7-1-2018

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

Andrew E. Christie
Pacific Biosciences Research Center

Andy Yu
Pacific Biosciences Research Center

Micah G. Pascual
Pacific Biosciences Research Center

Vittoria Roncalli
Pacific Biosciences Research Center

Matthew C. Cieslak
Pacific Biosciences Research Center

Marine Genomics 40 (2018) 25–44

Contents lists available at ScienceDirect

Marine Genomics

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

Method paper

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

Andrew E. Christie[a,*], Andy Yu[a], Micah G. Pascual[a], Vittoria Roncalli[a], Matthew C. Cieslak[a], Amanda N. Warner[b], Tess J. Lameyer[c], Meredith E. Stanhope[c], Patsy S. Dickinson[c], J. Joe Hull[b]

[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] Pest Management and Biocontrol Research Unit, US Arid Land Agricultural Research Center, USDA Agricultural Research Services, Maricopa, AZ 85138, USA
[c] Department of Biology, Bowdoin College, 6500 College Station, Brunswick, ME 04672, USA

ARTICLE INFO

Keywords:
Biological rhythm
Circadian rhythm
Transcriptomics
Crustacea
Decapoda

ABSTRACT

Essentially all organisms exhibit recurring patterns of physiology/behavior that oscillate with a period of ~24-h and are synchronized to the solar day. Crustaceans are no exception, with robust circadian rhythms having been documented in many members of this arthropod subphylum. However, little is known about the molecular underpinnings of their circadian rhythmicity. Moreover, the location of the crustacean central clock has not been firmly established, although both the brain and eyestalk ganglia have been hypothesized as loci. The American lobster, *Homarus americanus*, is known to exhibit multiple circadian rhythms, and immunodetection data suggest that its central clock is located within the eyestalk ganglia rather than in the brain. Here, brain- and eyestalk ganglia-specific transcriptomes were generated and used to assess the presence/absence of transcripts encoding the commonly recognized protein components of arthropod circadian signaling systems in these two regions of the lobster central nervous system. Transcripts encoding putative homologs of the core clock proteins clock, cryptochrome 2, cycle, period and timeless were found in both the brain and eyestalk ganglia assemblies, as were transcripts encoding similar complements of putative clock-associated, clock input pathway and clock output pathway proteins. The presence and identity of transcripts encoding core clock proteins in both regions were confirmed using PCR. These findings suggest that both the brain and eyestalk ganglia possess all of the molecular components needed for the establishment of a circadian signaling system. Whether the brain and eyestalk clocks are independent of one another or represent a single timekeeping system remains to be determined. Interestingly, while most of the proteins deduced from the identified transcripts are shared by both the brain and eyestalk ganglia, assembly-specific isoforms were also identified, *e.g.*, several period variants, suggesting the possibility of region-specific variation in clock function, especially if the brain and eyestalk clocks represent independent oscillators.

1. Introduction

Coordination of physiology and behavior to recurring changes in the environment is a requirement for all living organisms. In many cases, this coordination is achieved via the action of intrinsic, genetically-encoded timekeeping systems, so called “biological clocks”, which operate on a wide range of time scales, from sub-second to seasonal and even longer (Golombek et al., 2014). One of the best known clock systems is the circadian pacemaker, which is responsible for the timing of recurring patterns of physiology and behavior that oscillate with a period of ~24-h and are synchronized to the solar day. The molecular cascade responsible for the establishment of circadian rhythmicity has been well characterized for several species, for example the fruit fly, *Drosophila melanogaster* (Allada and Chung, 2010; Hardin, 2011; Mendoza-Viveros et al., 2017; Ozkaya and Rosato, 2012; Yoshii et al., 2015). In all species, circadian pacemakers involve interacting feedback loops of transcriptional activation and repression, as well as modulation of the feedback loops via processes such as phosphorylation and degradation of key protein components (see the abovementioned references for a complete description of the *D. melanogaster* cascade).

* Corresponding author at: 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.
E-mail address: crabman@pbrc.hawaii.edu (A.E. Christie).

https://doi.org/10.1016/j.margen.2018.03.002
Received 23 January 2018; Received in revised form 5 March 2018; Accepted 6 March 2018
Available online 12 April 2018
1874-7787/

One group of animals for which numerous circadian rhythms of physiology and behavior have been documented is the crustaceans. As reviewed by Strauss and Dircksen (2010), known crustacean physiological/behavioral systems that are under circadian control include, but are not limited to, feeding, locomotion, molting, pigment dispersion and reproduction. Despite the diversity of documented circadian rhythms in crustaceans, little is known about the molecular underpinnings of circadian signaling in these animals (Chen et al., 2017; Christie et al., 2013, Christie et al., 2018; Nesbit and Christie, 2014; Mazzotta et al., 2010; O'Grady et al., 2016; Roncalli et al., 2017; Sbragaglia et al., 2015; Tilden et al., 2011; Yang et al., 2006; Zhang et al., 2013). In fact, large collections of putative circadian genes/proteins have been identified and characterized for only a handful of crustacean species (Christie et al., 2013, 2018; Nesbit and Christie, 2014; Roncalli et al., 2017; Tilden et al., 2011; O'Grady et al., 2016; Sbragaglia et al., 2015). Moreover, the location of the crustacean central circadian pacemaker remains unclear, although several sites in the central nervous system (CNS) have been proposed as the locus of this clock system (Strauss and Dircksen, 2010). Proposed sites include both the eyestalk ganglia and the supraoesophageal ganglion, commonly referred to as the brain (Strauss and Dircksen, 2010).

One decapod crustacean for which multiple circadian rhythms have been documented is the American lobster, *Homarus americanus*. In this species, which has significance both for its commercial value (Overton, 2017) and as a biomedical model for investigating the basic principles governing the generation, maintenance and modulation of rhythmically active motor behavior (Blitz and Nusbaum, 2011; Hooper and DiCaprio, 2004; Marder and Bucher, 2007; Marder et al., 2017; Nusbaum et al., 2001; Otopalik et al., 2017; Schulz and Lane, 2017; Stein, 2009), circadian rhythms have been shown or are hypothesized to play roles in the control of locomotor activity (Jury et al., 2005), neurogenesis (Goergen et al., 2002), brain serotonin levels (Wildt et al., 2004), and heartbeat frequency (Chabot and Webb, 2008). Western blots of *H. americanus* eyestalk ganglia- and brain-derived protein extracts using an antibody generated against period (PER) protein, a key component of circadian signaling systems (*e.g.*, Allada and Chung, 2010), revealed immunoreactivity in the eyestalk ganglia but not in the brain (Grabek and Chabot, 2012). These data suggested that the central circadian clock of the lobster is located in the eyestalk ganglia rather than in the brain. However, the size of the protein detected in the eyestalk ganglia was significantly smaller than most other arthropod PERs, *i.e.*, ~70 *vs.* ~100^{+} kDa (Grabek and Chabot, 2012), raising the question of whether or not the immunoreactive protein was a true member of the PER family.

In the study presented here, a transcriptomics approach was used to assess the presence/absence of circadian signaling system transcripts and, by proxy, proteins, including PER, in the *H. americanus* brain and eyestalk ganglia. Specifically, region-specific transcriptomes were generated and searched for sequences encoding putative homologs of known arthropod circadian proteins. Transcripts encoding putative homologs of all of the circadian proteins searched for, including PER, were identified in both the brain and eyestalk ganglia assemblies, with a number of the identified sequences confirmed using PCR and traditional Sanger sequencing. These findings suggest that both the brain and eyestalk ganglia are likely to possess intrinsic circadian signaling systems. Interestingly, while most of the proteins deduced from the identified transcripts are shared by both the brain and eyestalk ganglia, there is also evidence for assembly-specific isoforms of some proteins, including PER, suggesting the possibility of region-specific functional variation in the brain and eyestalk clocks, especially if they represent distinct timekeeping systems rather than a single distributed one. Collectively, the data presented here provide a foundation for future investigations of circadian signaling in *H. americanus*, including studies designed to determine the diel cycling patterns of clock gene expression in the brain and eyestalk ganglia, as well as studies examining the location and identity of clock neurons in the lobster brain and eyestalk ganglia and whether these cells represent two distinct or one distributed timekeeping system.

2. Materials and methods

2.1. De novo transcriptome assembly

Two transcriptomes, one for the eyestalk ganglia and the other for the brain, were generated for use in determining the presence/absence of circadian signaling systems in these two regions of the lobster CNS. The production of the eyestalk ganglia-specific transcriptome is described in an earlier study (Christie et al., 2017); production of the brain-specific assembly is described below. Tissues from the same set of lobsters were used to produce both transcriptomes, with all dissection, RNA isolation, cDNA library production and Illumina sequencing conducted simultaneously and using identical methods.

2.1.1. Animals and tissue dissection

American lobsters, *H. americanus*, (n = 4) were purchased commercially from seafood retailers in Brunswick, Maine. Lobsters were maintained in recirculating natural seawater aquaria at 10–12 °C and were fed a diet of chopped shrimp approximately weekly. For the isolation of the brain, animals were anesthetized by packing in ice for approximately 30 min, after which the anterior portion of the thorax and its underlying tissue were isolated. The brain was dissected from this preparation in chilled (~4 °C) physiological saline (composition in mM/l: 479.12 NaCl, 12.74 KCl, 13.67 $CaCl_2$, 20.00 $MgSO_4$, 3.91 Na_2SO_4, 11.45 Trizma base, and 4.82 maleic acid [pH = 7.45]).

2.1.2. RNA isolation

Freshly dissected individual brains (n = 4) were placed into sterile RNase-free 1.5 ml microfuge tubes containing 300 µl of TRIzol Reagent (catalog no. 15596018; Thermo Fisher Scientific Inc., Waltham, MA, USA) and manually homogenized using a sterile RNase-free disposable pestle (catalog no. 9950–901; Argos Technologies Inc., Elgin, IL, USA). RNA was isolated from the resulting homogenate using a Direct-zol RNA MiniPrep spin column system (catalog no. R2052; Zymo Research, Irvine, CA, USA) according to the manufacturer-supplied protocol. RNA quality was assessed using an Agilent 2100 Bioanalyzer (Agilent, Santa Clara, CA, USA). All RNA samples were stored at −80 °C until being shipped on dry ice to the Georgia Genomics Facility (University of Georgia, Athens, GA, USA) for library preparation and sequencing.

2.1.3. cDNA library production and Illumina sequencing

Double-stranded cDNA libraries were prepared from total RNA using a KAPA Stranded mRNA-Seq kit (catalog No. KK8420; KAPA Biosystems, Wilmington, MA, USA) following the manufacturer's instructions; 3 µg of total RNA/sample was used for library generation. In brief, total RNA samples were purified with two oligo-dT selection (poly (A) enrichment using oligo-dT beads). Samples were then fragmented and reverse transcribed into double-stranded complementary cDNA using random primers, with second strand synthesis marked using dUPT. Each brain library was tagged with a unique indexed adapter. A Qubit 2.0 Fluorometer (Invitrogen, Carlsbad, CA, USA), an AATI Fragment Analyzer (Advanced Analytical Technologies, Ankeny, IA, USA), and KAPA qPCR assays were used to determine the quality and quantity of the final pool of libraries. Paired-end Illumina sequencing (150 base pairs [bp]) was performed on a NextSeq 500 system (Illumina, San Diego, CA, USA) using the high output kit v2 with 300 cycles.

2.1.4. Transcriptome assembly

Prior to transcriptome assembly, raw sequencing reads were assessed for quality using FASTQC (v1.0.0) software (Illumina Basespace Labs). Specifically, each RNA-Seq brain library was quality filtered using FASTQ Toolkit (v.2.0.0) by trimming the first 9 bp of each read,

removing all Illumina adapters (TruSeqLT universal primer), culling all low quality reads (Phred cutoff score = 30), and setting the minimum read length to 50 bp. This quality filtering resulted in the removal of < 1% of the reads present in each library, leaving from ~34 to ~39 million filtered reads per brain sample.

Filtered reads were *de novo* assembled using Trinity (v2.0.6) software (https://github.com/trinityrnaseq/trinityrnaseq/wiki; Grabherr et al., 2011) on the National Center for Genome Analysis Support's (Indiana University, Bloomington, IN, USA) Mason Linux cluster. Each node of the computer system includes four Intel Xeon L7555 8-core processors running at 1.87 GHz with 512 GB of memory. For the assembly, reads from all brain libraries were combined and the minimum sequence length in the assembly was set to 324 bp. For the *de novo* assembly, the initial parameters of Trinity were set as follows: maximum memory, 200 GB; CPU, 32; normalize maximum read coverage, 50; minimum contig length, 324. Assembly statistics were obtained using the script TrinityStat.pl (v2.0.6). Quality filtered raw reads were mapped against the *de novo* assembled transcriptome using Bowtie2 (v2.1.0) software (Langmead et al., 2009).

2.2. Identification and vetting of circadian signaling systems transcripts/proteins

2.2.1. Transcriptome mining

Searches of the *H. americanus* brain and eyestalk ganglia transcriptomes were conducted using BLAST software installed on an Intel-processor-based BEOWULF computer cluster (Pacific Biosciences Research Center, University of Hawaii at Manoa, Honolulu, HI, USA) and methods modified from a protocol used previously for the discovery of circadian transcripts/proteins in other crustaceans (Christie et al., 2013, 2018; Nesbit and Christie, 2014; Roncalli et al., 2017; Tilden et al., 2011). Specifically, known circadian proteins, primarily those from the fruit fly *D. melanogaster* (*e.g.*, Adams et al., 2000; Bae et al., 1998; Dombrádi et al., 1990; Myers et al., 1995), were input into tblastn as query sequences. The complete list of proteins searched for in this study, as well as the specific queries used, is provided in Supplemental Table 1.

2.2.2. Protein prediction and confirmation of protein attributions

A workflow developed to vet the identification of a variety of proteins, including those involved in circadian signaling (Christie et al., 2013, 2018; Nesbit and Christie, 2014; Roncalli et al., 2017; Tilden et al., 2011), was used to characterize the sequences deduced from the *H. americanus* transcripts. First, nucleotide sequences were translated using the "Translate" tool of ExPASy (http://web.expasy.org/translate/) and assessed for completeness. Proteins listed as "full-length" exhibit a "start" methionine and are flanked on their carboxyl (C)-termini by a stop codon. Proteins described here as "partial" lack a start methionine (referred to as C-terminal partial proteins) or a stop codon (referred to as amino [N]-terminal partial proteins). Next, to confirm that each of the proteins identified here is most similar to the *D. melanogaster* sequence used to identify the transcript encoding it, each *Homarus* protein was used as the input query in a BLAST search of the annotated *Drosophila* protein dataset present in FlyBase (version FB2016_05; Gramates et al., 2017). For cryptochrome 1 (CRY1) and cryptochrome 2 (CRY2), proteins from monarch butterfly, *Danaus plexippus,* were used as the original query sequences (legend of Supplemental Table 1), as this species possesses both proteins, while *D. melanogaster* lacks CRY2 (Yuan et al., 2007); thus, for the putative *H. americanus* CRYs, the extant *D. plexippus* protein dataset present in NCBI (taxid:13037) was used for the reciprocal BLAST rather than the FlyBase *D. melanogaster* dataset. The arthropod protein most similar to each *Homarus* sequence was subsequently determined by conducting a BLAST search of the non-redundant arthropod proteins curated at NCBI (taxid:6656), using each of the deduced *H. americanus* proteins as the input query. Finally, using the online program Pfam version 29.0 (http://pfam.xfam.org/; Finn et al., 2016), protein structural motifs were analyzed for each of the *H. americanus* proteins identified. This workflow was conducted for the deduced *Homarus* proteins on or before August 1, 2017.

Protein alignments were done using the online program MAFFT version 7 (http://mafft.cbrc.jp/alignment/software/; Katoh and Standley, 2013).

