6-1-2020

In silico analyses suggest the cardiac ganglion of the lobster, Homarus americanus, contains a diverse array of putative innexin/innexin-like proteins, including both known and novel members of this protein family

Andrew E. Christie
Pacific Biosciences Research Center

J. Joe Hull
USDA ARS U.S. Arid-Land Agricultural Research Center

Patsy S. Dickinson
Bowdoin College

Invertebrate Neuroscience (2020) 20:5
https://doi.org/10.1007/s10158-020-0238-6

SHORT COMMUNICATION

In silico analyses suggest the cardiac ganglion of the lobster, *Homarus americanus*, contains a diverse array of putative innexin/innexin-like proteins, including both known and novel members of this protein family

Andrew E. Christie[1] · J. Joe Hull[2] · Patsy S. Dickinson[3]

Received: 22 September 2019 / Accepted: 31 January 2020 / Published online: 2 March 2020

Abstract

Gap junctions are physical channels that connect adjacent cells, permitting the flow of small molecules/ions between the cytoplasms of the coupled units. Innexin/innexin-like proteins are responsible for the formation of invertebrate gap junctions. Within the nervous system, gap junctions often function as electrical synapses, providing a means for coordinating activity among electrically coupled neurons. While some gap junctions allow the bidirectional flow of small molecules/ions between coupled cells, others permit flow in one direction only or preferentially. The complement of innexins present in a gap junction determines its specific properties. Thus, understanding innexin diversity is key for understanding the full potential of electrical coupling in a species/system. The decapod crustacean cardiac ganglion (CG), which controls cardiac muscle contractions, is a simple pattern-generating neural network with extensive electrical coupling among its circuit elements. In the lobster, *Homarus americanus*, prior work suggested that the adult neuronal innexin complement consists of six innexins (Homam-Inx1-4 and Homam-Inx6-7). Here, using a *H. americanus* CG-specific transcriptome, we explored innexin complement in this portion of the lobster nervous system. With the exception of Homam-Inx4, all of the previously described innexins appear to be expressed in the *H. americanus* CG. In addition, transcripts encoding seven novel putative innexins (Homam-Inx8-14) were identified, four (Homam-Inx8-11) having multiple splice variants, e.g., six for Homam-Inx8. Collectively, these data indicate that the innexin complement of the lobster nervous system in general, and the CG specifically, is likely significantly greater than previously reported, suggesting the possibility of expanded gap junction diversity and function in *H. americanus*.

Keywords Gap junction · Electrical coupling · Electrical synapse · Central pattern generator · Cardiac neuromuscular system · Decapoda

Introduction

In decapod crustaceans, rhythmic control of the heart musculature is neurogenic rather than myogenic, with contractions driven by the cardiac ganglion (CG), an extension of the central nervous system located in the lumen of the heart (e.g., Cooke 2002). The decapod CG is one of the simplest rhythmically active neural networks known, consisting of just nine neurons: four small pacemaker neurons, alternatively referred to as the premotor neurons or small cells, and five larger motor neurons, alternatively termed the large cells (e.g., Cooke 2002). Due to its numerical simplicity, the comparatively large size and stereotyped positions of both the pacemaker and motor neurons, and its ability to

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s10158-020-0238-6) contains supplementary material, which is available to authorized users.

✉ Andrew E. Christie
crabman@pbrc.hawaii.edu

[1] 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

[2] Pest Management and Biocontrol Research Unit, US Arid Land Agricultural Research Center, USDA Agricultural Research Services, Maricopa, AZ 85138, USA

[3] Department of Biology, Bowdoin College, 6500 College Station, Brunswick, ME 04011, USA

be maintained in vitro for extended periods of time, the decapod CG has long served as one of the premiere models, invertebrate or vertebrate, for investigating the basic principles governing the generation, maintenance, and modulation of rhythmic motor behavior at the cellular and systems levels (e.g., Calabrese et al. 2016; Cooke 2002; Dickinson et al. 2016; Hooper and DiCaprio 2004; Otopalik et al. 2019; Schulz and Lane 2017).

The rhythmic output of the decapod CG is in part determined by a combination of electrical and chemical synapses among/between its circuit elements (e.g., Cooke 2002; Otopalik et al. 2019). Intrinsic rhythmic activity in the pacemaker neurons is responsible for driving burst rate and duration via excitatory electrical and chemical input to motor neurons. In turn, the motor neurons provide excitatory feedback to the pacemaker cells, via both electrical and chemical connections, allowing for prolongation of burst duration. With respect to electrical coupling, there is strong experimental evidence of electrical synapses among the four pacemaker neurons, among the five motor neurons, and between the pacemaker and motor neuron groups (e.g., Cooke 2002; Otopalik et al. 2019).

In both vertebrate and invertebrate nervous systems, gap junctions, protein pores that connect adjacent cells allowing for the flow of small molecules and ions between them, serve as the physical loci of the electrical synapses between neurons (e.g., Alcamí and Pereda 2019; Harris 2018; O'Brien 2019). Based on the specific properties/structures of the proteins involved in the formation of a given gap junction, some pores allow for a bidirectional flow of small molecules/ions between the coupled cells, while others are much more selective, with ion flow being exclusively or preferentially unidirectional (e.g., Alcamí and Pereda 2019; Harris 2018; O'Brien 2019). In addition, gap junctions are not static elements, but rather are capable of being modulated by a variety of mechanisms, e.g., modulation of their conductance by chemical compounds such as the amine dopamine (Lane et al. 2018), therein adding to neural network functional flexibility (e.g., Alcamí and Pereda 2019; Harris 2018; O'Brien 2019).

In vertebrate species, members of the connexin family are responsible for gap junction formation, while in invertebrates, innexin/innexin-like proteins serve this function (e.g., Beyer and Berthoud 2018; Pereda and Macagno 2017). Despite their common functional role, the connexins of vertebrates and the invertebrate innexins are not evolutionarily related (e.g., Beyer and Berthoud 2018; Pereda and Macagno 2017). In the species that have thus far been investigated thoroughly, multiple members of the connexin/innexin families have been identified (e.g., Kandarian et al. 2012; Starich et al. 2001; Stebbings et al. 2002; Willecke et al. 2002). The specific complement of connexins/innexins participating in the formation of a given gap junction determines the physical properties of the channel in question (e.g., Pereda et al. 2013; Phelan et al. 2008). Interestingly, while connexin genes typically consist of a single exon, multiple exons are present in most innexin genes, allowing for expanded protein diversity via alternative splicing (e.g., Phelan 2005). Thus, understanding the diversity of connexin/innexin family members and their variants is key to understanding the potential for gap junction diversity and function in a given species or tissue.

