.
- Monigatti, F., Gasteiger, E., Bairoch, A., Jung, E., 2002. The Sulfator: predicting tyrosine sulfation sites in protein sequences. *Bioinformatics* 18 (5), 769–770.
- Northcutt, A.J., Lett, K.M., Garcia, V.B., Diester, C.M., Lane, B.J., Marder, E., Schulz, D.J., 2016. Deep sequencing of transcriptomes from the nervous systems of two decapod crustaceans to characterize genes important for neural circuit function and modulation. *BMC Genomics* 17 (1). <https://doi.org/10.1186/s12864-016-3215-z>.
- Nguyen, T.V., Cummins, S.F., Elizur, A., Ventura, T., 2016. Transcriptomic characterization and curation of candidate neuropeptides regulating reproduction in the eyestalk ganglia of the Australian crayfish, *Cherax quadricarinatus*. *Sci. Rep.* 6 (1). <https://doi.org/10.1038/srep38658>.
- Oliphant, A., Alexander, J.L., Swain, M.T., Webster, S.G., Wilcockson, D.C., 2018. Transcriptomic analysis of crustacean neuropeptide signaling during the moult cycle in the green shore crab, *Carcinus maenas*. *BMC Genomics* 19 (1). <https://doi.org/10.1186/s12864-018-5057-3>.
- Oliphant, A., Hawkes, M.K.N., Cridge, A.G., Dearden, P.K., 2020. Transcriptomic characterisation of neuropeptides and their putative cognate G protein-coupled receptors during late embryo and stage-1 juvenile development of the Aotearoa-New Zealand crayfish, *Paranephrops zealandicus*. *Gen. Comp. Endocrinol.* 292, 113443. <https://doi.org/10.1016/j.ygcen.2020.113443>.
- Rahi, M.L., Mather, P.B., Ezaz, T., Hurwood, D.A., 2019. The molecular basis of freshwater adaptation in prawns: Insights from comparative transcriptomics of three *Macrobrachium* species. *Genome Biol. Evol.* 11, 1002–1018.
- Rzhetsky, A., Nei, M., 1992. A simple method for estimating and testing minimum evolution trees. *Mol. Biol. Evol.* 9, 945–967.
- Saitou, N., Nei, M., 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4, 406–425.
- Santos, C.A., Andrade, S.C.S., Teixeira, A.K., Farias, F., Kurkjian, K., Guerrelhas, A.C., Rocha, J.L., Galetti Jr., P.M., Freitas, P.D., 2018. *Litopenaeus vannamei* transcriptome profile of populations evaluated for growth performance and exposed to white spot syndrome virus (WSSV). *Front. Genet.* 9. <https://doi.org/10.3389/fgene.2018.00120.s001>.
- Skinner, W.S., Adams, M.E., Quistad, G.B., Kataoka, H., Cesarin, B.J., Enderlin, F.E., Schooley, D.A., 1989. Purification and characterization of two classes of neurotoxins from the funnel web spider, *Agelenopsis aperta*. *J. Biol. Chem.* 264, 2150–2155.
- Souza, C.A., Murphy, N., Strugnelli, J.M., 2018. *De novo* transcriptome assembly and functional annotation of the southern rock lobster (*Jasus edwardsii*). *Mar. Geonomics* 42, 58–62.
- Stemmler, E.A., Barton, E.E., Esonu, O.K., Polasky, D.A., Onderko, L.L., Bergeron, A.B., Christie, A.E., Dickinson, P.S., 2013. C-terminal methylation of truncated neuro-peptides: an enzyme-assisted extraction artifact involving methanol. *Peptides* 46, 108–125.
- Sturm, S., Ramesh, D., Brockmann, A., Neupert, S., Predel, R., 2016. Agatoxin-like peptides in the neuroendocrine system of the honey bee and other insects. *J. Proteomics* 132, 77–84.
- Sun, Y., Zhang, Y., Liu, Y., Xue, S., Geng, X., Hao, T., Sun, J., Lin, S., 2014. Changes in the organics metabolism in the hepatopancreas induced by eyestalk ablation of the Chinese mitten crab *Eriocheir sinensis* determined via transcriptome and DGE analysis. *PLoS ONE* 9 (4), e95827.
- Tom, M., Manfrin, C., Mosco, A., Gerdol, M., De Moro, G., Pallavicini, A., Giulianini, P.G., 2014. Different transcription regulation routes are exerted by L- and D-amino acid enantiomers of peptide hormones. *J. Exp. Biol.* 217 (24), 4337–4346.
- Torres, A.F.C., Huang, C., Chong, C.-M., Leung, S.W., Prieto-da-Silva, A.R.B., Havt, A., Quinet, Y.P., Martins, A.M.C., Lee, S.M.Y., Rádis-Baptista, G., Ho, P.L., 2014. Transcriptome analysis in venom gland of the predatory giant ant *Dinoponera quadricipes*: insights into the polypeptide toxin arsenal of hymenopterans. *PLoS ONE* 9 (1), e87556.
- Tran, N.M., Mykles, D.L., Elizur, A., Ventura, T., 2019. Characterization of G-protein coupled receptors from the blackback land crab *Gecarcinus lateralis* Y organ transcriptome over the molt cycle. *BMC Genomics* 20 (1). <https://doi.org/10.1186/s12864-018-5363-9>.
- Veenstra, J.A., 2000. Mono- and dibasic proteolytic cleavage sites in insect neuroendocrine peptide precursors. *Arch. Insect Biochem.* 43 (2), 49–63.
- Veenstra, J.A., 2015. The power of next-generation sequencing as illustrated by the neuropeptidome of the crayfish *Procambarus clarkii*. *Gen. Comp. Endocrinol.* 224, 84–95.
- Veenstra, J.A., 2016. Similarities between decapod and insect neuropeptidomes. *PeerJ* 4, e2043.
- Ventura, T., Cummins, S.F., Fitzgibbon, Q., Battaglene, S., Elizur, A., Vaudry, H., 2014. Analysis of the central nervous system transcriptome of the eastern rock lobster *Sagmariasus verreauxi* reveals its putative neuropeptidome. *PLoS ONE* 9 (5), e97323.
- Verbruggen, B., Bickley, L.K., Santos, E.M., Tyler, C.R., Stentiford, G.D., Bateman, K.S., van Aerle, R., 2015. *De novo* assembly of the *Carcinus maenas* transcriptome and characterization of innate immune system pathways. *BMC Genomics* 16 (1). <https://doi.org/10.1186/s12864-015-1667-1>.
- von Reumont, B.M., Blanke, A., Richter, S., Alvarez, F., Bleidorn, C., Jenner, R.A., 2014. The first venomous crustacean revealed by transcriptomics and functional morphology: remiped venoms express a unique toxin cocktail dominated by enzymes and a neurotoxin. *Mol. Biol. Evol.* 31, 48–58.
- Wang, Z., Luan, S., Meng, X., Cao, B., Luo, K., Kong, J., 2019. Comparative transcriptomic characterization of the eyestalk in Pacific white shrimp (*Litopenaeus vannamei*) during ovarian maturation. *Gen. Comp. Endocrinol.* 274, 60–72.
- Xu, Z., Zhao, M., Li, X., Lu, Q., Li, Y., Ge, J., Pan, J., 2015. Transcriptome profiling of the eyestalk of precocious juvenile Chinese mitten crab reveals putative neuropeptides and differentially expressed genes. *Gene* 569 (2), 280–286.