2.3. PCR confirmation of sequences

To confirm expression of *H. americanus* core clock genes (*i.e.*, clock [CLK], CRY2, cycle [CYC], PER and timeless [TIM]), fragments (~500 nucleotides) of each were amplified from brain and eyestalk ganglia cDNAs (n = 3 biological replicates for each portion of the CNS). Total RNA was purified from isolated brain and eyestalk ganglia samples as described above and then treated with DNase I (New England Biolabs, Ipswich, MA, USA) for 10 min at 37 °C to remove contaminating genomic DNA. cDNAs were synthesized from 500 ng total RNA using a SuperScript III First-Strand Synthesis System (Life Technologies, Carlsbad, CA, USA) with custom-made random pentadecamers (IDT, San Diego, CA, USA). Fragments of the transcripts of interest were amplified using SapphireAmp Fast PCR Master Mix (Takara Bio USA, Inc., Mountain View, CA, USA) in a 20 µl reaction volume with 0.4 µl cDNA and oligonucleotide primers (Supplemental Table 2) designed to the respective transcriptomic sequences using Primer 3 v2.3.7 (Rozen and Skaletsky, 2000) implemented in Geneious v10.1.3 (Biomatters Ltd., Auckland, New Zealand; Kearse et al., 2012). PCR conditions consisted of: 95 °C for 2 min, then 40 cycles of 95 °C for 20 s, 56 °C for 20 s, and 72 °C for 30 s, with a final extension at 72 °C for 5 min. The resulting products were electrophoresed on 1.5% agarose gels stained with SYBR Safe (Life Technologies), cloned into pCR2.1TOPO TA vector (Life Technologies), and sequenced at the Arizona State University DNA Core laboratory (Tempe, AZ, USA).

To further validate the *in silico* assemblies, oligonucleotide primers (Supplemental Table 2) designed to facilitate amplification of the complete ORF or longest possible fragment of core clock protein-encoding transcripts were designed with Primer 3 based on the transcriptome assembly data. Transcripts were amplified as above in a 20 µl reaction volume with 0.4 µl cDNA and PCR conditions consisting of: 95 °C for 2 min, then 40 cycles of 95 °C for 20 s, 56 °C for 20 s, and 72 °C for 90 s, with a final extension at 72 °C for 5 min. PCR products were cloned and sequenced; resulting sequences were compared with the *in silico* assembled transcripts.

RACE (rapid amplification of cDNA ends) PCR was performed to identify missing sequence data from the 5′ and 3′ ends of the transcriptome-derived TIM and CLK sequences, respectively. RACE cDNAs were generated using a SMARTer RACE cDNA Amplification kit (Takara Bio USA, Inc.) according to the manufacturer's instructions from brain and eyestalk-derived total RNAs. First-round 5′ RACE amplification of the *Homarus* TIM transcripts was performed using a SapphireAmp Fast PCR Master Mix (Takara Bio USA, Inc.) with thermocycler conditions consisting of 95 °C for 2 min followed by 35 cycles at 95 °C for 20 s, 68 °C for 20 s, and 72 °C for 90 s, and ended with a 5-min incubation at 72 °C. Products were cloned into pCR2.1TOPO TA (Life Technologies) and sequenced as described previously. Oligonucleotide primers (Supplemental Table 2) designed to putative start codons were used in subsequent amplification reactions to validate the RACE results with SapphireAmp Fast PCR Master Mix (Takara Bio USA, Inc.) and PCR conditions of 95 °C for 2 min followed by 40 cycles at 95 °C for 20 s, 56 °C for 20 s, and 72 °C for 2 min, and ended with a 5-min incubation at 72 °C. The resulting PCR products were electrophoresed on 1% agarose gels stained with SYBR Safe (Life Technologies); amplimers of the expected sizes were gel-excised and purified using an EZNA gel extraction kit (Omega Bio-Tek, Norcross, GA, USA). The excised products were cloned and sequenced as described previously.

Although the same conditions were initially used for 3′ RACE of the *Homarus* CLK transcript, no products were generated. Consequently, the

thermocycler conditions were modified to 95 °C for 2 min followed by 5 cycles at 94 °C for 30 s, 70 °C for 20 s, and 72 °C for 90 s, then 5 cycles at 94 °C for 30 s, 67 °C for 20 s, and 72 °C for 90 s, another 30 cycles at 94 °C for 30 s, 62 °C for 20 s, and 72 °C for 90 s, and ended with a 5-min incubation at 72 °C. The resulting products were then used as templates for nested RACE reactions using SapphireAmp Fast PCR Master Mix (Takara Bio USA, Inc.) with thermocycler conditions consisting of 95 °C for 2 min followed by 5 cycles at 94 °C for 30 s, 70 °C for 20 s, and 72 °C for 90 s, then 5 cycles at 94 °C for 30 s, 67 °C for 20 s, and 72 °C for 90s, 30 cycles at 94 °C for 30 s, 62 °C for 20 s, and 72 °C for 90 s with a final hold at 72 °C for 5 min. Products from the non-nested and nested reactions were cloned and sequenced; however, no CLK-like sequences were identified.

2.4. Generation of cladograms

To examine the phylogenetic relationship of *Homarus* PER relative to PER and CYC sequences from a diverse set of arthropods, multiple cladograms were constructed. Phylogenetic analyses (minimum evolution, UPGMA, neighbor joining, and maximum-likelihood) were performed in MEGA 6.06 (Tamura et al., 2013) with bootstrap support based on 1000 iterations, using protein sequences either previously described/annotated (*e.g.*, Adams et al., 2000; Chen et al., 2017; Chesmore et al., 2016; Christie et al., 2013, 2018; Moriyama et al., 2008; Nesbit and Christie, 2014; Roncalli et al., 2017; Sbragaglia et al., 2015; Shin et al., 2012; Tilden et al., 2011; *Tribolium* Genome Sequencing Consortium, 2008; Uryu et al., 2013; Zhan et al., 2011; Zhang et al., 2013) or identified here for the first time from publicly accessible transcriptome shotgun assembly data (Supplemental Fig. 1). PER and Bmal, a synonym for CYC, from the mollusk *Crassostrea gigas* (Accession Nos. AQM57604 and AQM57603, respectively [Perrigault and Tran, 2017]) were used as outgroups. Evolutionary distances were computed using the Jones-Taylor-Thornton matrix-based method (Jones et al., 1992) from 63 amino acid sequences in which positions containing gaps and/or missing data were eliminated for a total of 211 positions in the final dataset. Although only the maximum-likelihood analyses have been described in the text, those results were consistent with topologies generated by the other three analyses (minimum evolution, UPGMA, and neighbor joining). Similar phylogenetic inferences to support the initial *H. americanus* CRY1 and CRY2 annotations were likewise performed with bootstrap support based on 1000 iterations, using protein sequences either previously described/annotated (*e.g.*, Christie et al., 2018; Mat et al., 2016; Mazzotta et al., 2010; Nesbit and Christie, 2014; O'Grady et al., 2016; Roncalli et al., 2017; Zhan et al., 2011; Zhang et al., 2013; Zhu et al., 2005) or identified here for the first time from publicly accessible transcriptome shotgun assembly data (Supplemental Fig. 2). As for the PER-CYC cladogram, *C. gigas* CRY1 and CRY2 sequences (Accession Nos. ANJ02841 [Mat et al., 2016] and AQM57602 [Perrigault and Tran, 2017], respectively) were used as outgroups.

3. Results and discussion

3.1. Generation of brain- and eyestalk ganglia-specific transcriptomes

To identify circadian protein-encoding transcripts in the brain and eyestalk ganglia, a transcriptome for each of these regions of the *H. americanus* CNS was generated. The production of the brain-specific transcriptome is described here (see Materials and Methods); the generation of the eyestalk ganglia-specific assembly is described in an earlier study (Christie et al., 2017). Four samples consisting of either individual brains or eyestalk ganglia pairs, all collected from the same four lobsters, were used as the sources of RNA for the two transcriptomes (Table 1). Each brain or eyestalk ganglia sample was used to generate a unique cDNA library, which was subsequently sequenced using an Illumina NextSeq 500 system; the sequencing of all samples was done simultaneously. Illumina sequencing yielded ~35 million paired-end reads per library for the brain samples and ~33 million paired-end reads per library for eyestalk ganglia samples (Table 1). After trimming and quality filtering, these reads, 141,673,802 for the brain and 130,973,220 for the eyestalk ganglia, were assembled using Trinity (Grabherr et al., 2011) to produce the brain- and eyestalk ganglia-specific transcriptomes (Table 2). The brain transcriptome consists of 150,692 transcripts and 113,172 "Trinity predicted genes", while the eyestalk ganglia transcriptome consists of 147,542 transcripts and 110,841 "Trinity predicted genes" (Table 2). For both transcriptomes, 87% of the "Trinity predicted genes" are singletons, (98,392 for the brain transcriptome and 96,484 for the eyestalk ganglia transcriptome), with the remaining genes (14,780 for the brain and 14,357 for the eyestalk ganglia) having from two to either 52 (brain) or 55 (eyestalk ganglia) "Trinity predicted isoforms". The average transcript length for both the brain and eyestalk ganglia transcriptomes is essentially identical, 1227 bp and 1241 bp, respectively, as is their N50 length, 2134 bp and 2160 bp, respectively (Table 2). Similarly, the shortest and longest transcript lengths are nearly identical for the two transcriptomes, *i.e.*, 324 and 24,063 bp, respectively, for the brain transcriptome, and 324 and 27,389 bp, respectively, for the eyestalk ganglia transcriptome (Table 2). Mapping of the Illumina-generated reads from the brain or eyestalk ganglia against the respective complete assembly using Bowtie [Langmead et al., 2009] yielded an overall alignment rate of 91% for each assembly, with 48% of the brain and 52% of the eyestalk ganglia reads mapping just once (Table 3). Thus, all assembly and mapping metrics for both the brain and eyestalk ganglia transcriptomes are comparable, and suggest that the two transcriptomes are of high quality. The two transcriptomes (and their

Table 1
Summary of *Homarus americanus* brain and eyestalk ganglia samples and their Illumina sequencing output.

Lobster		Brain		Eyestalk ganglia[a]	
Individual	Sex	Total RNA concentration (ng/μl)	Raw read	Total RNA concentration (ng/μl)	Raw reads
1	M	363	38,982,128	223	26,829,856
2	M	360	34,750,554	268	38,085,928
3	F	436	33,823,408	316	32,879,458
4	M	328	34.117,712	262	33,177,978

[a] Data from Christie et al., 2017.

Table 2
Assembly statistics for the *Homarus americanus* brain and eyestalk ganglia transcriptomes.

Metric	Brain	Eyestalk ganglia[a]
Total number of bases assembled	184,921,128	183,106,109
Total number of reads assembled	141,673,802	130,973,220
Total number of assembled transcripts	150,692	147,542
Total number of Trinity "genes"	113,172	110,841
Minimum contig length (bp)	324	324
Maximum contig length (bp)	24,063	27,389
Average contig length (bp)	1227	1241
Median contig length (bp)	631	637
N25 (bp)	4157	4193
N50 (bp)	2134	2160
N75 (bp)	854	869
Total GC count (bp)	75,247,413	74,888,256
GC content for the complete assembly (%)	40.69	40.90

Abbreviations: bp, base pairs; GC, guanine-cytosine.
Transcriptome assembly statistics were generated using Trinity software. Reads used for the *de novo* assembly were trimmed for Illumina adapters and quality filtered (Phred score = 30).
[a] Data from Christie et al., 2017.

Table 3
Summary of the results of mapping RNA-Seq reads to the complete *Homarus americanus* brain and eyestalk ganglia assemblies.

Metric	Brain	Eyestalk ganglia[b]
Total reads used for mapping	141,673,802	130,973,220
Total mapped read[a]	129,552,197	119,120,860
Overall alignment (%)	91%	91%
Reads mapped 1 time (#)	68,633,504	68,454,575
Reads mapped 1 time (%)	48%	52%
Reads mapped > 1 time (#)	60,918,693	50,666,285
Reads mapped > 1 time (%)	43%	39%

[a] 88% of brain and 97% of eyestalk mapped reads aligned as clusters (read pairs).
[b] Data from Christie et al., 2017.

associated data) have been deposited in GenBank under BioProject Nos. PRJNA379629 (brain) and PRJNA338672 (eyestalk ganglia; Christie et al., 2017).

3.2. Strategy/rational for the discovery of Homarus circadian signaling system transcripts

Circadian signaling systems have been documented in essentially all living organisms, and the molecular machinery (genes/proteins) that underlies these timekeepers is highly conserved across taxa (Hut and Beersma, 2011). Given the level of conservation seen for most circadian genes/proteins, we used a strategy for identifying putative *H. americanus* clock components based on homology to known arthropod proteins, in particular those from the fruit fly, *D. melanogaster*, which has arguably the most thoroughly investigated circadian signaling system in the animal kingdom (Allada and Chung, 2010; Hardin, 2011; Mendoza-Viveros et al., 2017; Ozkaya and Rosato, 2012; Yoshii et al., 2015). To vet the *Homarus* protein identifications, reciprocal BLAST searches were conducted using the annotated *D. melanogaster* proteins curated in FlyBase as one target dataset; FlyBase is one of the largest, most complete, and most thoroughly characterized single-species arthropod protein databases extant (Gramates et al., 2017). The *H. americanus* protein identifications were further vetted by blastp searches against the non-redundant arthropod protein dataset curated in NCBI, which allowed for broad species comparisons, and by structural motif analysis using the online program Pfam (Finn et al., 2016). The rational for employing this workflow was that it has proven highly effective for protein discovery in a variety of crustacean species, including the identification of proteins putatively involved in the establishment of circadian signaling systems (Christie et al., 2013, 2018; Nesbit and Christie, 2014; Roncalli et al., 2017; Tilden et al., 2011).

3.3. Identification of putative circadian protein-encoding transcripts/proteins

Four categories of genes/proteins are responsible for the generation of circadian rhythms: core clock, clock-associated, clock input pathway, and clock output pathway genes/proteins (Allada and Chung, 2010). The genes/proteins that form the core clock are those that participate in the basic molecular feedback loop required for ~24-h cyclical timing, while clock-associated genes/proteins participate in modulating core clock function. Clock input pathway genes/proteins provide a means for synchronization of the core clock to the solar day, while clock output pathway proteins serve to transmit timing information from clock cells to affect physiology and behavior both directly and indirectly.

3.3.1. Core clock proteins

In *D. melanogaster* clock neurons, four interacting genes/proteins form the feedback loop of the core clock (Allada and Chung, 2010; Hardin, 2011; Mendoza-Viveros et al., 2017; Ozkaya and Rosato, 2012; Yoshii et al., 2015). Specifically, this molecular cascade is initiated through the binding of a CLK and CYC heterodimer to E-box elements located in the promoter regions of the *per* and *tim* genes, activating their transcription. Following transcription and translation, PER and TIM proteins accumulate and dimerize in the cytoplasm, with the resulting heterodimer translocated to the nucleus. In the nucleus, PER-TIM binds to CLK-CYC, inhibiting the latter heterodimer's activation of the genes encoding the former two proteins. Although it is the most extensively studied, the *Drosophila* circadian system is not necessarily stereotypical, as it is missing one protein, CRY2, that is a key component of "ancestral-type" arthropod core clocks (Tomioka and Matsumoto, 2010; Yuan et al., 2007), *e.g.*, that of the monarch butterfly *D. plexippus*. In *Danaus* and other species that possess it, CRY2 is postulated to function as an inhibitor of CLK-CYC-mediated transcription (Yuan et al., 2007).

BLAST searches using known *D. melanogaster* or *D. plexippus* proteins as query sequences identified transcripts encoding putative homologs of each of the five core clock protein families in both the *H. americanus* brain- and eyestalk ganglia-specific transcriptomes (Supplemental Table 1). As shown in Table 4, translation of these transcripts revealed a combined complement of one isoform of CLK (Homam-CLK; Fig. 1), one isoform of CRY2 (Homam-CRY2; Fig. 2), two isoforms of CYC (Homam-CYC-v1 and v2; Fig. 3 and Supplemental Fig. **3**), 12 isoforms of PER (Homam-PER-v1 through v11 [Homam-PER-v8 having v8a and v8b sub-variants]; Fig. 4 and Supplemental Fig. **4**), and three isoforms of TIM (Homam-TIM-v1a/b and v2; Fig. 5 and Supplemental Fig. **5**). Nearly all the *Homarus* proteins deduced from the brain and eyestalk ganglia transcriptomes appear to be full-length sequences. The exceptions are Homam-CLK, which is an N-terminal partial protein, and the three TIM variants, all of which appear to be C-terminal partial sequences (Table 4); RACE (see below) subsequently revealed the first methionine in the putative partial Homam-TIM-v1a to be the true start of this protein, indicating that it is a full-length sequence (Fig. 5). Homam-CYC-v2 is a truncated version of Homam-CYC-v1, being identical in sequence to the C-terminus of the longer protein (Supplemental Fig. **3**). The 12 *Homarus* PER variants differ from one another largely by the presence/absence of multiple insertions/deletions localized to four variable regions (Supplemental Fig. **4**). Homam-TIM-v1a and Homam-TIM-v1b, while of slightly different lengths, are identical in sequence over their region of overlap, except for five substituted residues (Supplemental Fig. **5**A); the N-terminus of Homam-TIM-v2 is distinct from that of the other two variants (Supplemental Fig. **5**A). Based on amino acid sequence variation and the unique Trinity-assigned transcript identifiers, it appears that alternative splicing of single genes is responsible for CYC, PER and TIM protein diversity in *H. americanus*.