Recently, the complement of innexin/innexin-like proteins in the nervous system of the American lobster, *Homarus americanus*, was assessed using a mixed nervous system region transcriptome (Shruti et al. 2014). The portions of the nervous system used to generate this assembly were the supraesophageal ganglion (brain), abdominal nerve cord, stomatogastric nervous system, and CG (BioProject No. PRJNA300643; Northcutt et al. 2016). Transcripts encoding six different putative innexins were identified using this resource, Homam-Inx1-4 and Homam-Inx6-7 (Shruti et al. 2014); the lack of an innexin specifically named Inx5 in *H. americanus* reflects a potential limitation of the current datasets. The comparative analysis of innexins in the two decapod species, *H. americanus* and the crab, *Cancer borealis*, revealed both have Inx1-4 and Inx6 homologs, whereas an Inx5 was found only in *C. borealis* and Inx7 only in *H. americanus* (Shruti et al. 2014), with the two proteins only having 26% sequence identity. It could be that homologs of Inx5 and Inx7 exist in each species but were underrepresented in the respective datasets due to conditional expression. For these six innexin proteins, no evidence of splice or other variants was found in the mixed nervous system region transcriptome (Shruti et al. 2014). In the current study, the sequences of those innexin proteins were used to probe a CG-specific assembly (Christie et al. 2018a) to assess the putative innexin complement in this portion of the *H. americanus* nervous system. As the data that follow will show, all of the previously described innexins appear to be expressed in the *H. americanus* CG except Homam-Inx4. Transcripts encoding seven previously undescribed innexins/innexin-like proteins (Homam-Inx8-14) were also identified from the CG transcriptome; four of these appear to have multiple splice variants, i.e., Homam-Inx8-11. Taken collectively, these data suggest that the innexin complement of the *H. americanus* nervous system generally, and in the CG specifically, is larger than initially thought. This expansion in innexin complement predicts expanded gap junction diversity (and potentially function) in the lobster nervous system, and more specifically in the cardiac neural network, a hypothesis that can now be tested at the molecular level using the new sequence data as a foundation.

Materials and methods

Transcriptome mining

Searches of the *H. americanus* CG-specific transcriptome were conducted using a protocol employed previously for the identification of a wide variety of protein-encoding transcripts in *H. americanus* and other decapod species (e.g., Christie et al. 2015, 2017, 2018a, b, c; Dickinson et al. 2019). Specifically, the database of the online program tblastn (http://blast.ncbi.nlm.nih.gov/Blast.cgi) was set to Transcriptome Shotgun Assembly (TSA) and restricted to data from the *H. americanus* CG-specific transcriptome (BioProject No. PRJNA412549; Christie et al. 2018a). Previously identified *H. americanus* innexin proteins (Shruti et al. 2014) were used as query sequences for all BLAST searches.

For partial *H. americanus* CG innexin proteins (see below), tblastn searches of other *H. americanus* datasets were conducted using the partial proteins as queries to identify full-length or extended amino acid coverage. The datasets used for these searches were an eyestalk ganglia-specific assembly (BioProject No. PRJNA338672; Christie et al. 2017), a brain-specific transcriptome (BioProject No. PRJNA379629; Christie et al. 2018b), and the transcriptome generated from multiple portions of the *H. americanus* nervous system (BioProject No. PRJNA300643; Northcutt et al. 2016) from which Homam-Inx1-4 and Homam-Inx6-7 were originally identified (Shruti et al. 2014).

Protein prediction and annotation

A workflow developed to provide provisional annotation for a diverse set of proteins from *H. americanus* and other decapod species was used to increase confidence in the annotations ascribed to the novel putative *H. americanus* innexins (e.g., Christie et al. 2015, 2018a, b, c; Dickinson et al. 2019). In brief, nucleotide sequences were translated using the Translate tool of ExPASy (http://web.expasy.org/translate/); the deduced proteins were then used to query the annotated fruit fly, *Drosophila melanogaster*, proteins in FlyBase (version FB2019_01; Thurmond et al. 2019) and the non-redundant arthropod proteins curated at NCBI (taxid:6656) for the most similar sequence in each dataset using the BLAST algorithm blastp (default settings used). In addition, protein structural motifs were analyzed for each *H. americanus* sequence using the online program SMART (http://smart.embl-heidelberg.de/; Letunic and Bork 2018). This workflow was conducted on or before August 30, 2019.

All protein alignments were done using the online program MAFFT version 7 (http://mafft.cbrc.jp/alignment/software/; Katoh and Standley 2013). Calculation of amino acid identity/similarity was done using MAFFT alignments. Specifically, percent identity was calculated as the number of identical amino acids divided by the total number of residues in the longest sequence (×100), while amino acid similarity was calculated as the number of identical and similar amino acids divided by the total number of residues in the longest sequence (×100).

Assessment of phylogenetic relationships among innexin proteins

The phylogenetic relationships of the putative *H. americanus* innexin/innexin-like proteins, both previously reported (Shruti et al. 2014) and described here for the first time, putative innexins from the crab, *Cancer borealis* (Shruti et al. 2014), and defined innexins from two model arthropod species (*D. melanogaster* and the beetle, *Tribolium castaneum*) were inferred from a multiple sequence alignment constructed using default MUSCLE (Edgar 2004) settings in Geneious v10.1.3 (Biomatters Ltd., Auckland, New Zealand; Kearse et al. 2012). Evolutionary analyses were conducted in MEGA X (Kumar et al. 2018) using the maximum likelihood method based on the Le and Gascuel model (2008). Initial trees for the heuristic search were obtained automatically by applying Neighbor-Joining and BioNJ algorithms to a matrix of pairwise distances estimated using a Jones, Taylor, and Thornton (JTT) model (Jones et al. 1992), 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 = 1.3881]). The analysis involved 33 amino acid sequences; all positions with less than 95% site coverage were eliminated such that fewer than 5% of alignment gaps, missing data, and ambiguous bases were allowed at any position. The final dataset consisted of a total of 235 positions. Phylogenetic inferences made using neighbor-joining (Saitou and Nei 1987) and minimum evolution (Rzhetsky and Nei 1992) approaches generated trees with similar topologies. The tree shown reflects the topology with the highest log likelihood (-10924.11). Accession numbers for sequences used in the phylogenetic analyses are provided in Supplemental Table 1.

Results

Identification of putative innexin-encoding transcripts and proteins in the cardiac ganglion of *Homarus americanus*

Using previously identified *H. americanus* innexins (i.e., Homam-Inx1-4 and Homam-Inx6-7; Shruti et al. 2014) as

input sequences, a CG-specific transcriptome was searched using the BLAST algorithm tblastn for transcripts encoding putative homologs (Table 1). Except for Homam-Inx4, sequences encoding each of the previously reported lobster innexins were found in the CG assembly (Table 1), indicating that they are expressed in this portion of the *H. americanus* nervous system. For Homam-Inx1-3 and Homam-Inx7, full-length proteins, identical in amino acid sequence to those reported by Shruti et al. (2014), were deduced from the CG transcripts (Table 1 and Fig. 1). For Homam-Inx6, transcripts encoding two 263 amino acid N-terminal partial proteins were deduced from the CG assembly (Table 1). The two partial CG Homam-Inx6 isoforms differ from one another at three substituted residues, and from the internal fragment of Homam-Inx6 previously reported by Shruti et al. (2014) at two substituted positions, which are different from those that differentiate the two CG isoforms; these variants are hypothesized to be the products of individual-specific variation in a single gene. For ease of future discussion, the two Inx6s identified from the CG assembly have been named Homam-Inx6a and Homam-Inx6b, while that reported by Shruti et al. (2014) is renamed Homam-Inx6c.

Interestingly, the transcripts encoding Homam-Inx1-3 and Homam-Inx6-7 were not the only putative innexin-encoding sequences found in the *H. americanus* CG-specific transcriptome. Specifically, transcripts encoding seven novel putative innexins (Homam-Inx8-14) were identified from the CG assembly, half having multiple putative splice variants (Table 1). For Homam-Inx8, transcripts encoding six full-length variants were identified from the CG transcriptome (Table 1 and Figs. 1, 2), while for Homam-Inx11, transcripts encoding three full-length variants were found (Table 1 and Figs. 1, 2, 3b). For Homam-Inx12 and Homam-Inx14, transcripts encoding single full-length proteins were identified from the CG dataset (Table 1 and Fig. 1). For Homam-Inx9-10 and Homam-Inx13, transcripts encoding partial proteins were identified from the CG transcriptome (Table 1), with two variants each for Homam-Inx9 and Homam-Inx10.

Table 1 Putative *Homarus americanus* innexin-encoding transcripts and deduced proteins identified using a cardiac ganglion-specific transcriptome

Transcript accession no.	Deduced protein		
	Name	Length	Type
GGPK01144812	Innexin 1	379	F
GGPK01125613	Innexin 2	362	F
GGPK01052903	Innexin 3	367	F
GGPK01118011	Innexin 6a	263	N
GGPK01118012	Innexin 6b	263	N
GGPK01032571	Innexin 7	387	F
GGPK01118861	Innexin 8 variant 1	405	F
GGPK01118860	Innexin 8 variant 2	399	F
GGPK01118864	Innexin 8 variant 3	387	F
GGPK01118859	Innexin 8 variant 4	441	F
GGPK01118862	Innexin 8 variant 5	435	F
GGPK01118863	Innexin 8 variant 6	423	F
GGPK01049379	Innexin 9 variant 1	113	C
GGPK01003307	Innexin 9 variant 2	308	N
GGPK01040498	Innexin 10 variant 1	272	C
GGPK01040497	Innexin 10 variant 2	272	C
GGPK01090090	Innexin 11 variant 1	441	F
GGPK01090093	Innexin 11 variant 2	434	F
GGPK01090091	Innexin 11 variant 3	434	F
GGPK01169339	Innexin 12	413	F
GGPK01169338	Innexin 12	413	F
GGPK01160047	Innexin 13	144	I
GGPK01085081	Innexin 14	394	F

Deduced protein type: F, full-length; N, amino-terminal partial; I, internal fragment; C, carboxyl-terminal partial

Filling in sequence gaps using other lobster datasets

As alluded to in the previous section, only partial amino acid sequences for Homam-Inx6, Homam-Inx9-10, and Homam-Inx13 were obtained from the CG-specific transcriptome. To fill in the missing sequence data for these proteins, the partial sequences deduced from the CG assembly were used to search other *H. americanus* transcriptomes for transcripts that could extend amino acid coverage for them. The datasets searched were an eyestalk ganglia-specific assembly (Christie et al. 2017), a brain-specific transcriptome (Christie et al. 2018b), and the mixed nervous system assembly (Northcutt et al. 2016) from which Homam-1-4 and Homam-6-7 were originally identified (Shruti et al. 2014).

A brain transcript (Accession No. GFUC01007852) provided the missing C-terminal sequence of Homam-Inx6a, which when joined with the previous sequence read yielded full sequence coverage for this protein (Fig. 1). Similarly, via a mixed nervous system region transcript (Accession No. GEBG01001573), the missing N-terminal portion of Homam-Inx9-v1 was obtained (Fig. 1). For both Homam-Inx10-v1 and v2, missing N-terminal sequence information was provided by proteins deduced from brain transcripts GFUC01007851 and GFUC01007850, respectively (Figs. 1 and 3a). For Homam-Inx13, the eyestalk ganglia transcript GFDA01118394 provided 107 amino acids of additional N-terminal sequence coverage, converting the internal protein fragment identified from the CG assembly into a significantly longer N-terminal partial protein (Supplemental Fig. 1). Searches yielded no additional sequence information for Homam-Inx6b-c or Homam-Inx9-v2. The longest sequence for each of the putative innexins/innexin variants identified here from the CG, both previously identified