To strengthen our confidence that the proteins described above truly represent *H. americanus* members of the CLK, CRY2, CYC, PER and TIM families, each was used as the query sequence in a blastp search of the annotated *D. melanogaster* proteins in FlyBase, as well as in a blastp search of the non-redundant arthropod proteins curated in NCBI. The expectations here were that each putative *Homarus* sequence would return an isoform of the relevant protein family as the top hit in each BLAST search. For the FlyBase searches using the putative *H. americanus* CLK, CYC, and TIM proteins as input queries, this was indeed the case (Supplemental Table 3), with an isoform of the relevant *Drosophila* protein family returned as the top BLAST hit for each sequence, *e.g.*, Drome-TIM-B (Accession No. AAF51098) for Homam-TIM-v1a and Homam-TIM-v1b and Drome-TIM-T (Accession No. API64979) for Homam-TIM-v2. Interestingly, and quite unexpectedly, searches of FlyBase using the 12 *Homarus* PERs as the input sequences returned PER as the top *D. melanogaster* BLAST hit for only four of the isoforms (Homam-PER-v4, Homam-PER-v8a/b and Homam-PER-v9), with the remaining eight lobster proteins returning CYC as the top hit (Supplemental Table 3), despite the fact that all of the proteins appear to be derived from a common gene; this conundrum is addressed and resolved below (see Section *3.5*). As *D. melanogaster* does not possess a CRY2 homolog, a blastp search of the extant *D. plexippus* (monarch

Table 4
Tissue distribution of deduced *Homarus americanus* circadian signaling system proteins.

Clock component	Protein family	*Homarus* isoform	Tissue detection	
			Brain	Eyestalk ganglia
Core clock protein	CLK	Homam-CLK	+	+
	CRY2	Homam-CRY2	+	+
	CYC	Homam-CYC-v1	+	+
		Homam-CYC-v2	−	+
	PER	Homam-PER-v1	+	+
		Homam-PER-v2	+	−
		Homam-PER-v3	+	+
		Homam-PER-v4	−	+
		Homam-PER-v5	−	+
		Homam-PER-v6	+	+
		Homam-PER-v7	+	+
		Homam-PER-v8a	+	−
		Homam-PER-v8b	−	+
		Homam-PER-v9	−	+
		Homam-PER-v10	−	+
		Homam-PER-v11	−	+
	TIM	Homam-TIM-v1a	+	−
		Homam-TIM-v1b	−	+
		Homam-TIM-v2	+	−
Clock-associated protein	CKIIα	Homam-CKIIα-v1	+	+
		Homam-CKIIα-v2	+	−
	CKIIβ	Homam-CKIIβ	+	+
	CWO	Homam-CWO-v1	+	+
		Homam-CWO-v2	+	+
		Homam-CWO-v3	−	+
		Homam-CWO-v4	−	+
	DBT	Homam-DBT-v1	+	+
		Homam-DBT-v2	+	+
		Homam-DBT-v3	+	+
		Homam-DBT-v4	+	+
		Homam-DBT-v5	−	+
	JET	Homam-JET-v1	+	+
		Homam-JET-v2	+	+
		Homam-JET-v3	+	+
		Homam-JET-v4	−	+
		Homam-JET-v5	+	+
		Homam-JET-v6	+	−
		Homam-JET-v7	+	+
		Homam-JET-v8	+	+
		Homam-JET-v9	+	−
		Homam-JET-v10	−	+
		Homam-JET-v11	−	+
	PDP1	Homam-PDP1-v1	+	+
		Homam-PDP1-v2	+	+
		Homam-PDP1-v3	+	+
		Homam-PDP1-v4	+	+
		Homam-PDP1-v5	+	+
		Homam-PDP1-v6	+	+
		Homam-PDP1-v7	−	+
		Homam-PDP1-v8	−	+
	PP1	Homam-PP1-I	+	+
		Homam-PP1-II-v1	+	−
		Homam-PP1-II-v2	−	+
		Homam-PP1-III	+	+
	PP2A MTS	Homam-MTS-I	+	+
		Homam-MTS-II-v1	+	+
		Homam-MTS-II-v2	+	−
	PP2A TWS	Homam-TWS-v1	+	+
		Homam-TWS-v2	+	+
		Homam-TWS-v3	+	+
		Homam-TWS-v4	+	−
	PP2A WDB	Homam-WDB	+	+
	SGG	Homam-SGG	+	+
	SLIMB	Homam-SLIMB-v1	+	−
		Homam-SLIMB-v2	−	+
		Homam-SLIMB-v3	+	+
		Homam-SLIMB-v4	−	+
	VRI	Homam-VRI	+	+

(*continued on next page*)

Table 4 (*continued*)

Clock component	Protein family	*Homarus* isoform	Tissue detection	
			Brain	Eyestalk ganglia
Clock input pathway protein	CRY1	Homam-CRY1-v1	+	+
		Homam-CRY1-v2	+	−
		Homam-CRY1-v3	−	+
		Homam-CRY1-v4	+	+
		Homam-CRY1-v5	+	−
		Homam-CRY1-v6	+	−
		Homam-CRY1-v7	+	+
		Homam-CRY1-v8	+	+
		Homam-CRY1-v9	+	−
		Homam-CRY1-v10	+	−
Clock output pathway protein	Prepro-PDH	Homam-prepro-PDH-I	+	+
		Homam-prepro-PDH-II	−	+
		Homam-prepro-PDH-III	+	−
	PDHR	Homam-PDHR-I-v1	−	+
		Homam-PDHR-I-v2	−	+
		Homam-PDHR-I-v3	+	−
		Homam-PDHR-II	+	+

Protein family abbreviations: CLK, clock; CRY2, cryptochrome 2; CYC, cycle; PER, period; TIM, timeless; CKIIα, casein kinase IIα; CKIIβ, casein kinase IIβ; CWO, clockwork orange, DBT, doubletime; JET, jetlag; PDP1, PAR-domain protein 1; PP1, protein phosphatase 1; PP2A MTS, protein phosphatase 2A catalytic subunit microtubule star; PP2A TWS, protein phosphatase 2A regulatory subunit twins; PP2A WDB, protein phosphatase 2A regulatory subunit widerborst; SGG, shaggy; SLIMB, supernumerary limbs; VRI, vrille; CRY1, cryptochrome 1; Prepro-PDH, pigment dispersing hormone precursor; PDHR, pigment dispersing hormone receptor. "+" indicates the putative presence of a protein isoform in the brain and/or eyestalk ganglia, while "-" indicates the putative absence of a protein isoform in the brain and/or eyestalk ganglia.

```
Homam-CLK       MAMSHELIETSPGKSVMDDDGDEKDDVKRKSRNLSEKKRRDQFNLLINELSGMVASNNRK
Macro-CLK       ----------------MDDDADEKDDVKRKSRNLSEKKRRDQFNLLINELSSMVASNNRK
                              ****.*******************************.********

Homam-CLK       MDKSTVLKATIAFLKNQKEISVRSQNQEIREDWKPSFLSNEEFTHLMLEALDGFIMTVSC
Macro-CLK       MDKSTVLKATIAFLKNQKEISVRTQSNEIREDWKPSFLSNEEFTHLMLEALDGFIMTVSC
                ************************:*.:**********************************

Homam-CLK       SGRVLYTSESITPLLGHLPGDISGTPLYDLMLEEERGDMKRFLSNPALAPNPSTWLEDTK
Macro-CLK       SGRVLYTSESITPLIGHLPSDLAGTPLYDLMLEEERGEMRRFLSNPALAPNPSTCFDNTK
                **************:****.*::***************:*:************** :::**

Homam-CLK       EKYTVAIHLKRGMTNASDSIHYERIHLTGYFERYNCPSEDGVLDFSCSEAEDSVSLSSCS
Macro-CLK       EKYTIAVHLRRGTINSSDATNYERVHLMGYFERYSGPSEDGVLDFSCSEAEDSVSVSSCS
                ****:*:**:**  *:**: :***:** ******. ********************:****

Homam-CLK       RS---------------AGGGLFQSSLVHPQLTQETTKLVFVAIARMERPQLVREMMIIE
Macro-CLK       RSMFGGNAGGGVTGSSNAGSGLCQSSLVQTNPSQEPTKLVFVAIGRLERPQLVREMMIIE
                **                **.** *****:.: :**.********.*:**************

Homam-CLK       PTKTEFTSRHSLEWKFLFLDHRAPTIIGYMPFEVLGTSGYDYYHVDDLEKIAACHEQLMK
Macro-CLK       PSKTEFTSRHSLEWKFLFLDHRAPTIIGYLPFEVLGTSGYDYYHVEDLDKVASCHEQLMK
                *:**************************:****************:**:*:*:*******

Homam-CLK       TGKGTSCYYRFLTKGQQWIWLQTQYYITYHQWNSKPEFIVCTNTVVSYSDVKAELSQEVI
Macro-CLK       TGKGTSCYYRFLTKGQQWIWLQTQYYITYHQWNSKPEFIVCTNTVVSYSDVKAELVKEQM
                ********************************************************* :* :

Homam-CLK       PDSCQSEIEIHQPEGSMVLNNAGA----SVAGPSGLAL-----LAPVGELSQ-QLPDEKP
Macro-CLK       PNGL-SELEINQSESSMGLSGAGPSGTNSMSGQTGVGVGVSVSVSGVQEHSQSQLSEEDT
                *:.  **:**:*.*.** *..**.    *::* :*:.:     :: * * ** **.:*..

Homam-CLK       VIQPSRAPLPGPSHL----------------------QPQHLQQQQTQQQLQADQQTQ
Macro-CLK       SLQPARTPQPGPSHLLLQHQHQEHNIPQTHLPLTQNQQQHIQHLQQQQQQQQQQQQQQQQ
                 :**:*:* ******                      : ******* *** * :** *

Homam-CLK       QQQQHSLQQKQLQQQQIHHHQQQQLFAPP-------------------------------
Macro-CLK       QQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ
                ****:. **:* **** :::****

Homam-CLK       ------------------------------------------------------------
Macro-CLK       QQQQQQQQQQQQQQQQQQQQQQQQQQQEQQQQQHQQQQQQQQQQQQQQQQLQQQQLQQHL

Homam-CLK       ------TTAVASVTTTAVAPTPPAAVAATARGPVPPPAVGQTHRKRVGEGQRG-------
Macro-CLK       QQQQHMTQCMPQQTLQLAAPSSPSSKISCS------PSSKQTARRWCSKITRYFKCRTTI
                      * .:.. *   .**:.*::  : :      *:  ** *:  .:  *
Homam-CLK       -------
Macro-CLK       SWPSVSK
```

Fig. 1. MAFFT alignment of the known portion of *Homarus americanus* clock (Homam-CLK) and *Macrobrachium rosenbergii* clock (Macro-CLK). In the line immediately below each sequence grouping, "*" indicates identical amino acid residues, while ":" and "." denote amino acids that are similar in structure between sequences. In this figure, helix-loop-helix DNA-binding, PAS fold, and PAS domains identified by Pfam analyses are highlighted in yellow, light green, and light blue, respectively. Macro-CLK (Accession No. AAX44045; Yang et al., 2006). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

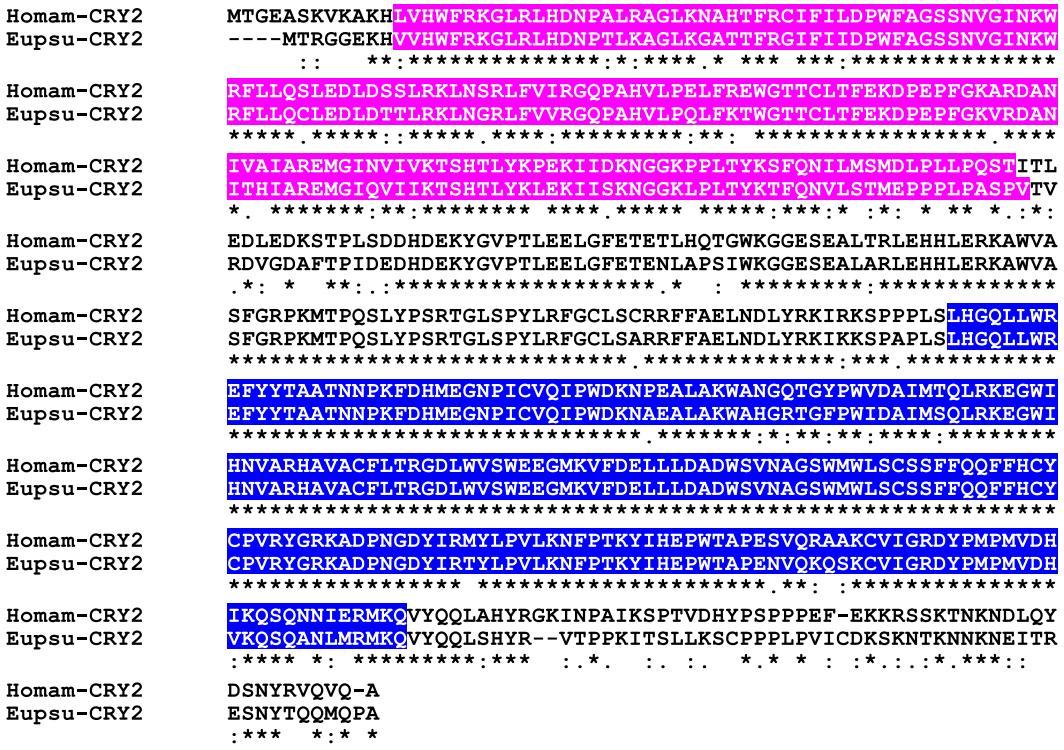


Fig. 2. MAFFT alignment of *Homarus americanus* cryptochrome 2 (Homam-CRY2) and *Euphausia superba* CRY2 (Eupsu-CRY2). In the line immediately below each sequence grouping, "*" indicates identical amino acid residues, while ":" and "." denote amino acids that are similar in structure between sequences. In this figure, DNA photolyase domains and FAD binding domains of DNA photolyase identified by Pfam analyses are highlighted in pink and blue, respectively. Eupsu-CRY2 (Accession No. CAQ86665; Mazzotta et al., 2010). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

butterfly) proteins curated in NCBI was substituted for the Homam-CRY2 reciprocal BLAST, as this was the species from which the original query protein was obtained. As expected, CRY2 was confirmed as the most similar *Danaus* protein to Homam-CRY2 (Supplemental Table 3).

BLAST searches of the non-redundant arthropod proteins curated in NCBI also support each of the putative *H. americanus* core clock proteins described earlier as being members of the protein families for which they were named. Specifically, CLK from the caridean shrimp *Macrobrachium rosenbergii* (Macro-CLK [Fig. 1]; Accession No. AAX44045; Yang et al., 2006) was returned as the top BLAST hit when Homam-CLK was used to probe the NCBI non-redundant arthropod protein dataset (Supplemental Table 4), while a "mammalian-type CRY", a synonym for CRY2, from the krill *Euphausia superba* (Eupsu-CRY2 [Fig. 2]; Accession No. CAQ86665; Mazzotta et al., 2010) was returned as the top BLAST hit for Homam-CRY2 (Supplemental Table 4). The top non-redundant arthropod BLAST hit for both Homam-CYC-v1 and Homam-CYC-v2 was Bmal1, a synonym for CYC, from the crayfish *Pacifastacus leniusculus* (Pacle-Bmal1 [Fig. 3]; Accession No. AFV39705; Soderhall and Lin, unpublished direct GenBank submission), while searches of the non-redundant arthropod protein dataset using the 12 *Homarus* PER variants as input queries identified PER from the Norway lobster *Nephrops norvegicus* (Nepno-PER [Fig. 4]; Accession No. ALC74274; Sbragaglia et al., 2015) as the most similar sequence (Supplemental Table 4). TIM from *N. norvegicus* (Nepno-TIM [Fig. 5]; Accession No. ALC74275; Sbragaglia et al., 2015) was returned as the most similar non-redundant arthropod sequence to each of the three *Homarus* TIM variants (Supplemental Table 4).