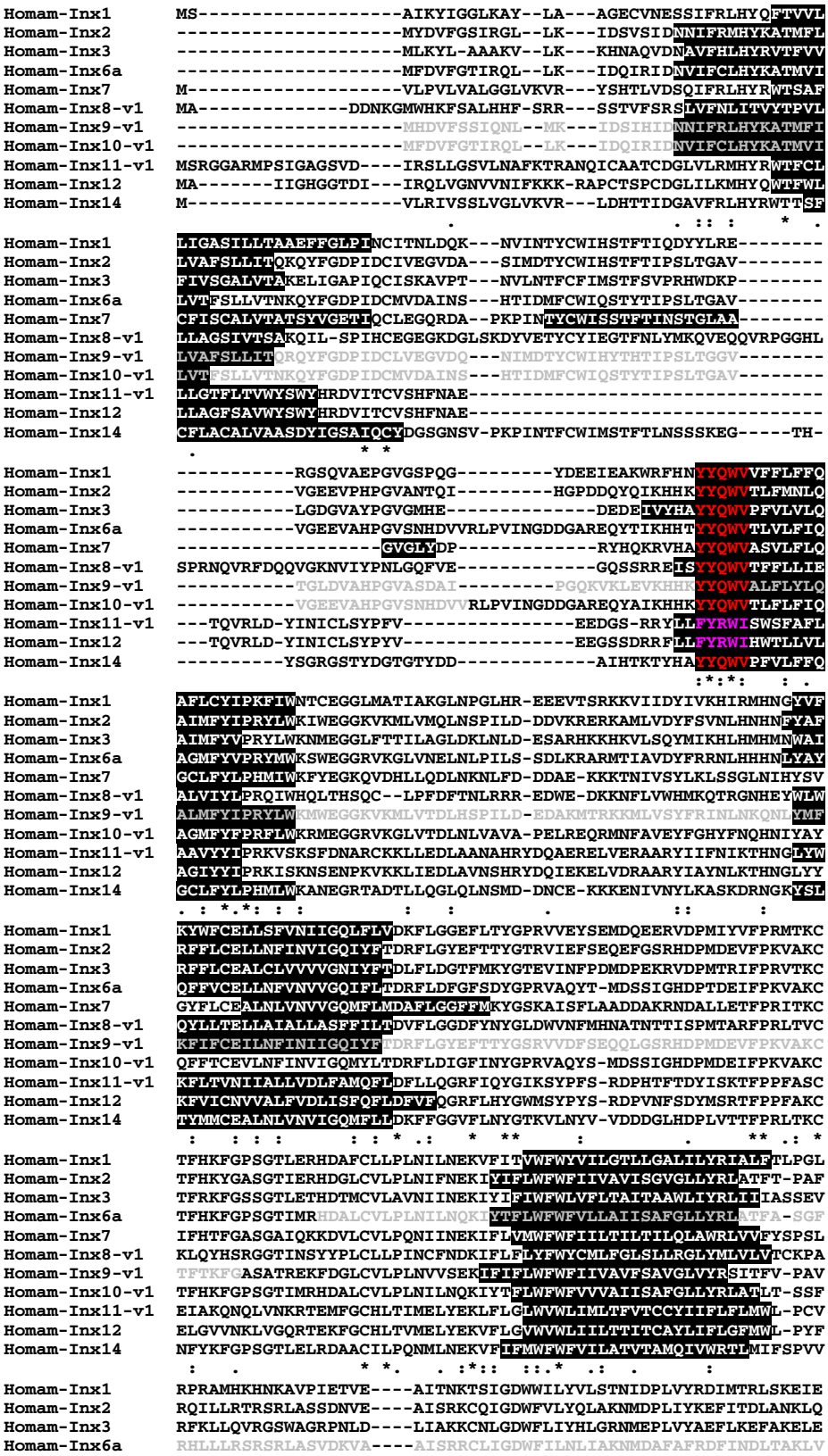

Fig. 1 MAFFT alignment of selected full-length putative *Homarus americanus* innexin (Homam-Inx) proteins. In the line immediately below each sequence grouping, the symbol "*" indicates amino acids that are identical in all proteins, while "." and ":" denote amino acids that are similar in structure among all sequences. Membrane-spanning domains identified by the online program SMART are highlighted in black. Innexin signature motifs (–YYQWV–) are shown in red font, while variants of this motif (–FYRWI–) are shown in pink font. Amino acids in innexin proteins shown in gray were identified from sources other than the cardiac ganglion (see "Filling in sequence gaps using other lobster datasets" section) (color figure online)

Fig. 1 (continued)

```
Homam-Inx7      RIRLLEHRALMNFSPRTE----HAVRRMHLGDYCLLDGIGRNLGTLNFKAVLQGYTEASE
Homam-Inx8-v1   RRLRLKFSAKLVPEDTLD----RFINAHNLSDWFVLCNLAPTMDPVLIAELVTQLVYEVQ
Homam-Inx9-v1   RRILLRVRSRLSTGDTVE----YIADKCAIGDWFFLYQLAKNMDPLIYKEFINELAYDLQ
Homam-Inx10-v1  RHLLLRGRSRLASAEKVE----SISRRCQIGDWFILYLLAKNMDPFVYKDFVNDLSEKLG
Homam-Inx11-v1  RVYLLRVAKPVHASDKVRGVVHAVTQNCKIGDIYLLYRLKGHLSHARFYELMVRLSDPNL
Homam-Inx12     RLMMLRVAKPLNAKDTVSNTIVSVVNCCKIGDVYLLYRLKQHLSHARFYELLTRLSDPEL
Homam-Inx14     RFRMLERRGKLMSLPKLE----QALRQLHLGDFFLLDILGCNLDASTFKDILLKATDCDD
                *   :.                        :.*  .:  :   :       .:
Homam-Inx1      TANSNSPYNSAGLYSSSSV-----------------------------------------
Homam-Inx2      GKGPV-------------------------------------------------------
Homam-Inx3      NSVSTLERKPMLVGS---------------------------------------------
Homam-Inx6a     EKEHLS------------------------------------------------------
Homam-Inx7      EIDDVPSAPSLGAYRTFTTPNEPDTLPRKRP-----------------------------
Homam-Inx8-v1   GGESSDSKPS-------------------RL-----------------------------
Homam-Inx9-v1   GKSA----------------------------------------------------------
Homam-Inx10-v1  DKEHLS--------------------------------------------------------
Homam-Inx11-v1  CNKQIQQSGP------GVMPESKDKVPPNKQADTLRQRRPNMPQDPPVNPEYLHQLLAGN
Homam-Inx12     IKTMLE--------------DPADRAVHARNQDNMRNRKPNMT--------------IGG
Homam-Inx14     VIANNNS------YRPFYEPGDDDPVAYKRQ-----------------------------

Homam-Inx1      ------------------------------
Homam-Inx2      ------------------------------
Homam-Inx3      ------------------------------
Homam-Inx6a     ------------------------------
Homam-Inx7      -----------------LYQDTTL------
Homam-Inx8-v1   -----------------GKQEKSLQSVNYI
Homam-Inx9-v1   ------------------------------
Homam-Inx10-v1  ------------------------------
Homam-Inx11-v1  PEMMGRHPQHPDQRTPLLKTNTSILIE---
Homam-Inx12     PKGKFNNPNDLFISQDYGRPNTSILVE---
Homam-Inx14     -----------------LAAEDATAV----
```

(Shruti et al. 2014) and novel, is provided in Supplemental Fig. 1.

Annotation vetting of novel and formerly partial *Homarus americanus* innexin proteins

As one means of increasing confidence in the innexin annotations ascribed to the novel putative family members identified here, i.e., Homam-Inx8-14, as well as Homam-Inx6, which was formerly an N-terminal partial protein (Shruti et al. 2014), each sequence was used to search the annotated *D. melanogaster* proteins in FlyBase and the non-redundant arthropod proteins in NCBI for the most similar sequences. The expectations for these searches were that the top hit for each *H. americanus* sequence in each dataset would be a protein annotated as an innexin family member. As Tables 2 and 3 show, this prediction was borne out for all searches. For example, the top FlyBase hit for each of the six putative Homam-Inx8, and the three putative Homam-Inx11 isoforms was *D. melanogaster* innexin 2 (Table 2). Similarly, the top NCBI non-redundant arthropod protein hit for each of the six putative *H. americanus* Inx8s was a protein annotated as either innexin 1 or innexin 2 from the amphipod *Hyalella azteca*, with a shrimp, *Penaeus vannamei,* protein annotated as innexin 2 being the top hit for each of the three *Homarus* Inx11s (Table 3).

As a second means of vetting the annotations of Homam-Inx6 and Homam-Inx8-14, the sequence of each protein/isoform was scanned for structural/functional domains considered characteristic of innexins. Previous work has shown that members of the innexin family, including those of the lobster, are typified by the presence of four membrane-spanning domains and the signature innexin motif –YYQWV– in the second membrane-spanning region (e.g., Shruti et al. 2014). In addition, a number of highly conserved residues, or close approximations thereof, are present in other portions of the proteins, i.e., –S/TX_{17}GX_4C$X_{13\text{-}19}$C$X_{25\text{-}92}$YXWX_{14}PX_3W– (where X_X represents one or more variable residues) spanning transmembrane domains 1 and 2 and the first extracellular loop and –FX_4C$X_{16\text{-}24}$CX_4NX_4KXY/FX_3Y/F/W–, present between the second extracellular loop and the fourth membrane-spanning domain (e.g., Shruti et al. 2014). For Homam-Inx6, Homam-Inx8, Homam-Inx9, and Homam-Inx14, these features appear largely present, at least in the full-length proteins/isoforms identified here (Fig. 1 and Supplemental Fig. 1). Homam-Inx13 also appears to possess the features of a stereotypical innexin, at least over the portion of the protein that has been identified (Supplemental Fig. 1). More interesting are the Homam-Inx10-12 proteins/variants. Both isoforms of Homam-Inx10 appear to be missing the third transmembrane domain, at least as predicted using the online program SMART (Figs. 1 and 3a). Interestingly, the portion of Homam-Inx10 that corresponds to the third membrane-spanning region in the other innexins, and which is identically conserved in Homam-Inx10-v1 and v2 (Fig. 3a), is similar in amino acid sequence (70% identity/100% similarity) to that region of the other innexins, i.e., –IYAYQFFTCEVLNFINVIGQMYL– in the Homam-Inx10s versus –LYAYQFFVCELLNFVNVVGQIFL– in the Homam-Inx6 variants. However, the sequence

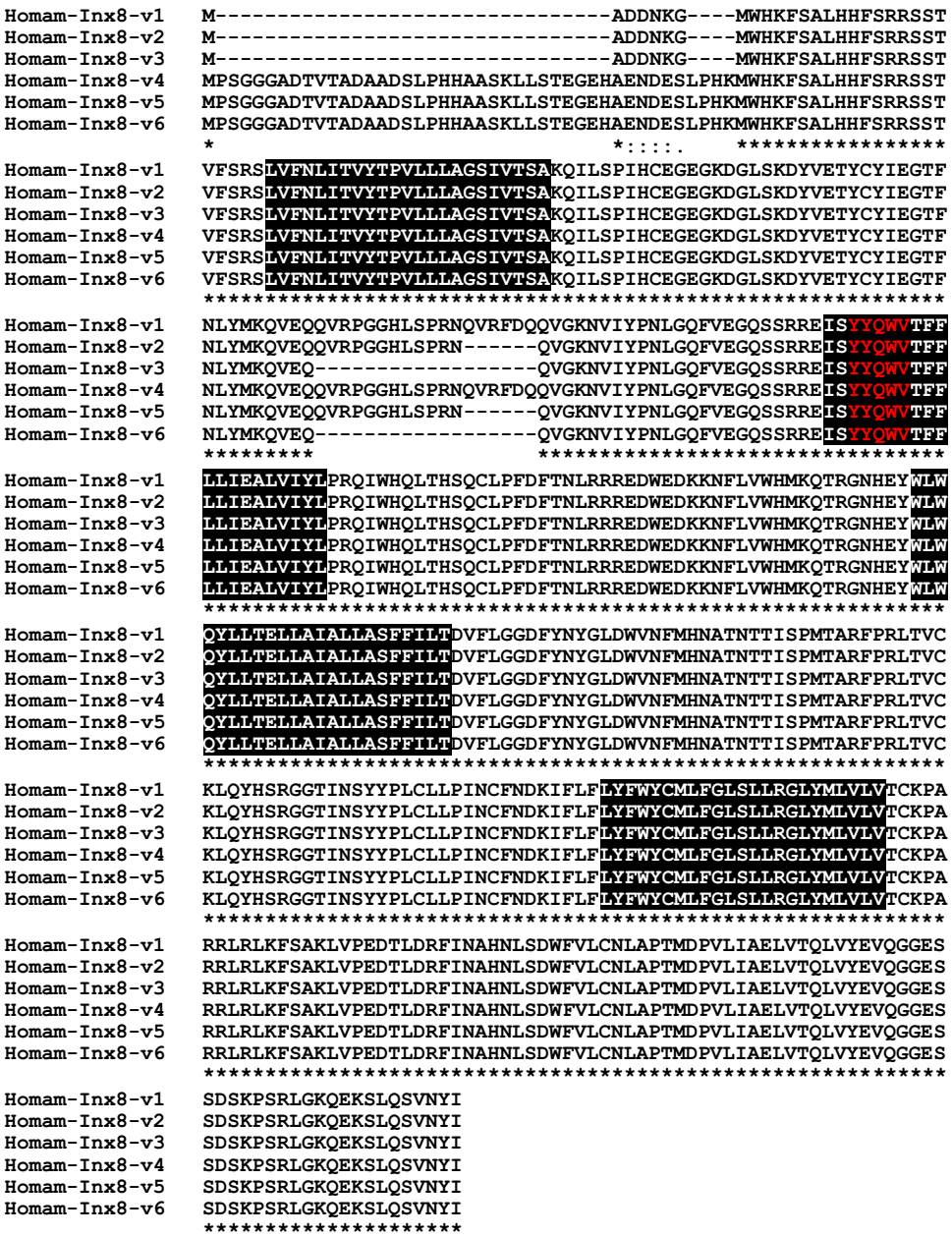

Fig. 2 MAFFT alignment of putative *Homarus americanus* innexin 8 (Homam-Inx8) variants. In the line immediately below each sequence grouping, the symbol "*" indicates amino acids that are identical in all six variants, while "." and ":" denote amino acids that are similar in structure among all six sequences. Membrane-spanning domains identified by the online program SMART are highlighted in black. Innexin signature motifs (–YYQWV–) are shown in red font (color figure online)

in Homam-Inx6 is predicted to be a transmembrane domain by SMART (Fig. 1). This suggests that the two Homam-Inx10s do possess a third membrane-spanning region, but that this feature simply did not reach the threshold value to be identified as such by SMART. In the three Homam-Inx11 isoforms and in Homam-Inx12 (Figs. 1 and 3b), the stereotypical –YYQWV– innexin motif present in the second membrane-spanning region is replaced with the variant sequence –FYRWI–. While three of the five residues vary between the two sequences, all are highly conserved substitutions (Fig. 1). Thus, taken collectively, the structural domain and reciprocal BLAST results obtained for Homam-Inx6 and Homam-Inx8-14 support the provisional innexin family annotations ascribed to them here, although functional analyses are needed to confirm this hypothesis for these proteins, as they are for the previously reported (Shruti et al. 2014) Homam-Inx1-4 and Homam-7.