As a final means of vetting the identifications of the putative *H. americanus* core clock proteins reported here, each was subjected to structural domain analysis using the online program Pfam (Finn et al., 2016), and the identified domains were compared to those predicted by the program for each sequence's top arthropod non-redundant protein hit (see above). Pfam analysis of Homam-CLK and Homam-CYC-v1 identified one helix-loop-helix DNA-binding, one PAS fold and one PAS domain in each protein (Supplemental Table 5 and Figs. 1 and 3), which are the same domain complements predicted by the program for Macro-CLK and Pacle-Bmal1 (Figs. 1 and 3). Consistent with its truncated nature, only a single PAS domain was predicted by Pfam in Homam-CYC-v2 (Supplemental Table 5). One DNA photolyase domain and one FAD binding domain of DNA photolyase were predicted by Pfam in Homam-CRY2 (Supplemental Table 5 and Fig. 2), which is the same domain complement identified by the program for Eupsu-CRY2 (Fig. 2). One PAS domain and one period protein 2/3C-terminal region were identified by Pfam in each of the 12 *Homarus* PER variants (Fig. 4 and Supplemental Table 5), as well as in Nepno-PER (Fig. 4). A single timeless protein domain was identified by Pfam in each of the *Homarus* TIM variants (Fig. 5 and Supplemental Table 5); no structural/functional domains were identified in Nepno-TIM using Pfam, though this was not surprising as it is a partial sequence, missing the N-terminus that contains the timeless domain identified in the three lobster variants (Fig. 5). Taken collectively, these structural domain predictions/comparisons, in conjunction with the FlyBase and non-redundant arthropod protein BLAST search results, strongly support the *H. americanus* CLK, CRY2, CYC, PER or TIM isoforms identified here as being true members of their respective core clock protein families.

3.3.2. Clock-associated proteins

The molecular feedback loop of the core clock is modulated by a broad collection of proteins commonly referred to as clock-associated proteins (Allada and Chung, 2010; Hardin, 2011; Mendoza-Viveros et al., 2017; Ozkaya and Rosato, 2012; Yoshii et al., 2015). In *D. melanogaster*, the clock-associated proteins known at this time include three kinases, *i.e.*, casein kinase II (CKII), doubletime (DBT) and shaggy (SGG), two phosphatases, *i.e.*, protein phosphatase 1 (PP1) and protein

```
Homam-CYC-v1    MFGMGNFDFPKYRSECNSVASFSSD-GSKKRRGSFIDSNDDDGESIKIPRTAGDWNKRQN
Pacle-BMAL1a    MFGMSNFDYPKYRSECNSLASFSSDGGSKKRRGSFIDSNDDDADSIKIPRTAGDWNKRQN
                ****.***:*********:****** ***************.:****************

Homam-CYC-v1    HSEIEKRRRDKMNTYIMELSSIIPVCTSRKLDKLTVLRMAVQHMKMLRGSLNSYTEGHYK
Pacle-BMAL1a    HSEIEKRRRDKMNTYIMELSSIIPVCTSRKLDKLTVLRMAVQHMKMLRGSLNSYTEGHYK
                ************************************************************

Homam-CYC-v1    PAFLSDDELKNLILQAADGFLFVVGCDRGRILYVSESVYQTLHYTQGELLGTSWFDILHP
Pacle-BMAL1a    PAFLSDDELKNLILQAADGFLFVVGCDRGRILYVSESVYQTLHYTQGELLGTSWFDILHP
                ************************************************************

Homam-CYC-v1    KDLTKVKEQLSCSDISRRERLVDAKTLLPVKTDVPQGLTKLCPGSRRAFFCRMRCKSVPV
Pacle-BMAL1a    KDLTKVKEQLSCSDISRRERLVDAKTLLPVKTDVPQGLTKLCPGSRRAFFCRMRCKTSPV
                ************************************************************

Homam-CYC-v1    LKEEADSSTGCQKKKSKSQSSDKKYSVIHFTGYLKSWAPTKDPLEEDSGSDSDSCNLSCL
Pacle-BMAL1a    LKEEADSSTGCPRKKSKSQSSDKKYSVIHFTGYLKSWAPTKDPLEEDSGSDSDSCNLSCL
                *********** :************************************************

Homam-CYC-v1    VAVGRVHQPLLASSVRETSRLGRAAPTQPIEFISKHTNDGKFVFIDQRASLLLGWLPQEL
Pacle-BMAL1a    VAVGRVHQPLLASSVLDTSRLGSAAPTQPIEFISKHTNDGKFVFIDQRASLLLGWLPQEL
                *************** :***** ************************************

Homam-CYC-v1    LGSSMYEYFHQDDIPFLAETHRATLQNSESCNTQVYRFRTKEGSFVRLQSNWWTFKNPWT
Pacle-BMAL1a    LGSSMYEYFHQDDIAFLAETHKTTLQNSESCNTQVYRFRTKEGSFVRLQSNWWTFKNPWT
                **************.******::************************************

Homam-CYC-v1    KEIEYIISKNSVVTSEAGMVESTMANESVSQSYNNFNEFLNSPDASHNQDNSPLGCAGSS
Pacle-BMAL1a    KEIEYIISKNSVVTSEAELVESTMANESVSQSYSNFSEYLNSPDPNHNQDTSPIGGAGSN
                ***************** :**************.**.*:*****..****.**:* ***.

Homam-CYC-v1    NSNSRLMGGGIQARKIGRQIADEVLDNQRRNDSTSNSPVSPFEGILGSGASDRSFAALLR
Pacle-BMAL1a    NSNSRLIGGGIQAGKIGRGIADEVLDNQWKNDSTCKSPVSPFEGILGSGASDRSFAALLR
                ******:****** **** ********* :****.:************************

Homam-CYC-v1    SDMTAHRTNVKNNVLLSSNTSTCSGSDTSRSTVPNNNHRRLSSMH-NNLANSGERGGSYN
Pacle-BMAL1a    SDMTAHRTNVKNNAVLSSNTSTCSESDTSRSTQLSNNHHRLPSMHNNNLINSSERGGSFS
                *************.:********* *******  .***:**.*** *** **.*****:.

Homam-CYC-v1    HNNNHRQPVSNEGDLMDVVSGRDLETDGTSDSDEAAMAVIMSLLEADAGLGGPVDFSHLP
Pacle-BMAL1a    HNNSHRQPVSNEGDLMDVVSGRDLETDGTSDSDEAAMAVIMSLLEADAGLGGPVDFSHLP
                ***.********************************************************

Homam-CYC-v1    WPLP
Pacle-BMAL1a    WPLP
                ****
```

Fig. 3. MAFFT alignment of *Homarus americanus* cycle (Homam-CYC) and *Pacifastacus leniusculus* Bmal1a (Pacle-Bmal1a), a synonym for CYC. In the line immediately below each sequence grouping, "*" indicates identical amino acid residues, while ":" and "." denote amino acids that are similar in structure between sequences. In this figure, helix-loop-helix DNA-binding, PAS fold, and PAS domains identified by Pfam analyses are highlighted in yellow, light green, and light blue, respectively. Pacle-Bmal1a (Accession No. AFV39705: Soderhall and Lin, direct GenBank submission). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

phosphatase 2A (PP2A; which consists of one catalytic subunit, *i.e.*, microtubule star [MTS] and either of two regulatory subunits, *i.e.*, twins [TWS] or widerborst [WDB]), a ligase, *i.e.*, supernumerary limbs (SLIMB), and an F-box protein, *i.e.*, jetlag (JET), all of which play key roles in controlling the state of phosphorylation and/or the degradation of core clock elements. In addition, several other clock-associated genes/proteins, *i.e.*, clockwork orange (CWO), PAR-domain protein 1 (PDP1), and vrille (VRI), function as regulators of distinct but interdependent feedback loops involving the CLK-CYC heterodimer; these genes/proteins are postulated to play roles in modulating the phase, amplitude and rhythmic output of the core clock.

BLAST searches using known *D. melanogaster* proteins as the query sequences identified transcripts encoding putative homologs of each of the abovementioned clock-associated protein families in both the *H. americanus* brain- and eyestalk ganglia-specific transcriptomes (Supplemental Table 1). As shown in Table 4, translation of the collective sets of transcripts revealed two CKII α-subunits (Homam- CKIIα-v1 and v2; Supplemental Fig. 6A1–2), one CKII β-subunit (Homam-CKIIβ; Supplemental Fig. **6**B), four CWOs (Homam-CWO-v1 through v4; Supplemental Fig. **6**C1-4), five DBTs (Homam-DBT-v1 through v5; Fig. 6), 11 JETs (Homam-JET-v1 through v11; Supplemental Fig. 6D1–11), eight PDP1s (Homam-PDP1-v1 through v8; Supplemental Fig. 6E1–8), four PP1s (Homam-PP1-I, Homam-PP1-II-v1 and v2 and Homam-PP1-III; Fig. 7), three PP2A MTSs (Homam-MTS-I and Homam-MTS-II-v1 and v2; Fig. 8), four PP2A TWSs (Homam-TWS-v1 through v4; Supplemental Fig. 6F1–4), one PP2A WDB (Homam-WDB; Supplemental Fig. **6**G), one SGG (Homam-SGG; Supplemental Fig. **6**H), four SLIMBs (Homam-SLIMB-v1 through v4; Supplemental Fig. 6I1–4), and one VRI (Homam-VRI; Supplemental Fig. **6**J). These proteins are all full-length sequences except for Homam-PP1-III and Homam-VRI, both of which are N-terminal fragments (Table 4). With the exception of PP1 and MTS, sequence diversity within the predicted families appears to arise from alternative splicing and/or individual-specific variation in single genes (*e.g.*, the five DBT variants shown in Fig. 6). In contrast, the protein diversity observed in the *H. americanus* PP1 and MTS sequences is, at least in part, the result of gene duplication, with three genes predicted for PP1 and two for MTS (Figs. 7 and 8). The protein products of these sets of putative paralog genes share significant amino conservation. For example, Homam-PP1-I is 85.1%/95.1% and 84.8%/94.8% identical/similar in amino acid sequence to Homam-PP1-II-v1 and Homam-PP1-II-v2, respectively (Fig. 7); Homam-PP1-II-v1 and Homam-PP1-II-v2 differ from one another only by the presence of an additional lysine residue at the C-terminal of the latter protein (Fig. 7). Homam-PP1-III also shares considerable amino acid conservation with the other *Homarus* PP1s (Fig. 7), though due to its partial nature, no formal sequence comparisons were made for it. Similarly, Homam-MTS-I is 57.1%/83.1% and 47.6%/70.4% identical/similar in amino acid sequence to Homam-MTS-II-v1 and Homam-MTS-II-v2, respectively (Fig. 8); the two MTS-II variants differ from one another in that Homam-MTS-II-v2 is N-terminally extended by 59 amino acids (Fig. 8). Expansions of PP1 and MTS genes have been noted in several other crustaceans, *i.e.*, the krill *Meganyctiphanes norvegica*, the copepods *Tigriopus californicus* and *Labidocera madurae* and the cladoceran *Daphnia pulex* for PP1 (Christie et al., 2018; Nesbit and Christie, 2014; Roncalli

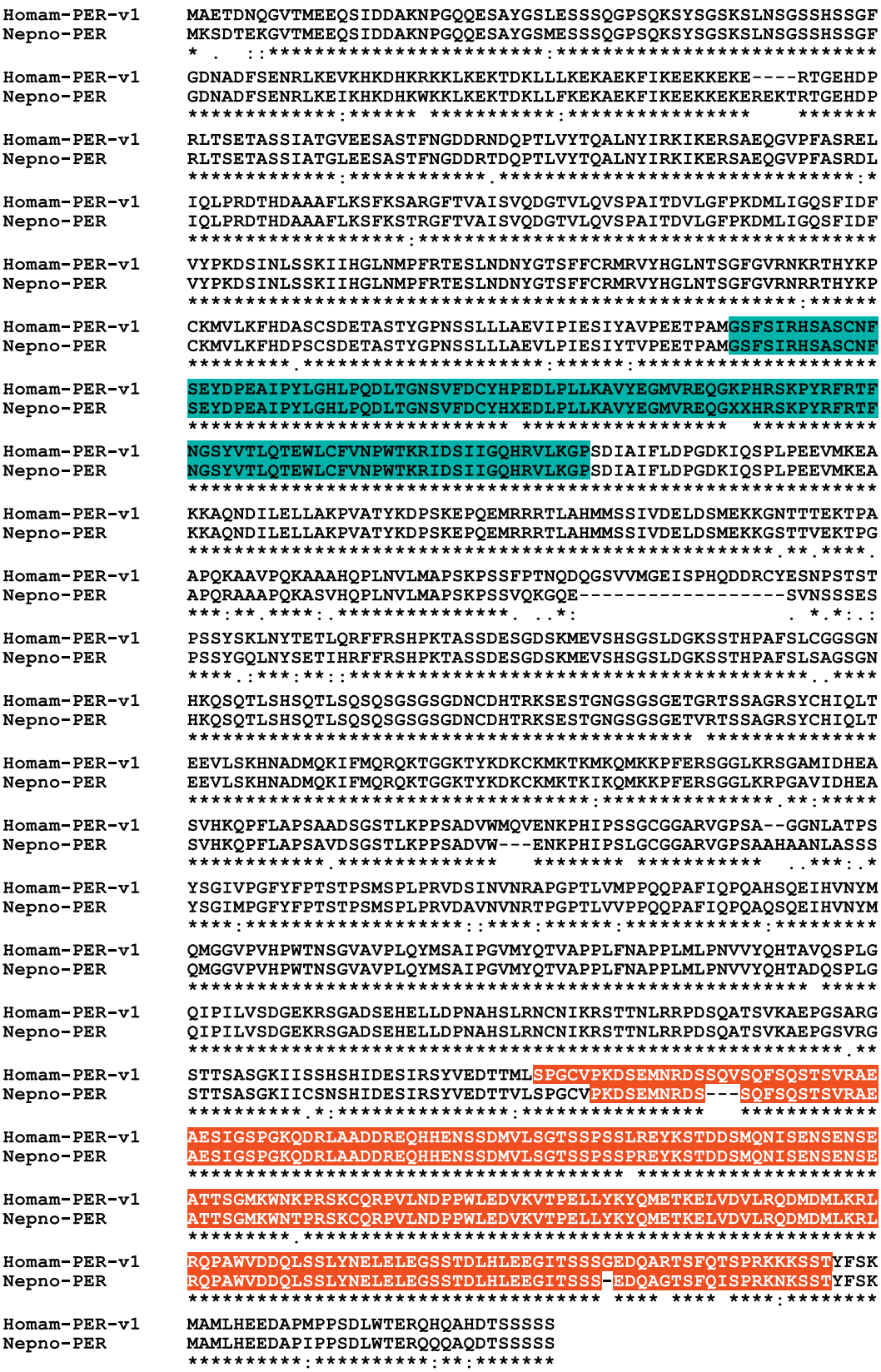


Fig. 4. MAFFT alignment of *Homarus americanus* period variant 1 (Homam-PER-v1) and *Nephrops norvegicus* period (Nepno-PER). In the line immediately below each sequence grouping, "*" indicates identical amino acid residues, while ":" and "." denote amino acids that are similar in structure between sequences. In this figure, PAS domains and period protein 2/3C-terminal regions identified by Pfam analyses are highlighted in light blue and red, respectively. Nepno-PER (Accession No. ALC74274; Sbragaglia et al., 2015). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

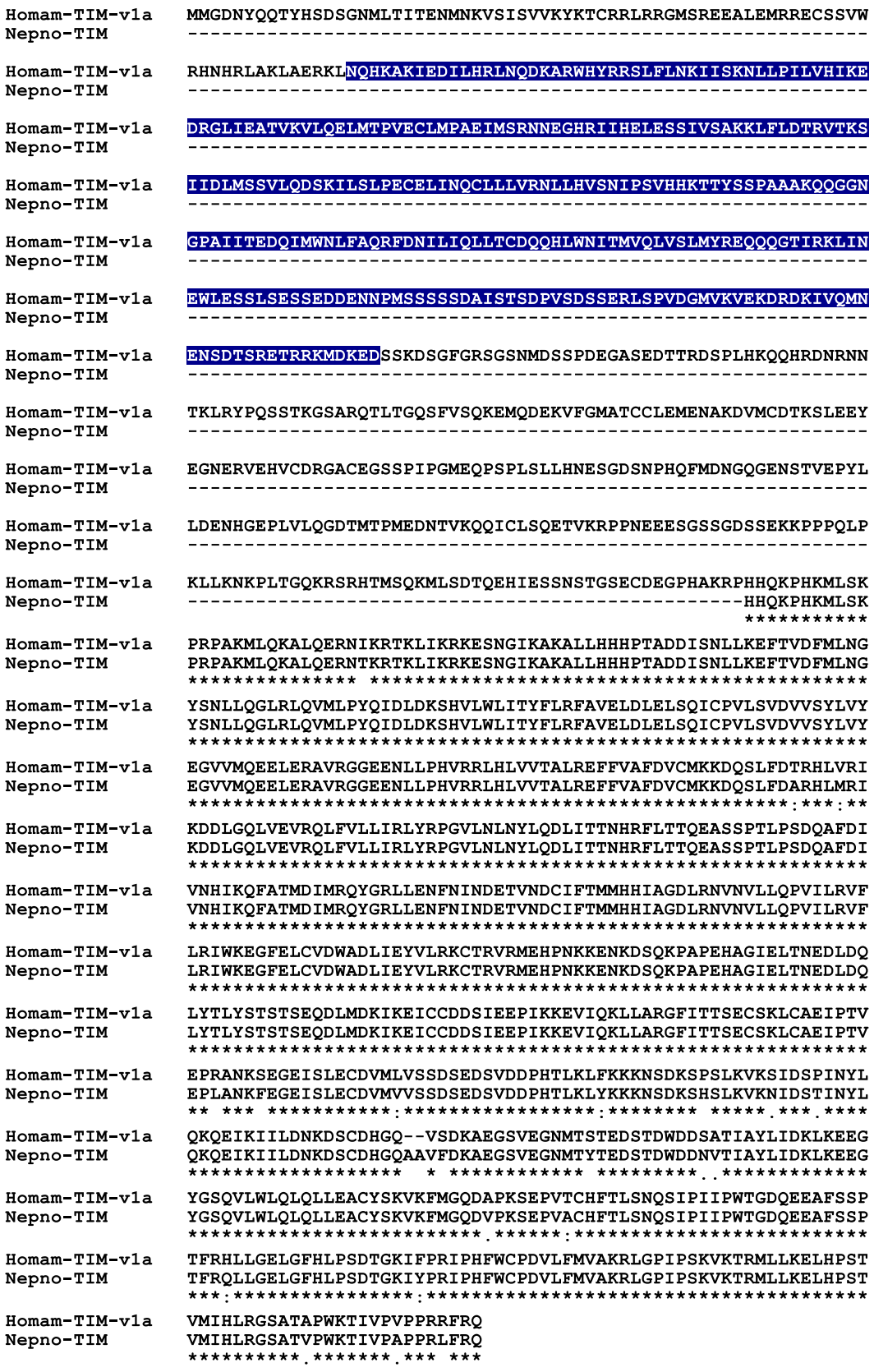


Fig. 5. MAFFT alignment of *Homarus americanus* timeless variant 1a (Homam-TIM-v1a) and the known portion of *Nephrops norvegicus* timeless (Nepno-TIM). In the line immediately below each sequence grouping, "*" indicates identical amino acid residues, while ":" and "." denote amino acids that are similar in structure between sequences. In this figure, a timeless protein domain identified by Pfam analysis is highlighted in dark blue. Nepno-TIM (Accession No. ALC74275; Sbragaglia et al., 2015). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

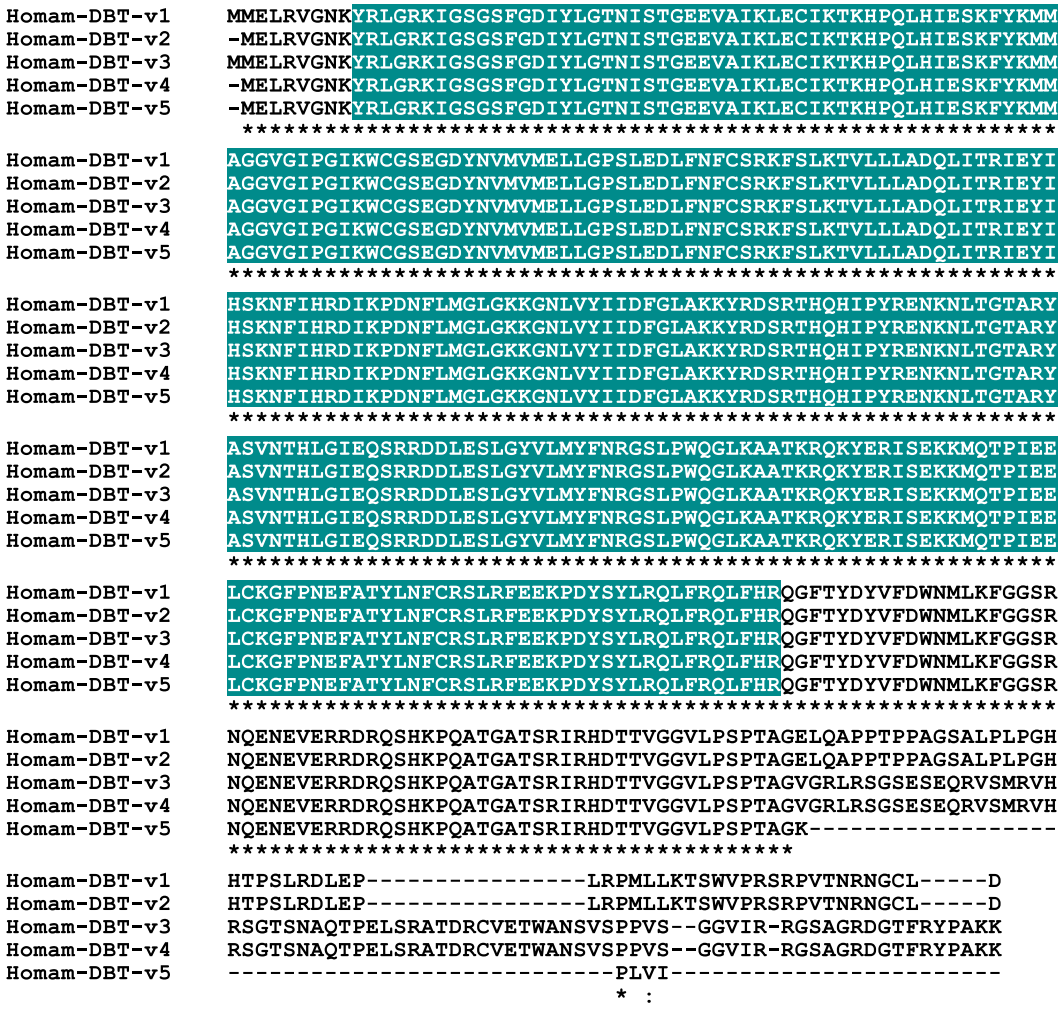


Fig. 6. MAFFT alignment of *Homarus americanus* doubletime (Homam-DBT) variants. In the line immediately below each sequence grouping, "*" indicates identical amino acid residues, while ":" and "." denote amino acids that are similar in structure between sequences. In this figure, protein kinase domains identified by Pfam analyses are highlighted in teal. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

et al., 2017; Tilden et al., 2011), and the copepods L. *maderae* and *Calanus finmarchicus* for MTS (Christie et al., 2013; Roncalli et al., 2017).

Searches of FlyBase confirmed an isoform of the relevant protein family as the top *D. melanogaster* hit for each of the *Homarus* clock-associated proteins identified here (Supplemental Table 3). Similarly, the top hit from BLAST searches of the non-redundant arthropod proteins curated in NCBI provide support for the hypothesis that each of the putative *H. americanus* clock-associated proteins is a member of the protein family for which it was named (Supplemental Table 4). Finally, structural domain analyses using Pfam identified domain complements in the *Homarus* clock-associated proteins that are consistent with these proteins being true members of their respective families (Supplemental Table 5).

3.3.3. Clock input pathway proteins

CRY1 is an ultraviolet/blue light photoreceptor that serves as a clock input pathway protein to entrain the core clock to the solar day (Allada and Chung, 2010; Hardin, 2011; Mendoza-Viveros et al., 2017; Ozkaya and Rosato, 2012; Yoshii et al., 2015). BLAST searches using CRY1 from the monarch butterfly *D. plexippus* as the query template identified multiple transcripts encoding putative CRY1 homologs in both the *H. americanus* brain and eyestalk ganglia transcriptomes (Supplemental Table 1). Translation of this collective set of sequences revealed ten distinct isoforms of CRY1 (Table 4, Fig. 9 and Supplemental Fig. 6K1–9); eight of these are full-length proteins (Homam-CRY1-v1 through v8), one is an N-terminal partial protein (Homam-CRY1-v9), and one is a C-terminal partial sequence (Homam-CRY1-v10). A single *cry1* gene is predicted for *Homarus*, with isoform diversity likely derived largely from alternative splicing of that gene. BLAST searches of the extant *D. plexippus* proteins curated at NCBI confirmed that CRY1, the original query protein, is the most similar *Danaus* sequence to all ten of the *Homarus* CRY1s (Supplemental Table 3). An insect isoform of CRY1 was identified as the most similar non-redundant arthropod protein to each of the *Homarus* sequences (Supplemental Table 4), *e.g.*, an isoform from the Western tarnished plant bug *Lygus hesperus* (Lyghe-CRY1 [Fig. 9]; Accession No. JAQ08121; Tassone et al., 2016) for Homam-CRY-v1. Pfam analyses identified one DNA photolyase domain and one FAD binding domain of DNA photolyase in Homam-CRY1-v1, v2 and v3 (Fig. 9 and Supplemental Table 5), a domain complement also predicted by the program for Lyghe-CRY1 (Fig. 9); one or the other of these domains, but not both, were identified by Pfam in the other *Homarus* CRY1 variants (Supplemental Table 5).

3.3.4. Clock output pathway proteins

In *D. melanogaster*, the peptide hormone pigment dispersing factor (PDF), known in crustaceans as pigment dispersing hormone (PDH) (Christie et al., 2010), is produced/released by a subset of clock neurons, and is one example of a *Drosophila* clock output pathway protein

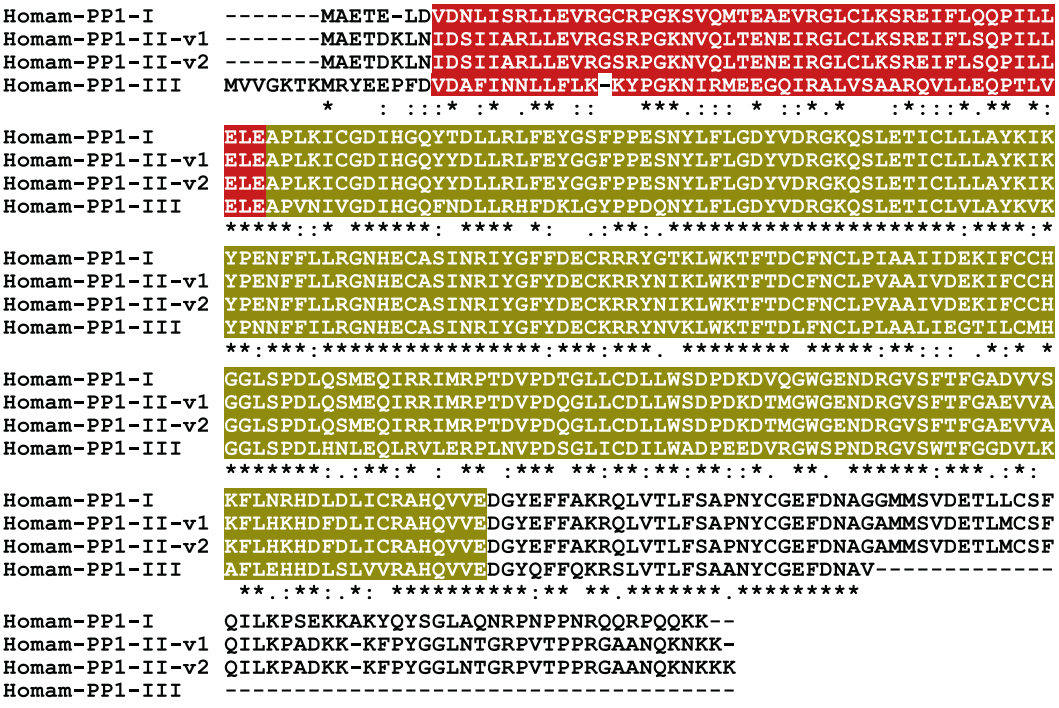


Fig. 7. MAFFT alignment of *Homarus americanus* protein phosphatase 1 (Homam-PP1) proteins and their variants. In the line immediately below each sequence grouping, "*" indicates identical amino acid residues, while ":" and "." denote amino acids that are similar in structure between sequences. In this figure, serine-threonine protein phosphatase N-terminal and calcineurin-like phosphoesterase domains identified by Pfam analyses are highlighted in dark red and dark yellow, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

(Shafer and Yao, 2014). This peptide, in concert with its G-protein-coupled receptor, pigment dispersing factor receptor (PDFR), plays a key role both in synchronizing activity among clock neurons and in controlling the relative daily balance of circadian activity in the *Drosophila* clock (Shafer and Yao, 2014).

BLAST searches using either a known *H. americanus* PDH precursor or a *D. melanogaster* PDFR protein as the query sequence identified transcripts encoding putative homologs of these two protein families in both the *H. americanus* brain- and eyestalk ganglia-specific transcriptomes (Supplemental Table 1). Translation of the collective set of

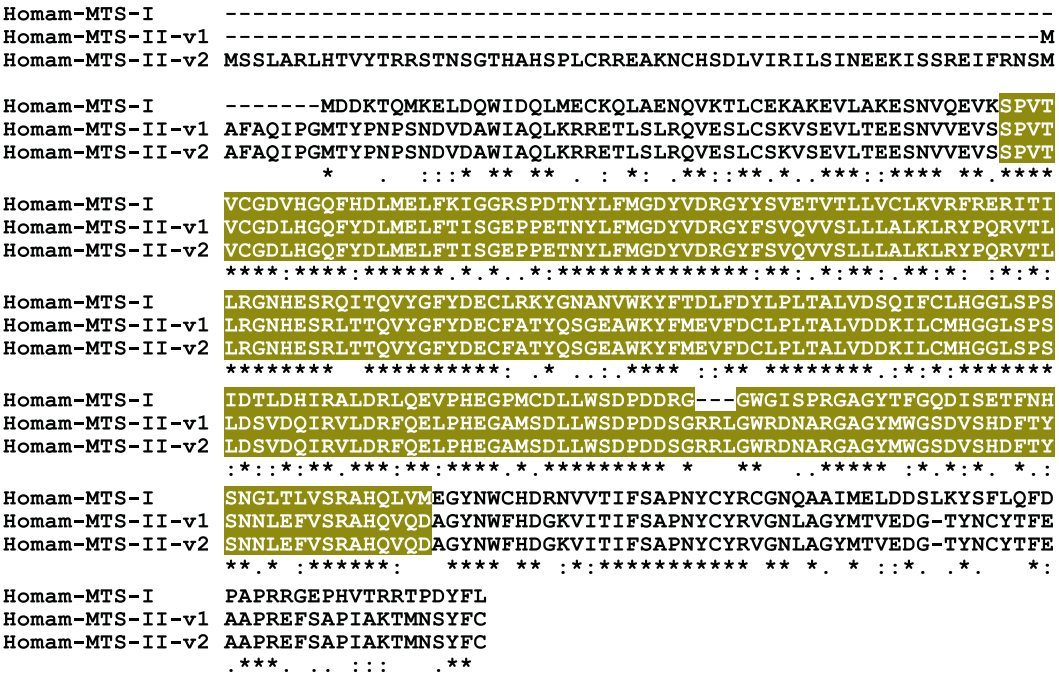


Fig. 8. MAFFT alignment of *Homarus americanus* protein phosphatase 2A catalytic subunit microtubule star (Homam-MTS) proteins and their variants. In the line immediately below each sequence grouping, "*" indicates identical amino acid residues, while ":" and "." denote amino acids that are similar in structure between sequences. In this figure, calcineurin-like phosphoesterase domains identified by Pfam analyses are highlighted in dark yellow. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