Phylogenetic relationships among *Homarus americanus* innexins

Maximum likelihood-based phylogenetic inferences of the putative lobster innexin sequences identified here with protein sequences from crab, *C. borealis,* and two model arthropods (*D. melanogaster* and *T. castaneum*) revealed six reasonably supported (bootstrap values ranging from 27 to 99)

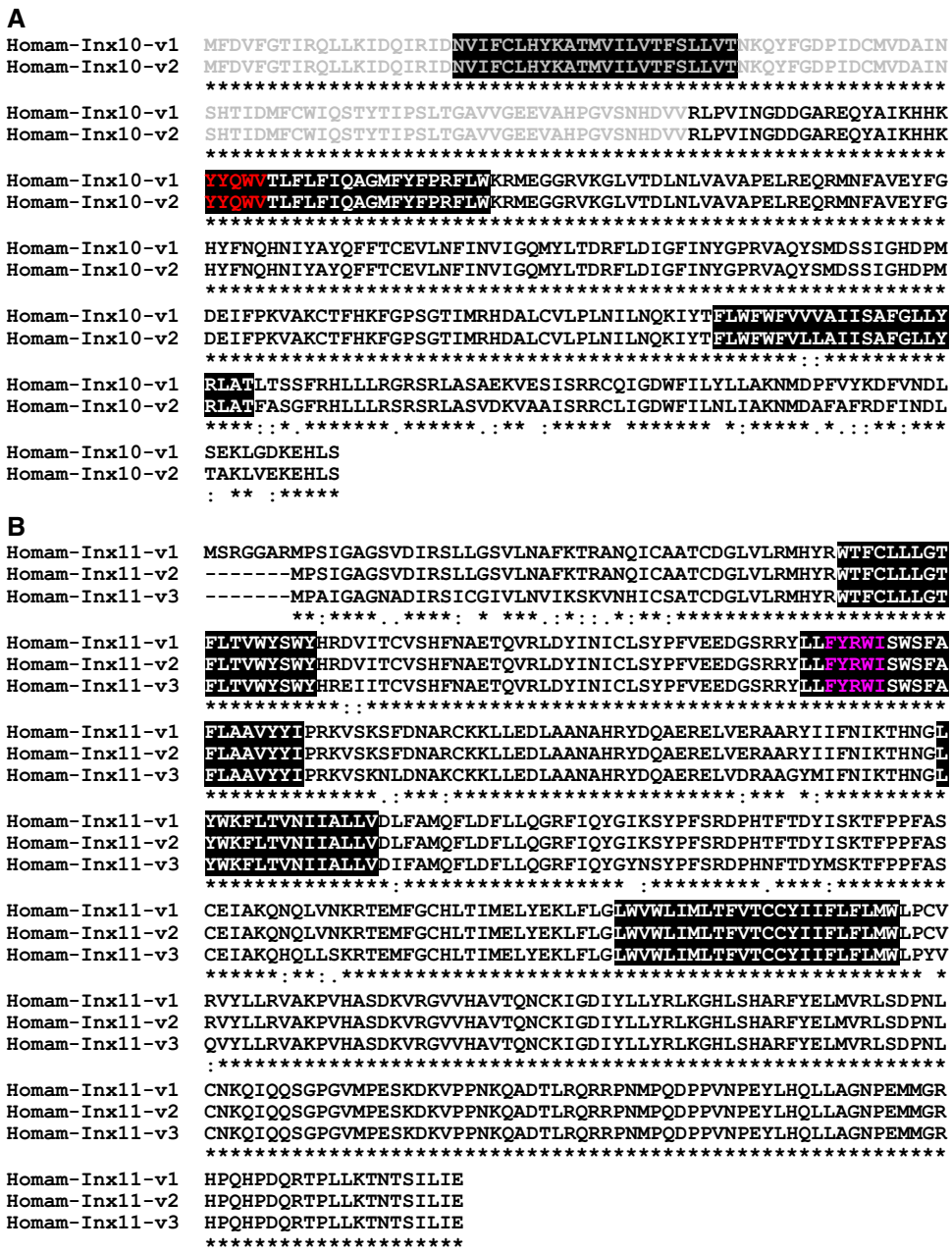

Fig. 3 MAFFT alignment of putative *Homarus americanus* innexin (Homam-Inx) 10 and 11 variants. **a** Alignment of Homam-Inx10 variants 1 and 2, **b** alignment of Homam-Inx11 variants 1–3. In the line immediately below each sequence grouping, the symbol "*" indicates amino acids that are identical between/among proteins, while "." and ":" denote amino acids that are similar in structure among all sequences in a given alignment. Membrane-spanning domains identified by the online program SMART are highlighted in black. In **a**, innexin signature motifs (–YYQWV–) are shown in red font, while variants of this motif (–FYRWI–) in **b** are shown in pink font. In **a**, amino acids shown in gray were identified from sources other than the cardiac ganglion (see "Filling in sequence gaps using other lobster datasets" section) (color figure online)

innexin-specific clades (labeled A-F in Fig. 4) rather than class-specific clades, underscoring the evolutionary conservation of at least a subset of innexins. Three apparent gene expansion events (Homam-Inx7/13/14, Homam-Inx2/6/9/10, and Homam-Inx11/12) may account for the diversity of *H. americanus* innexins relative to the insect models.

Discussion

In the study presented here, putative innexin complement in the CG of *H. americanus* was assessed using a transcriptome specific for this portion of the lobster nervous system. Employing known *H. americanus* innexins as queries, the products of 12 putative innexin genes were identified in the CG assembly. These included transcripts and corresponding proteins for five of the six known *H. americanus* innexins (i.e., Homam-Inx1-3 and Homam-Inx6-7, but not Homam-Inx4), and those derived from seven novel putative innexin genes (i.e., Homam-Inx8-14). Reciprocal BLAST and structural domain analyses support the collective set of proteins as being actual members of the innexin family. These data suggest that the complement of innexins present in the lobster nervous system is larger than the six originally identified by Shruti and colleagues (Shruti et al. 2014). For Homam-Inx6 and Homam-Inx8-11, multiple

Table 2 Most similar *Drosophila melanogaster* protein to each of the novel putative *Homarus americanus* innexin sequence deduced from cardiac ganglion transcripts[a]

H. americanus innexin	Top FlyBase hit			
	Accession no.	Name	BLAST statistics	
			Score	*E*-value
Innexin 6a	AAF46229	Innexin 2, isoform A	419	4e−117
Innexin 6b	AAF46229	Innexin 2, isoform A	295	5e−80
Innexin 8 variant 1	AAF46229	Innexin 2, isoform A	168	2e−41
Innexin 8 variant 2	AAF46229	Innexin 2, isoform A	168	1e−41
Innexin 8 variant 3	AAF46229	Innexin 2, isoform A	174	2e−43
Innexin 8 variant 4	AAF46229	Innexin 2, isoform A	168	2e−41
Innexin 8 variant 5	AAF46229	Innexin 2, isoform A	168	2e−41
Innexin 8 variant 6	AAF46229	Innexin 2, isoform A	173	3e−43
Innexin 9 variant 1	AAF46229	Innexin 2, isoform A	451	5e−127
Innexin 9 variant 2	AAF46229	Innexin 2, isoform A	377	8e−105
Innexin 10 variant 1	AAF46229	Innexin 2, isoform A	416	3e−116
Innexin 10 variant 2	AAF46229	Innexin 2, isoform A	400	2e−111
Innexin 11 variant 1	AAF46229	Innexin 2, isoform A	107	5e−23
Innexin 11 variant 2	AAF46229	Innexin 2, isoform A	106	6e−23
Innexin 11 variant 3	AAF46229	Innexin 2, isoform A	104	3e−22
Innexin 12	AAF46229	Innexin 2, isoform A	104	3e−22
Innexin 13	AAF46229	Innexin 2, isoform A	233	3e−61
Innexin 14	AAF56822	Innexin 3, isoform A	280	3e−75