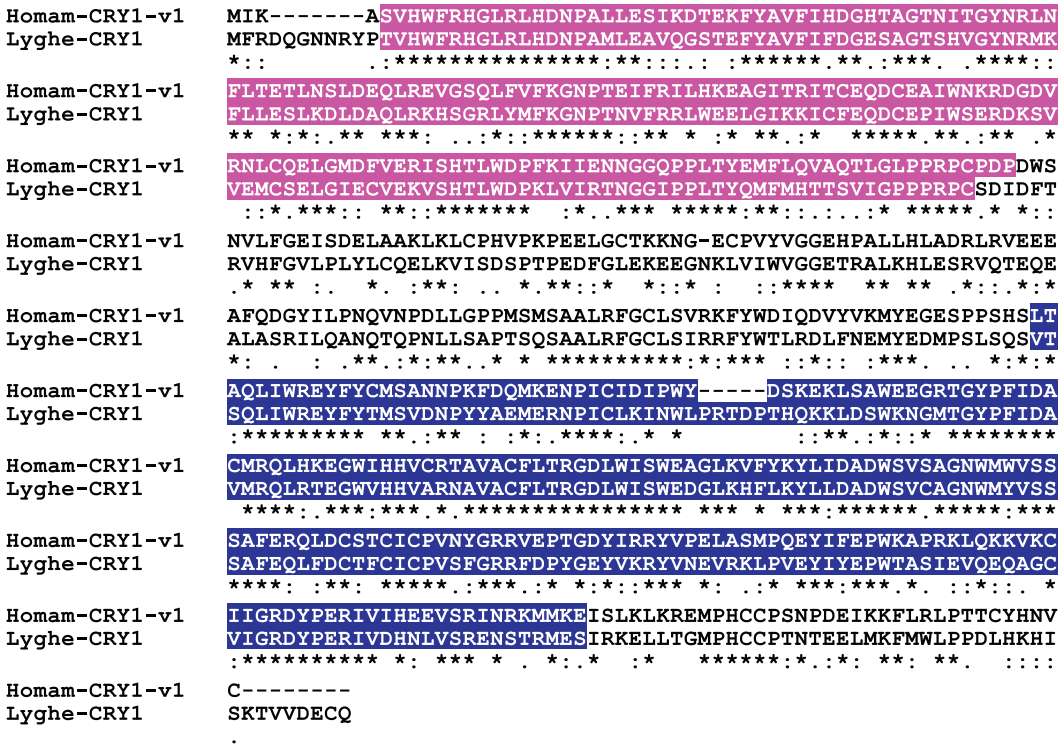


Fig. 9. MAFFT alignment of *Homarus americanus* cryptochrome 1 variant 1 (Homam-CRY1-v1) and *Lygus hesperus* cryptochrome 1 (Lyghe-CRY1). In the line immediately below each sequence grouping, "*" indicates identical amino acid residues, while ":" and "." denote amino acids that are similar in structure between sequences. In this figure, DNA photolyase domains and FAD binding domains of DNA photolyase identified by Pfam analyses are highlighted in pink and blue, respectively. Lyghe-CRY1 (Accession No. JAQ08121; Tassone et al., 2016). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

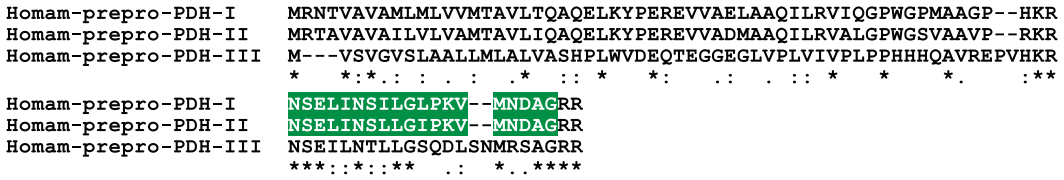


Fig. 10. MAFFT alignments of *Homarus americanus* pigment dispersing hormone precursor (Prepro-PDH) proteins. In the line immediately below each sequence grouping, "*" indicates identical amino acid residues, while ":" and "." denote amino acids that are similar in structure between sequences. In this figure, pigment dispersing hormone domains identified by Pfam analyses are highlighted in dark green. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

putative PDH precursor-encoding transcripts revealed three distinct prepro-PDHs (Homam-prepro-PDH-I, Homam-prepro-PDH-II and Homam-prepro-PDH-III; Table 4 and Fig. 10), each a full-length protein and likely the product of a different gene. Two of the PDH precursors, Homam-prepro-PDH-I and Homam-prepro-PDH-II, are preprohormones previously identified from the *H. americanus* eyestalk transcriptome (Christie et al., 2017); Homam-prepro-PDH-III is described here for the first time. The mature PDH isoforms NSELINSILGLPKVMNDAamide and NSELINSLLGIPKVMNDAamide were predicted from Homam-prepro-PDH-I and II, respectively; both of these PDHs are known *H. americanus* isoforms (Christie et al., 2015, 2017; Ma et al., 2008). The putative mature PDH isoform predicted from Homam-prepro-PDH-III, NSEILNTLLGSQDLSNMRSAamide, is a novel lobster neuropeptide that diverges from typical decapod PDHs in that it is 20 rather than 18 amino acids in length (for review see Christie et al., 2010).

Translation of the collective set of PDHR-encoding sequences revealed four distinct receptors, which are likely the products of two different genes (Homam-PDHR-I-v1 through v3 and Homam-PDHR-II; Table 4, Fig. 11 and Supplemental Fig. 6L). Homam-PDHR-I-v1, Homam-PDHR-I-v2 and Homam-PDHR-II are full-length proteins, while Homam-PDHR-I-v3 is an N-terminal partial sequence (Table 4). Homam-PDHR-I-v1 (Supplemental Fig. 6 L1) is identical in sequence to a previously described *H. americanus* PDHR, *i.e.* Homam-PDHR-I (Christie et al., 2015). Homam-PDHR-I-v2 (Fig. 11) is identical in sequence to Homam-PDHR-I-v1, except for the presence of a 26 amino acid N-terminal extension, while Homam-PDHR-I-v3 (Supplemental Fig. 6L2) is, with the exception of a missing methionine residue, identical in sequence to the N-terminus of Homam-PDHR-I-v2. Homam-PDHR-II (Supplemental Fig. 6L3) is also a previously described *H. americanus* protein (Christie et al., 2015).

Searches of FlyBase confirmed that isoforms of PDF precursor or PDHR are the most similar *Drosophila* proteins to each of the above-mentioned lobster prepro-PDHs or PDHRs, respectively (Supplemental Table 3). Similarly, a PDH precursor or a PDHR was returned as the top BLAST hit when each of the putative lobster clock output pathway proteins was used to query the non-redundant arthropod protein dataset curated in NCBI (Supplemental Table 4). For example, neuropeptide G-protein coupled receptor B2, a synonym for PDHR, from the brown planthopper *Nilaparvata lugens* (Nillu-GPCR-B2 [Fig. 11]; Accession No. BAO01102; Tanaka et al., 2014) was identified as the non-

```
Homam-PDHR-I-v2 MMVT-------------------------PCSVIN-------------------------
Nillu-GPCR-B2   MPVRRVSSDGNRDPLVAHRASLLAVPGLHPTSGVGSDVGSCRQLDLSLTSPADTTAAAMS
                * *                           * * :.

Homam-PDHR-I-v2 ---TPVC-----------------------------------------EPEDHPLSQS-
Nillu-GPCR-B2   DDISPECVSPAVLLVPPPYPEASDEVSSTSKSDQGIQITVVEDTTEDDKQPQSQPQPQSL
                   :* *                                        :*:.:* .**

Homam-PDHR-I-v2 --PPMTYSIYLQACVDNYEHVNLSTHESWCNATWDLVLCWPPTPAGGSTRLPCPPVKGVD
Nillu-GPCR-B2   HHRPVANKRSLELCAARYHEYSAPLGVLYCNWTWDQVLCWPPTRAGTTATQRCPRDKGID
                   *:: .  *: *. .*.. . .    :** *** ******* ** ::   **  **:*

Homam-PDHR-I-v2 PSKYVYKHCDGSGRWTGKTPG-DFTLPQGWTNYTVCFTKAIQEIMQELYKQSDEDAQTKL
Nillu-GPCR-B2   PTKFATKKCSAEGRWEGKHPGEEATSPQGWTNYTPCYTPEMLELFRKLYVGSEKAAMKKL
                *:*:. *:*...*** ** ** : * ******** *:*  : *::::**  *:: * .**

Homam-PDHR-I-v2 NIALGTRIMEIVGLSLSLASLCASLAIFFHFRSLKNNRTRIHRNLFVAMVIQVMIRLVLY
Nillu-GPCR-B2   EIAERTRTLEIVGFSISLAALLISLAIFCHFRSLRNNRTRIHKNLFVAMVIQVVIRLTLY
                :**  ** :****:*:***:*  ***** ***** :*******:**********:***.**

Homam-PDHR-I-v2 IDQAIIR-----GHIVGHSPTNTNAPRQGIDNTPMLCEASYVLLEYARTAMFMWMFIEGL
Nillu-GPCR-B2   IDQALVKGNSLTGQRVSSSSSSSGGFRQGIDNTPVLCEASYVLLEYARTAMFMWMFIEGL
                ****:::      *: *. *.:.:.. ********:************************

Homam-PDHR-I-v2 YLHNKITVTVFHHKFYYSVYHAVGWGVPVLLTAAWATATAMHYGNSRCWWGYNFTSYFWI
Nillu-GPCR-B2   YLHNMVTVTVFQEKSYYAAYTMVGWGVPVLMTAAWAMTTASTLGSTRCWWGYNLNKFFWI
                **** :*****:.* **:.*  ********:***** :**   *.:*******:..:***

Homam-PDHR-I-v2 LEGPRFSVISMNMVFLLNIIRVLVTKLRQSNSSEALQVRKAVKAAIVLLPLLGITNVLNM
Nillu-GPCR-B2   LEGPRLIFILMNFLFLLNIIRVLVVKLRQSHTSEIEQVRKAVRAAVVLLPLLGITNLVNM
                *****: .* **::**********.*****::**  ******:**:**********::**

Homam-PDHR-I-v2 VAAPLGRSAAEFGLWSYATHFLTSFQGFFIALLYCFLNGEVRTAVRKYVDNYLLHRSAGV
Nillu-GPCR-B2   TEAPLTRAVWEFGLWSYTTHFLTSFQGLFIAILYCFLNGEVRMAVRKSVSTYLSLRPQHF
                . *** *:. *******:*********:***:********** **** *..**  *.  .

Homam-PDHR-I-v2 ---RRGSGLSSVFLT---TVTDLPRDRHGRTRH-LCACLRGNHSPTHGSYL
Nillu-GPCR-B2   TPQRRNSAFVSANMTEQVVSTPPPRPATSAPRSWFRSCCSHPSQPTPETRV
                   **.*.: *. :*   . *  **   . .*  : :*     .**  : :
```

Fig. 11. MAFFT alignment of *Homarus americanus* pigment dispersing hormone receptor I variant 2 (Homam-PDHR-I-v2) and *Nilaparvata lugens* neuropeptide G-protein coupled receptor B2 (Nillu-GPCR-B2), a synonym for the PDHR. In the line immediately below each sequence grouping, "*" indicates identical amino acid residues, while ":" and "." denote amino acids that are similar in structure between sequences. In this figure, hormone receptor and secretin family seven-transmembrane receptor domains identified by Pfam analyses are highlighted in gray and black, respectively. Nillu-GPCR-B2 (Accession No. BAO01102; Tanaka et al., 2014).

redundant arthropod protein most similar to each of the Homam-PDHR-I variants. Finally, Pfam analyses identified a pigment dispersing hormone domain in both Homam-prepro-PDH I and II, but not in Homam-prepro-PDH III (Fig. 10 and Supplemental Table 5); it is likely that the two additional amino acid residues in the PDH derived from Homam-prepro-PDH III are responsible for the lack of identification of the PDH domain in this protein. Pfam analyses identified one hormone receptor and one secretin family seven-transmembrane receptor domain in each of the lobster PDHRs (Fig. 11 and Supplemental Table 5). Pfam identified the same complements of domains in the top non-redundant arthropod protein hit for each PDH precursor or receptor sequence, *e.g.*, Nillu-GPCR-B2 (Fig. 11).

3.4. Confirmation and extension of core clock transcript sequences

To provide increased confidence in the presence of the core clock proteins (*i.e.*, CLK, CRY2, CYC, PER and TIM) in both the brain and eyestalk ganglia, ~ 500 bp fragments of the transcripts encoding them were amplified from multiple brain and eyestalk cDNA samples ($n = 3$ samples for each CNS region; Fig. 12); all of the amplified PCR products exhibited > 99% identity to their respective transcriptome-derived sequences. Unexpectedly, variants of CLK and CYC were also amplified from both templates, albeit with much lower efficiency, on some, but not all, tissue collection dates (Fig. 12). Sequence analysis of the respective amplimers suggested intron retention-based alternative splicing, as both products contained insertions (CYC, 99 nucleotides; CLK, 225 nucleotides) that introduced premature stop codons. The generation of unproductive alternative splice variants is a common phenomenon in plant circadian genes (Filichkin and Mockler, 2012), and may reflect nonsense-mediated decay within the circadian system (Staiger et al., 2013) and/or stress responses (Pegoraro et al., 2014; Kwon et al., 2014).

To further confirm the transcripts identified from the transcriptome assemblies, the longest full-length open reading frames for the *Homarus* CRY2-, CYC- and PER-encoding transcripts were amplified from brain/eyestalk cDNAs. For all three, nucleotide sequence identity between the

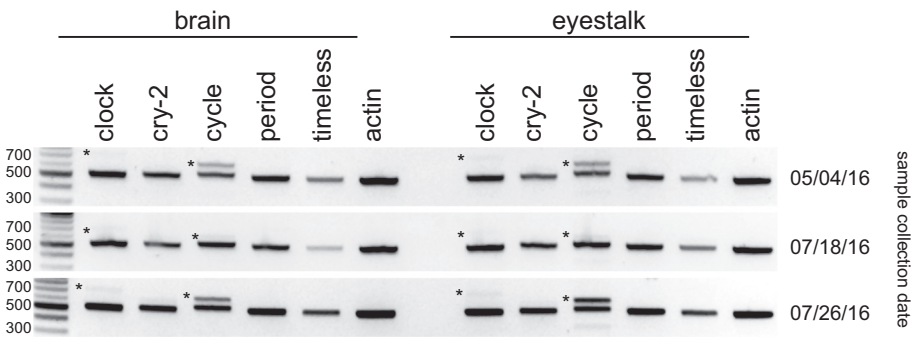


Fig. 12. Expression profile of *Homarus americanus* core clock genes in brain and eyestalk ganglia. Fragments, ~ 500 base pairs (bp) of the five core clock genes, clock, cryptochrome 2 (CRY2), cycle, period, and timeless, were amplified from cDNAs corresponding to brain and eyestalk ganglia isolated from lobsters collected on the indicated dates. Asterisks correspond to nonsense splice variants containing premature stop codons presumably generated *via* intron retention. A 500-bp fragment of *H. americanus* actin was amplified as a positive control for cDNA template quality. For enhanced band clarity, the negative of the gel image is shown.