BLAST searches of FlyBase were conducted on or before August 30, 2019

[a]While innexin 6 was identified previously from *H. americanus*, the proteins deduced from cardiac ganglion transcripts are longer, and hence have been included here for revetting

Table 3 Most similar non-redundant arthropod protein to each of the novel putative *Homarus americanus* innexin sequence deduced from cardiac ganglion transcripts[a]

H. americanus innexin	Top NCBI non-redundant arthropod protein hit				
	Accession no.	Species	Protein name	BLAST statistics	
				Score	*E*-value
Innexin 6a	AIJ10714	*Cancer borealis*	Innexin 6	510	3e−180
Innexin 6b	AIT97137	*Homarus americanus*	Innexin 6	415	4e−147
Innexin 8 variant 1	XP_018023776	*Hyalella azteca*	Innexin 1-like isoform X1	744	0.0
Innexin 8 variant 2	XP_018023776	*Hyalella azteca*	Innexin 1-like isoform X1	746	0.0
Innexin 8 variant 3	XP_018023784	*Hyalella azteca*	Innexin 2-like isoform X2	748	0.0
Innexin 8 variant 4	XP_018023776	*Hyalella azteca*	Innexin 1-like isoform X1	748	0.0
Innexin 8 variant 5	XP_018023776	*Hyalella azteca*	Innexin 1-like isoform X1	749	0.0
Innexin 8 variant 6	XP_018023784	*Hyalella azteca*	Innexin 2-like isoform X2	752	0.0
Innexin 9 variant 1	AIT97134	*Homarus americanus*	Innexin 2	570	0.0
Innexin 9 variant 2	AIT97134	*Homarus americanus*	Innexin 2	482	3e−170
Innexin 10 variant 1	XP_027221824	*Penaeus vannamei*	Innexin 2-like	519	0.0
Innexin 10 variant 2	AIJ10714	*Cancer borealis*	Innexin 6	510	2e−180
Innexin 11 variant 1	XP_027232619	*Penaeus vannamei*	Innexin 2-like isoform X1	763	0.0
Innexin 11 variant 2	XP_027232619	*Penaeus vannamei*	Innexin 2-like isoform X1	760	0.0
Innexin 11 variant 3	XP_027232619	*Penaeus vannamei*	Innexin 2-like isoform X1	732	0.0
Innexin 12	XP_027239060	*Penaeus vannamei*	Innexin 2-like	783	0.0
Innexin 13	XP_027207462	*Penaeus vannamei*	Innexin 2-like	451	1e−158
Innexin 14	XP_027207461	*Penaeus vannamei*	Innexin 3-like	543	0.0

BLAST searches of FlyBase were conducted on or before August 30, 2019

[a]While innexin 6 was identified previously from *H. americanus*, the proteins deduced from cardiac ganglion transcripts are longer, and hence have been included here for revetting

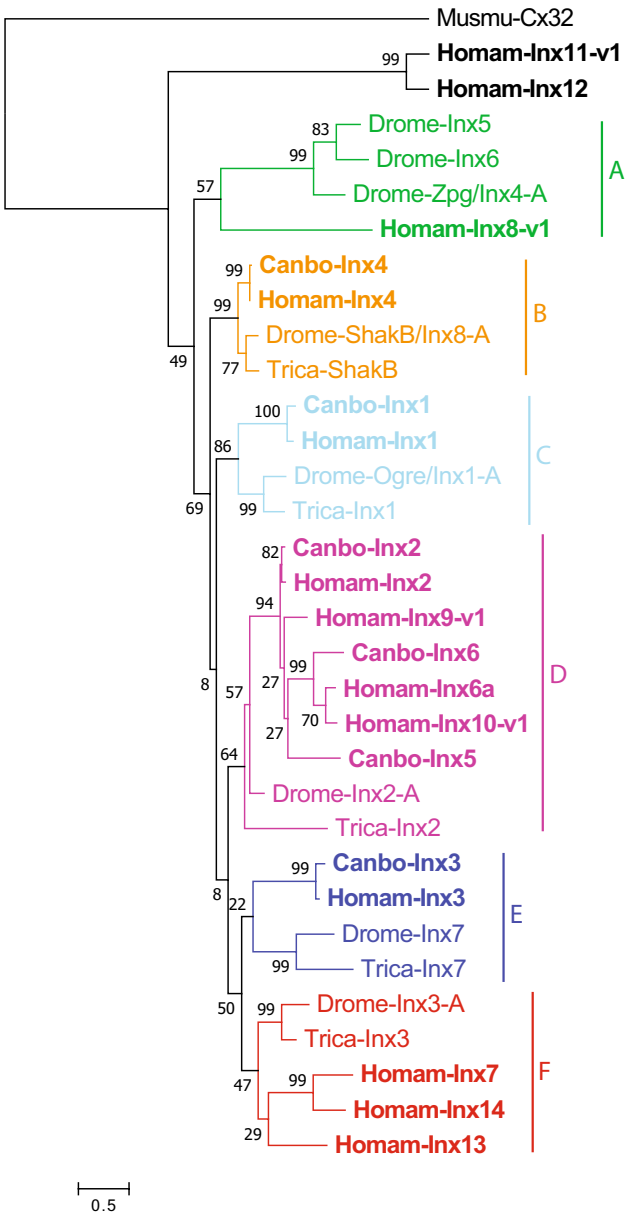

Fig. 4 Phylogenetic relationship of putative *Homarus americanus* innexins with innexin sequences from three other arthropods. Maximum likelihood tree depicting the inferred evolutionary history of the respective innexin sequences. The tree with the highest log likelihood is drawn to scale with branch lengths measured in the number of substitutions per site. The percentage of trees in which the associated taxa clustered together across 1000 replicates is shown next to the branches. Innexin-specific clades have been colored and labeled accordingly. The tree was rooted to the mouse connexin sequence (Musmu-Cx32; Accession No. AAA37496). Decapod sequences are shown in bold. Species abbreviations are: Canbo, *Cancer borealis*; Drome, *Drosophila melanogaster*; Homam, *Homarus americanus*; Musmu, *Mus musculus*; Trica, *Tribolium castaneum*

variants were identified from the CG dataset, further expanding putative innexin diversity in the lobster generally, and in the CG specifically. Since innexin diversity is a requirement for gap junction diversity in any given species or tissue, these findings support the possibility of expanded gap junction diversity and function in the nervous system of *H. americanus*, a hypothesis that can now be tested biochemically, molecularly, and physiologically using the sequence data presented here as a foundation.