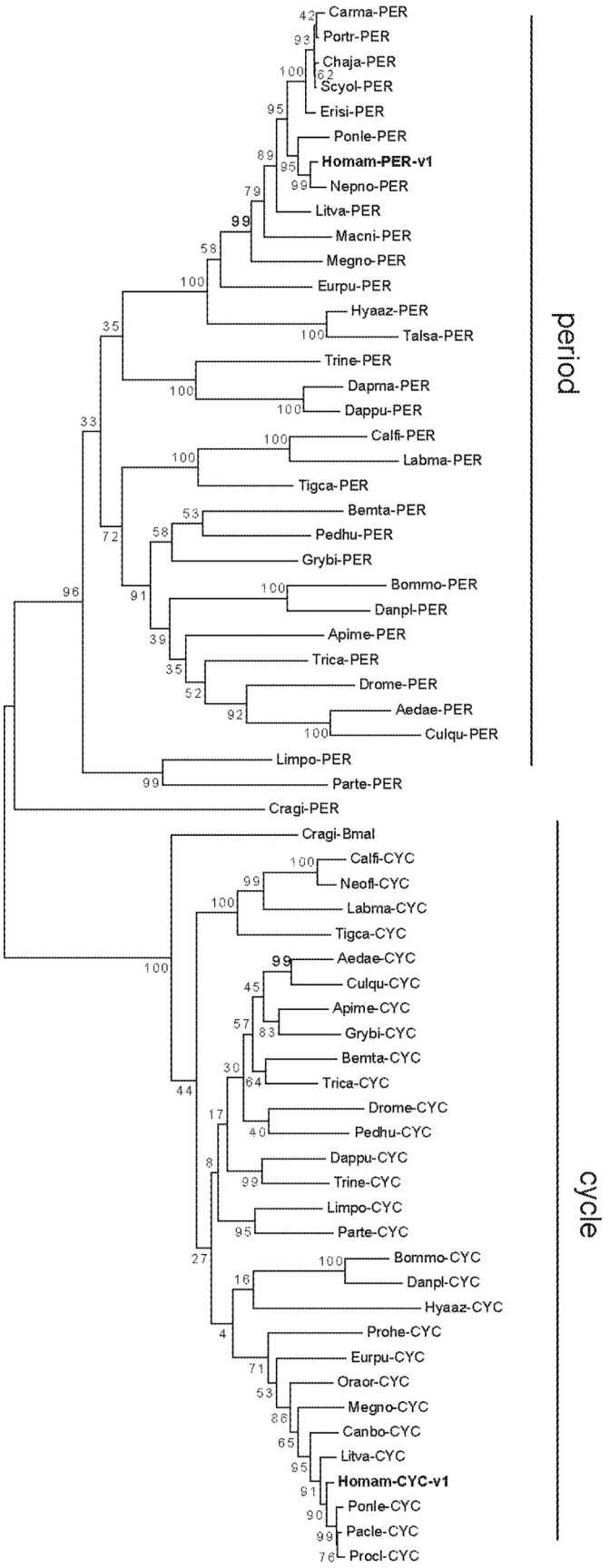
consensus sequences and the transcriptomic data exceeded 99%. These cloned sequences were submitted to GenBank under Accession Nos. MG280773 (CRY2), MG280774 (CYC) and MG280776 (PER-v1).

Because the assembly data suggested that the *in silico Homarus* TIM transcripts corresponded to fragments truncated at their 5′ ends (*i.e.*, the lack of an upstream stop codon to facilitate identification of the first in-frame start methionine), RACE methods were used to complete the coding sequences. Two variants, differentiated by alternative exon usage, were identified. One variant is 99.7% identical at the nucleotide level to the transcriptomic assembly sequence encoding Homam-TIM-v1 and is predicted to encode a 1344-amino acid protein with a start site that is present in the assembly sequence (Supplemental Fig. 5B); this cloned sequence was deposited in GenBank under Accession No. MG280777. The second cloned variant (Accession No. MG280778) utilizes an alternative initiation exon that is 133 nucleotides shorter than that encoding Homam-TIM-v1, resulting in a predicted protein of 1300 amino acids (Supplemental Fig. 5C), which corresponds to Homam-TIM-v2. Intriguingly, the first 103 nucleotides of the initiation exon are identical to a transcribed sequence (Accession No. GEBG01019623) from a different *H. americanus* transcriptome assembly (BioProject No. PRJNA300643; Northcutt et al., 2016); however, GEBG01019623 lacks the 132-nucleotide deletion observed in MG280778, thus resulting in a slightly different deduced protein (Supplemental Fig. 5C). The use of alternative translational sites for core clock genes is not without precedence, as *Drosophila* utilizes a similar process (Rosato et al., 1997) to generate the shorter TIM isoform associated with seasonal diapause (Tauber et al., 2007). In addition to these two variants, a fragment of a *Homarus* TIM-encoding transcript with a 94-nucleotide insertion at position 3909 was also amplified (data not shown); this insertion changes the downstream reading frame and removes the previous stop codon.

The transcriptome-derived *H. americanus* CLK-encoding transcript is a partial sequence (1703 nucleotides) that lacks a stop codon and encodes a protein that is likely truncated by > 100 amino acids. Despite repeated efforts, RACE methods to complete the 3′ end of the transcript failed to yield a CLK-like sequence. Transcriptomic data for CLK-encoding transcripts in other crustacean species likewise represent partial sequences, with most truncated 5′ to the characteristic poly-glutamine stretch associated with transcriptional activation. We speculate that the inability to generate full-length clock sequences for some species may be inherent to the CAA/CAG trinucleotide repeats that comprise the poly-glutamine stretch and contribute to the formation of nucleic acid hairpin structures that can limit both PCR amplification and sequence library construction (Nelms and Labosky, 2011). Nonetheless, a significant portion of the transcriptome-derived CLK-encoding transcript (nucleotides 1–1397) was validated by conventional PCR with > 99% sequence identity (Accession No. MG280775).

3.5. *Homarus PER and CYC cluster appropriately in an arthropod cladogram that includes members of both protein families*

As described earlier, reciprocal BLAST searches of FlyBase using the *H. americanus* PER proteins identified here as query sequences returned *D. melanogaster* PER as the top BLAST hit for only four of the 12 isoforms, *i.e.*, Homam-PER-v4, Homam-PER-v8a/b and Homam-PER-v9. The remaining eight *Homarus* PERs returned *Drosophila* CYC as the top hit, even though all twelve of the lobster proteins appear to be derived from the same gene. Alignment of the full set of *H. americanus* PER sequences (Supplemental Fig. 4) revealed that the four that return PER as the top *Drosophila* hit all possess a common 16 amino acid deletion, namely residues 639–654 (DGKSSTHPAFSLCGGS) in Homam-PER-v1, a stretch of amino acids that is present in all of the sequences that return *Drosophila* CYC as the top hit. The DGKSSTHPAFSLCGGS insertion/deletion is responsible for the lobster proteins returning PER or CYC as the top FlyBase hit, as when the sequence is inserted into the proteins that are missing it, they switch from returning PER as the top *D.*



(caption on next page)

Fig. 13. Maximum-likelihood tree of putative period (PER) and cycle (CYC) sequences from diverse arthropods. The tree was generated using the Jones-Taylor-Thornton matrix-based substitution model (Jones et al., 1992) with bootstrap support for 1000 iterations shown at the nodes. The longest *Homarus americanus* PER variant, *Homarus americanus* PER variant 1 (Homam-PER-v1; shown in bold) sorted with other PER proteins despite the presence of a 16-amino acid CYC-related sequence (see Section *3.5*). Similarly, the longest *H. americanus* CYC variant, *H. americanus* CYC variant 1 (Homam-CYC-v1; shown in bold) sorted with CYC proteins. The PER and CYC sequences used to generate the cladogram, and their associated accession numbers, can be found in Supplemental Fig. 1. Crustacean species abbreviations: Calfi, *Calanus finmarchicus*; Canbo, *Cancer borealis*; Chaja, *Charybdis japonica*; Carma, *Carcinus maenas*; Dapma, *Daphnia magna*; Dappu, *Daphnia pulex*; Erisi, *Eriocheir sinensis*; Eurpu, *Eurydice pulchra*; Homam, *Homarus americanus*; Hyaaz, *Hyalella azteca*; Labma, *Labidocera maderae*; Litva, *Litopenaeus vannamei*; Macni, *Macrobrachium nipponense*; Megno, *Meganyctiphanes norvegica*; Neofl, *Neocalanus flemingeri*; Nepno, *Nephrops norvegicus*; Oraor, *Oratosquilla oratoria*; Pacle, *Pacifastacus leniusculus*; Ponle, *Pontastacus leptodactylus*; Portr, *Portunus trituberculatus*; Prohe, *Proasellus hercegovinensis*; Procl, *Procambarus clarkii*; Scyol, *Scylla olivacea*; Talsa, *Talitrus saltator*; Tigca, *Tigriopus californicus*; Trine, *Triops newberryi*. Hexapod species abbreviations: Acypi, *Acyrthosiphon pisum*; Aedae, *Aedes aegypti*; Apime, *Apis mellifera*; Bemta, *Bemisia tabaci*; Bommo, *Bombyx mori*; Culqu, *Culex quinquefasciatus*; Danpl, *Danaus plexippus plexippus*; Drome, *Drosophila melanogaster*; Grybi, *Gryllus bimaculatus*; Pedhu, *Pediculus humanus corporis*; Trica, *Tribolium castaneum*. Chelicerate species abbreviations: Limpo, *Limulus polyphemus*; Parte, *Parasteatoda tepidariorum*. Outgroup species abbreviation: Cragi, *Crassostrea gigas*.

melanogaster hit to returning CYC as the most similar *Drosophila* protein. Likewise, when the insertion is removed from the *H. americanus* PERs that possess it, they return a PER rather than CYC as their top FlyBase hit.

To conclusively show that the *H. americanus* proteins reported here as PER truly cluster with members of the PER family rather than with CYC proteins, a cladogram that includes the two protein groups was constructed (Fig. 13). This cladogram includes Homam-PER-v1 (the longest lobster PER isoform and one that includes the 16 amino acid insertion described above), Homam-CYC, Drome-PER-A (Accession No. AAF45804 [Adams et al., 2000]) and Drome-CYC (Accession No. AAF49107 [Adams et al., 2000]), as well as a number of other arthropod PER and CYC proteins, both previously described and identified here for the first time from publicly accessible transcriptome shotgun assembly data (see Supplemental Fig. 1); PER and Bmal from the mollusk *C. gigas* (Accession Nos. AQM57604 and AQM57603, respectively; Perrigault and Tran, 2017) were used as outgroups. Despite the presence of the 16 CYC-related amino acids, Homam-PER-v1 sorted to the PER clade and clustered with high bootstrap support among other decapod PER sequences (Fig. 13). Interestingly, other Astacidae infraorder sequences comprising the phylogenetic branch that includes Homam-PER-v1 contained highly similar insertions (DGKSSYHPAFSLCDGS in *Pontastacus leptodactylus* and DGKSSTHPAFSLSAGS in *Nephrops norvegicus*), as did the more evolutionary distant decapod *Litopenaeus vannamei* (DGKSSTHPAFSLSAGS). The sister branch, comprised of sequences from the infraorder Brachyura, lacked the insertion. No sequences annotated as PER aligned to the CYC clade and *vice versa*.

Maximum-likelihood phylogenetic analyses using CRY1 and CRY2 sequences from various arthropod species, both previously described and identified here for the first time (Supplemental Fig. 2), confirmed the initial *Homarus* CRY annotations, with the lobster sequences sorting to the respective CRY1 and CRY2 clades (Supplemental Fig. 7); as was the case for the PER/CYC cladogram, *C. gigas* CRY1 and CRY2 sequences (Accession Nos. ANJ02841 [Mat et al., 2016] and AQM57602 [Perrigault and Tran, 2017], respectively) were used as outgroups.

3.6. Shared and region-specific circadian protein variants

Although many of the circadian proteins described here were identified from both the brain and eyestalk ganglia transcriptomes, and hence are likely present in both portions of the CNS, assembly-, and possibly region-specific isoforms of some protein families (including core clock, clock-associated, clock input pathway and clock output pathway components) were also identified (Table 4). For example, while both the brain and eyestalk ganglia share the same CLK and CRY2, CYC and PER have both shared and assembly-specific variants. Specifically, Homam-CYC-v1 was deduced from both brain and eyestalk transcripts, while Homam-CYC-v2 was predicted only from the eyestalk transcriptome. Similarly, Homam-PER-v1, v3, v6 and v7 were deduced from both assemblies, while Homam-PER-v2 and v8a appear to be brain-specific, and Homam-PER-v4, v5, v8b, v9, v10 and v11 were predicted only from the eyestalk ganglia. For TIM, although transcripts were identified from both the brain and eyestalk ganglia, the deduced proteins are all assembly-specific, with Homam-TIM-v1a and v2 predicted only from the brain, and Homam-TIM-v1b identified only from the eyestalk ganglia. Assembly-specific variants were also found for a number of the clock-associated proteins, *e.g.*, CWO, DBT and JET, as well as for the clock input pathway protein CRY1 and the clock output pathway proteins PDH precursor and PDHR (Table 4).

The effects, if any, that region-specific circadian protein variants may have on clock function remain unknown. For some proteins, *e.g.*, a number of the CRY1 isoforms, the lack of a subset of standard functional domains suggests that they are likely non-functional. However, for others, *e.g.*, PER and TIM, common sets of functional domains were identified; in these cases, it is possible that the different isoforms are functionally distinct, as has been reported for circadian protein splice variants in other species, *e.g.*, TIM in *D. melanogaster* (Montelli et al., 2015) and Bmal (CYC) in humans (Schoenhard et al., 2002). Given the differential distribution of variants between the brain and eyestalk ganglia assemblies, region-specific functional variation in the brain and eyestalk clocks is a possibility, especially if the two CNS regions have independent timekeeping systems rather than a single distributed one.

3.7. The previously reported PER-like immunoreactivity in the eyestalk is unlikely due to the PER proteins predicted here

A previous study used PER Western blotting to identify the putative location of the central circadian pacemaker in *H. americanus* (Grabek and Chabot, 2012). Specifically, eyestalk ganglia- and brain-derived protein extracts were analyzed, with PER immunoreactivity detected only in the eyestalk ganglia (Grabek and Chabot, 2012). These findings were interpreted as indicating that the eyestalk, and not the brain, is the locus of the central clock in *H. americanus*. Interestingly, the size of the putative PER protein detected *via* Western blotting in the eyestalk ganglia was smaller than that of most other arthropod PERs; *i.e.*, it was ~70 kDa, while most other identified arthropod PER isoforms are in the > 100^{+} kDa range (Grabek and Chabot, 2012). The small size of the protein detected in the eyestalk raises the question of whether or not the protein detected by Western analyses is actually PER.

Our mining of the *H. americanus* brain and eyestalk ganglia transcriptomes identified PER-encoding transcripts in both assemblies (Supplemental Table 1), strongly suggesting that PER is produced in both regions of the lobster CNS. Multiple full-length PER variants were found in both the brain and eyestalk ganglia, with some variants present in both locations, while others are assembly-specific (Table 4). The variants are characterized by the presence/absence of insertions/deletions in four segments (referred to as variable regions 1–4) of the PER-encoding transcript (Supplemental Fig. 4). To examine the transcriptional relevance of the variants, the respective variable regions were amplified from brain and eyestalk ganglia cDNAs; clear differences in amplimer sizes, consistent with the observed transcriptomic deletions, were found (Supplemental Fig. 8). Sequence analyses of the cloned amplimers confirmed alternative splicing of variable regions 1, 3, and 4; no products corresponding to the deletion in variable region 2 were identified, suggesting that transcripts with this splicing pattern may be

of low abundance or are differentially regulated.

The largest of the *H. americanus* PER variants, Homam-PER-v1, is 1286 amino acids in length, while the smallest, Homam-PER-v11, is 1235 amino acids long; both of these proteins are predicted to exist in the eyestalk ganglia. Based solely on amino acid sequence, the molecular masses of lobster PERs are calculated to range from 135.51 kDa (Homam-PER-v11) to 141.39 kDa (Homam-PER-v1), which is about double that of the protein detected *via* Western blotting by Grabek and Chabot (2012). Thus, we believe that the immunoreactivity reported by Grabek and Chabot (2012) is not truly a detection of PER, but is rather non-specific interaction with a non-PER, ~636 amino acid protein, and that the hypothesis that the eyestalk ganglia, not the brain, are the locus of the *Homarus* central circadian clock, is incorrect.

3.8. All expected components of a putative "ancestral-type" arthropod circadian signaling system are present in both the Homarus brain and eyestalk ganglia

In the study presented here, transcripts encoding putative homologs of 21 different circadian protein families were searched for in region-specific transcriptomes generated from *H. americanus* brain or eyestalk ganglia. The targeted protein groups included five core clock proteins (CLK, CRY2, CYC, PER and TIM), thirteen clock-associated proteins (CKIIα, CKIIβ, CWO, DBT, JET, PDP1, PP1, MTS, TWS, WDB, SGG, SLIMB and VRI), one clock input pathway protein (CRY1) and two clock output pathway proteins (PDH and PDHR). For each of the targeted groups, transcripts encoding putative family members were found in both the brain and eyestalk ganglia transcriptomes, providing the first description of the molecular components of putative circadian signaling systems in *H. americanus*, and suggesting that an intrinsic timekeeping system is potentially present in both regions of the lobster CNS. These *Homarus* genes/proteins represent the first "complete" description of the molecular components of a circadian signaling system from any decapod crustacean, and more generally represent one of just a handful of full circadian gene/protein complements described from any member of the Crustacea.