Acknowledgements Lisa Baldwin and Dr. Colin Brent are thanked for reading and commenting on earlier versions of this article. This work was supported by funds from the National Science Foundation (IOS-1353023 and IOS-1856307 [to AEC]; IOS-1354567 and IOS-1856433 [to PSD]), 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). 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 USDA; the USDA is an equal opportunity provider and employer.

Compliance with ethical standards

Conflict of interest None.

References

Alcamí P, Pereda AE (2019) Beyond plasticity: the dynamic impact of electrical synapses on neural circuits. Nat Rev Neurosci 20:253–271

Beyer EC, Berthoud VM (2018) Gap junction gene and protein families: connexins, innexins, and pannexins. Biochim Biophys Acta Biomembr 1860:5–8

Calabrese RL, Norris BJ, Wenning A (2016) The neural control of heartbeat in invertebrates. Curr Opin Neurobiol 41:68–77

Christie AE, Chi M, Lameyer TJ, Pascual MG, Shea DN, Stanhope ME, Schulz DJ, Dickinson PS (2015) Neuropeptidergic signaling in the American lobster *Homarus americanus*: new insights from high-throughput nucleotide sequencing. PLoS One 10:e0145964

Christie AE, Roncalli V, Cieslak MC, Pascual MG, Yu A, Lameyer TJ, Stanhope ME, Dickinson PS (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 AE, Yu A, Roncalli V, Pascual MG, Cieslak MC, Warner AN, Lameyer TJ, Stanhope ME, Dickinson PS, Hull JJ (2018a) 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 Genom 41:19–30

Christie AE, Yu A, Pascual MG, Roncalli V, Cieslak MC, Warner AN, Lameyer TJ, Stanhope ME, Dickinson PS, Hull JJ (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 Genom 40:25–44

Christie AE, Stanhope ME, Gandler HI, Lameyer TJ, Pascual MG, Shea DN, Yu A, Dickinson PS, Hull JJ (2018c) Molecular characterization of putative neuropeptide, amine, diffusible gas and small molecule transmitter biosynthetic enzymes in the eyestalk ganglia of the American lobster, *Homarus americanus*. Invertebr Neurosci 18:12

Cooke IM (2002) Reliable, responsive pacemaking and pattern generation with minimal cell numbers: the crustacean cardiac ganglion. Biol Bull 202:108–136

Dickinson PS, Qu X, Stanhope ME (2016) Neuropeptide modulation of pattern-generating systems in crustaceans: comparative studies and approaches. Curr Opin Neurobiol 41:149–157

Dickinson PS, Hull JJ, Miller A, Oleisky ER, Christie AE (2019) To what extent may peptide receptor gene diversity/complement contribute to functional flexibility in a simple pattern-generating neural network? Comp Biochem Physiol Part D Genom Proteom 30:262–282

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792–1797

Harris AL (2018) Electrical coupling and its channels. J Gen Physiol 150:1606–1639

Hooper SL, DiCaprio RA (2004) Crustacean motor pattern generator networks. Neurosignals 13:50–69

Jones DT, Taylor WR, Thornton JM (1992) The rapid generation of mutation data matrices from protein sequences. Comput Appl Biosci 8:275–282

Kandarian B, Sethi J, Wu A, Baker M, Yazdani N, Kym E, Sanchez A, Edsall L, Gaasterland T, Macagno E (2012) The medicinal leech genome encodes 21 innexin genes: different combinations are expressed by identified central neurons. Dev Genes Evol 222:29–44

Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30:772–780

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 28:1647–1649

Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018) MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol 35:1547–1549

Lane BJ, Kick DR, Wilson DK, Nair SS, Schulz DJ (2018) Dopamine maintains network synchrony via direct modulation of gap junctions in the crustacean cardiac ganglion. Elife 7:e39368

Le SQ, Gascuel O (2008) An improved general amino acid replacement matrix. Mol Biol Evol 25:1307–1320

Letunic I, Bork P (2018) 20 years of the SMART protein domain annotation resource. Nucleic Acids Res 46:D493–D496

Northcutt AJ, Lett KM, Garcia VB, Diester CM, Lane BJ, Marder E, Schulz DJ (2016) Deep sequencing of transcriptomes from the nervous systems of two decapod crustaceans to characterize genes important for neural circuit function and modulation. BMC Genom 17:868

O'Brien J (2019) Design principles of electrical synaptic plasticity. Neurosci Lett 695:4–11

Otopalik AG, Lane B, Schulz DJ, Marder E (2019) Innexin expression in electrically coupled motor circuits. Neurosci Lett 695:19–24

Pereda AE, Macagno E (2017) Electrical transmission: two structures, same functions? Dev Neurobiol 77:517–521

Pereda AE, Curti S, Hoge G, Cachope R, Flores CE, Rash JE (2013) Gap junction-mediated electrical transmission: regulatory mechanisms and plasticity. Biochim Biophys Acta 1828:134–146

Phelan P (2005) Innexins: members of an evolutionarily conserved family of gap-junction proteins. Biochim Biophys Acta 1711:225–245

Phelan P, Goulding LA, Tam JL, Allen MJ, Dawber RJ, Davies JA, Bacon JP (2008) Molecular mechanism of rectification at identified electrical synapses in the *Drosophila* giant fiber system. Curr Biol 18:1955–1960

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

Schulz DJ, Lane BJ (2017) Homeostatic plasticity of excitability in crustacean central pattern generator networks. Curr Opin Neurobiol 43:7–14

Shruti S, Schulz DJ, Lett KM, Marder E (2014) Electrical coupling and innexin expression in the stomatogastric ganglion of the crab *Cancer borealis*. J Neurophysiol 112:2946–2958

Starich T, Sheehan M, Jadrich J, Shaw J (2001) Innexins in *C. elegans*. Cell Commun Adhes 8:311–314

Stebbings LA, Todman MG, Phillips R, Greer CE, Tam J, Phelan P, Jacobs K, Bacon JP, Davies JA (2002) Gap junctions in *Drosophila*: developmental expression of the entire innexin gene family. Mech Dev 113:197–205

Thurmond J, Goodman JL, Strelets VB, Attrill H, Gramates LS, Marygold SJ, Matthews BB, Millburn G, Antonazzo G, Trovisco V, Kaufman TC, Calvi BR, FlyBase Consortium (2019) FlyBase 2.0: the next generation. Nucleic Acids Res 47:D759–D765

Willecke K, Eiberger J, Degen J, Eckardt D, Romualdi A, Güldenagel M, Deutsch U, Söhl G (2002) Structural and functional diversity of connexin genes in the mouse and human genome. Biol Chem 383:725–737