In insects, several different models of circadian clock systems have been proposed, based primarily on the fact that one or more of the core clock/clock input pathway proteins, CRY2, TIM or CRY1, have been lost in a number of species (Yuan et al., 2007). In the "ancestral-type" clock model, which is proposed for species such as *D. plexippus*, in which all three proteins are present, CRY1 serves as an ultraviolet/blue light photoreceptor, which entrains the clock to the solar day, and CRY2 acts as a transcriptional repressor in the core clock feedback loop. In *Drosophila*, however, CRY2 has been lost, with the result that CRY2-mediated transcriptional repression has been taken over by PER. As in the ancestral clock, CRY1 serves to detect light and synchronize the clock with the solar day in this model. In a third group of insects, which includes both the beetle *Tribolium* and the bee *Apis*, CRY1 has been lost, but CRY2 is retained. The clockwork model proposed for these species suggests that CRY2 serves as a transcriptional repressor, as it does in the ancestral model. The protein that is involved in light detection is less clear, and may involve retinal or extraretinal opsins. In *Tribolium*, it is thought that the light input may be mediated through TIM. In *Apis*, however, TIM also appears to have been lost, so the mechanisms by which photic input gains access to the clock remain murky (Yuan et al., 2007).

Although the number of crustacean species examined to date is relatively small, and the details of clock function have not been studied in these species, all of those that have been examined using genomic or transcriptomic methods appear to possess both CRY1 and CRY2, as well as TIM (Christie et al., 2013, 2018; Nesbit and Christie, 2014; O'Grady et al., 2016; Roncalli et al., 2017; Sbragaglia et al., 2015; Tilden et al., 2011). Our transcriptomic data for both the brain and the eyestalk ganglia of *H. americanus* indicate that transcripts encoding all three proteins, CRY1, CRY2, and TIM, are present in both of these regions, suggesting that an "ancestral-type" clock is likely in this species, as it appears to be in other crustaceans.

4. Summary/conclusions

In the study reported here, region-specific transcriptomes were generated and searched for sequences encoding putative circadian proteins in order to determine if the brain, eyestalk ganglia or both regions of the CNS are likely to possess an intrinsic clock system. Transcripts encoding putative homologs of all of the core clock, clock-associated, clock input pathway and clock output pathway proteins searched for were identified in both the brain and eyestalk ganglia assemblies, suggesting that both regions of the nervous system are likely to possess intrinsic circadian signaling systems. As TIM, CRY1 and CRY2 were identified in *H. americanus*, this species appears to possess a typical, ancestral-type core clock. Although most of the circadian signaling system protein isoforms identified here appear to be present in both the brain and eyestalk ganglia, there is evidence for several assembly-specific variants, including both PER and TIM, suggesting the possibility of region-specific variation in clock function in the brain and eyestalk, especially if distinct pacemakers are present in these two regions of the CNS. The data presented here now provide a means for initiating functional studies of circadian signaling in *H. americanus*, including determining the diel cycling patterns of clock genes in the brain and eyestalk ganglia, the location and identity of clock neurons in these two portions of the CNS, and whether the brain and eyestalk clocks represent two distinct or one distributed timekeeping system.

Acknowledgements

The authors would like to thank the Georgia Genomics Facility for cDNA library production and Illumina sequencing services, and National Center for Genome Analysis Support for access to their Mason Linux cluster (supported by National Science Foundation [NSF] ABI-1062432 to Indiana University), which was used for assembling the transcriptome reported here. Dr. Petra Lenz (University of Hawaii at Manoa) is thanked for her helpful input on the design of the sequencing and transcriptome assembly experiments. 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 U. S. Department of Agriculture. USDA is an equal opportunity provider and employer.

Funding

This work was supported by the National Science Foundation [grant numbers IOS-1353023, IOS-1354567 and OCE-1459235]; the National Institutes of Health [grant numbers 5P20RR016463-12 and 8P20GM103423-12]; base CRIS funding [Project #2020-22620-022-00D]; the Cades Foundation of Honolulu, Hawaii; the University of Hawaii at Manoa's Undergraduate Research Opportunities Program; the American Physiological Society Undergraduate Research Fellowship Program; and the Henry L. and Grace Doherty Charitable Foundation.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.margen.2018.03.002.

References

Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Scherer, S.E., Li, P.W., Hoskins, R.A., Galle, R.F., George, R.A., Lewis, S.E., Richards, S., Ashburner, M., Henderson, S.N., Sutton, G.G., Wortman, J.R., Yandell, M.D., Zhang, Q., Chen, L.X., Brandon, R.C., Rogers, Y.H., Blazej, R.G., Champe, M., Pfeiffer,

B.D., Wan, K.H., Doyle, C., Baxter, E.G., Helt, G., Nelson, C.R., Gabor, G.L., Abril, J.F., Agbayani, A., An, H.J., Andrews-Pfannkoch, C., Baldwin, D., Ballew, R.M., Basu, A., Baxendale, J., Bayraktaroglu, L., Beasley, E.M., Beeson, K.Y., Benos, P.V., Berman, B.P., Bhandari, D., Bolshakov, S., Borkova, D., Botchan, M.R., Bouck, J., Brokstein, P., Brottier, P., Burtis, K.C., Busam, D.A., Butler, H., Cadieu, E., Center, A., Chandra, I., Cherry, J.M., Cawley, S., Dahlke, C., Davenport, L.B., Davies, P., de Pablos, B., Delcher, A., Deng, Z., Mays, A.D., Dew, I., Dietz, S.M., Dodson, K., Doup, L.E., Downes, M., Dugan-Rocha, S., Dunkov, B.C., Dunn, P., Durbin, K.J., Evangelista, C.C., Ferraz, C., Ferriera, S., Fleischmann, W., Fosler, C., Gabrielian, A.E., Garg, N.S., Gelbart, W.M., Glasser, K., Glodek, A., Gong, F., Gorrell, J.H., Gu, Z., Guan, P., Harris, M., Harris, N.L., Harvey, D., Heiman, T.J., Hernandez, J.R., Houck, J., Hostin, D., Houston, K.A., Howland, T.J., Wei, M.H., Ibegwam, C., Jalali, M., Kalush, F., Karpen, G.H., Ke, Z., Kennison, J.A., Ketchum, K.A., Kimmel, B.E., Kodira, C.D., Kraft, C., Kravitz, S., Kulp, D., Lai, Z., Lasko, P., Lei, Y., Levitsky, A.A., Li, J., Li, Z., Liang, Y., Lin, X., Liu, X., Mattei, B., McIntosh, T.C., McLeod, M.P., McPherson, D., Merkulov, G., Milshina, N.V., Mobarry, C., Morris, J., Moshrefi, A., Mount, S.M., Moy, M., Murphy, B., Murphy, L., Muzny, D.M., Nelson, D.L., Nelson, D.R., Nelson, K.A., Nixon, K., Nusskern, D.R., Pacleb, J.M., Palazzolo, M., Pittman, G.S., Pan, S., Pollard, J., Puri, V., Reese, M.G., Reinert, K., Remington, K., Saunders, R.D., Scheeler, F., Shen, H., Shue, B.C., Sidén-Kiamos, I., Simpson, M., Skupski, M.P., Smith, T., Spier, E., Spradling, A.C., Stapleton, M., Strong, R., Sun, E., Svirskas, R., Tector, C., Turner, R., Venter, E., Wang, A.H., Wang, X., Wang, Z.Y., Wassarman, D.A., Weinstock, G.M., Weissenbach, J., Williams, S.M., Woodage, T., Worley, K.C., Wu, D., Yang, S., Yao, Q.A., Ye, J., Yeh, R.F., Zaveri, J.S., Zhan, M., Zhang, G., Zhao, Q., Zheng, L., Zheng, X.H., Zhong, F.N., Zhong, W., Zhou, X., Zhu, S., Zhu, X., Smith, H.O., Gibbs, R.A., Myers, E.W., Rubin, G.M., Venter, J.C., 2000. The genome sequence of *Drosophila melanogaster*. Science 287, 2185–2195.

Allada, R., Chung, B.Y., 2010. Circadian organization of behavior and physiology in *Drosophila*. Annu. Rev. Physiol. 72, 605–624.

Bae, K., Lee, C., Sidote, D., Chuang, K.Y., Edery, I., 1998. Circadian regulation of a *Drosophila* homolog of the mammalian Clock gene: PER and TIM function as positive regulators. Mol. Cell. Biol. 18, 6142–6151.

Blitz, D.M., Nusbaum, M.P., 2011. Neural circuit flexibility in a small sensorimotor system. Curr. Opin. Neurobiol. 21, 544–552.

Chabot, C.C., Webb, L.K., 2008. Circadian rhythms of heart rate in freely moving and restrained American lobsters, *Homarus americanus*. Mar. Freshw. Behav. Physiol. 41, 29–41.

Chen, S., Qiao, H., Fu, H., Sun, S., Zhang, W., Jin, S., Gong, Y., Jiang, S., Xiong, W., Wu, Y., 2017. Molecular cloning, characterization, and temporal expression of the clock genes period and timeless in the oriental river prawn *Macrobrachium nipponense* during female reproductive development. Comp. Biochem. Physiol. A Mol. Integr. Physiol. 207, 43–51.

Chesmore, K.N., Watson 3rd, W.H., Chabot, C.C., 2016. Identification of putative circadian clock genes in the American horseshoe crab, *Limulus polyphemus*. Comp. Biochem. Physiol. Part D Genomics Proteomics 19, 45–61.

Christie, A.E., Stemmler, E.A., Dickinson, P.S., 2010. Crustacean neuropeptides. Cell. Mol. Life Sci. 67, 4135–4169.

Christie, A.E., Fontanilla, T.M., Nesbit, K.T., Lenz, P.H., 2013. Prediction of the protein components of a putative *Calanus finmarchicus* (Crustacea, Copepoda) circadian signaling system using a *de novo* assembled transcriptome. Comp. Biochem. Physiol. Part D Genomics Proteomics 8, 165–193.

Christie, A.E., Chi, M., Lameyer, T.J., Pascual, M.G., Shea, D.N., Stanhope, M.E., Schulz, D.J., Dickinson, P.S., 2015. Neuropeptidergic signaling in the American lobster *Homarus americanus*: new insights from high-throughput nucleotide sequencing. PLoS One 10, e0145964.

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., Yu, A., Pascual, M.G., 2018. Circadian signaling in the northern krill *Meganyctiphanes norvegica*: prediction of the protein components of a putative clock system using a publicly accessible transcriptome. Mar. Genomics 37, 97–113.

Dombrádi, V., Axton, J.M., Brewis, N.D., da Cruz e Silva, E.F., Alphey, L., Cohen, P.T., 1990. *Drosophila* contains three genes that encode distinct isoforms of protein phosphatase 1. Eur. J. Biochem. 194, 739–745.

Filichkin, S.A., Mockler, T.C., 2012. Unproductive alternative splicing and nonsense mRNAs: a widespread phenomenon among plant circadian clock genes. Biol. Direct 7, 20.

Finn, R.D., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Mistry, J., Mitchell, A.L., Potter, S.C., Punta, M., Qureshi, M., Sangrador-Vegas, A., Salazar, G.A., Tate, J., Bateman, A., 2016. The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res. 44, D279–D285.

Goergen, E.M., Bagay, L.A., Rehm, K., Benton, J.L., Beltz, B.S., 2002. Circadian control of neurogenesis. J. Neurobiol. 53, 90–95.

Golombek, D.A., Bussi, I.L., Agostino, P.V., 2014. Minutes, days and years: molecular interactions among different scales of biological timing. Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci. 369, 20120465.

Grabek, K.R., Chabot, C.C., 2012. Daily rhythms of PERIOD protein in the eyestalk of the American lobster, *Homarus americanus*. Mar. Freshw. Behav. Physiol. 45, 269–279.

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

Gramates, L.S., Marygold, S.J., Santos, G.D., Urbano, J.M., Antonazzo, G., Matthews, B.B., Rey, A.J., Tabone, C.J., Crosby, M.A., Emmert, D.B., Falls, K., Goodman, J.L., Hu, Y., Ponting, L., Schroeder, A.J., Strelets, V.B., Thurmond, J., Zhou, P., the FlyBase Consortium, 2017. FlyBase at 25: looking to the future. Nucleic Acids Res. 45, D663–D671.

Hardin, P.E., 2011. Molecular genetic analysis of circadian timekeeping in *Drosophila*. Adv. Genet. 74, 141–173.

Hooper, S.L., DiCaprio, R.A., 2004. Crustacean motor pattern generator networks. Neurosignals 13, 50–69.

Hut, R.A., Beersma, D.G., 2011. Evolution of time-keeping mechanisms: early emergence and adaptation to photoperiod. Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci. 366, 2141–2154.

Jones, D.T., Taylor, W.R., Thornton, J.M., 1992. The rapid generation of mutation data matrices from protein sequences. Comput. Appl. Biosci. 8, 275–282.

Jury, S.H., Chabot, C.C., Watson, W.H., 2005. Daily and circadian rhythms of locomotor activity in the American lobster, *Homarus americanus*. J. Exp. Mar. Biol. Ecol. 318, 61–70.

Katoh, K., Standley, D.M., 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., 2012. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics (12), 1647–1649.

Kwon, Y.J., Park, M.J., Kim, S.G., Baldwin, I.T., Park, C.M., 2014. Alternative splicing and nonsense-mediated decay of circadian clock genes under environmental stress conditions in *Arabidopsis*. BMC Plant Biol. 14, 136.

Langmead, B., Trapnell, C., Pop, M., Salzberg, S.L., 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 10, R25.

Ma, M., Chen, R., Sousa, G.L., Bors, E.K., Kwiatkowski, M.A., Goiney, C.C., Goy, M.F., Christie, A.E., Li, L., 2008. Mass spectral characterization of peptide transmitters/hormones in the nervous system and neuroendocrine organs of the American lobster *Homarus americanus*. Gen. Comp. Endocrinol. 156, 395–409.

Marder, E., Bucher, D., 2007. Understanding circuit dynamics using the stomatogastric nervous system of lobsters and crabs. Annu. Rev. Physiol. 69, 291–316.

Marder, E., Gutierrez, G.J., Nusbaum, M.P., 2017. Complicating connectomes: electrical coupling creates parallel pathways and degenerate circuit mechanisms. Dev. Neurobiol. 77, 597–609.

Mat, A.M., Perrigault, M., Massabuau, J.C., Tran, D., 2016. Role and expression of cry1 in the adductor muscle of the oyster *Crassostrea gigas* during daily and tidal valve activity rhythms. Chronobiol. Int. 33, 949–963.

Mazzotta, G.M., De Pittà, C., Benna, C., Tosatto, S.C., Lanfranchi, G., Bertolucci, C., Costa, R., 2010. A cry from the krill. Chronobiol. Int. 27, 425–445.

Mendoza-Viveros, L., Bouchard-Cannon, P., Hegazi, S., Cheng, A.H., Pastore, S., Cheng, H.M., 2017. Molecular modulators of the circadian clock: lessons from flies and mice. Cell. Mol. Life Sci. 74, 1035–1059.

Montelli, S., Mazzotta, G., Vanin, S., Caccin, L., Corrà, S., De Pittà, C., Boothroyd, C., Green, E.W., Kyriacou, C.P., Costa, R., 2015. *period* and *timeless* mRNA splicing profiles under natural conditions in *Drosophila melanogaster*. J. Biol. Rhythm. 30, 217–227.

Moriyama, Y., Sakamoto, T., Karpova, S.G., Matsumoto, A., Noji, S., Tomioka, K., 2008. RNA interference of the clock gene period disrupts circadian rhythms in the cricket *Gryllus bimaculatus*. J. Biol. Rhythm. 23, 308–318.

Myers, M.P., Wager-Smith, K., Wesley, C.S., Young, M.W., Sehgal, A., 1995. Positional cloning and sequence analysis of the *Drosophila* clock gene, timeless. Science 270, 805–808.

Nelms, B.L., Labosky, P.A., 2011. A predicted hairpin cluster correlates with barriers to PCR, sequencing and possibly BAC recombineering. Sci. Rep. 1, 56–57.

Nesbit, K.T., Christie, A.E., 2014. Identification of the molecular components of a *Tigriopus californicus* (Crustacea, Copepoda) circadian clock. Comp. Biochem. Physiol. Part D Genom. Proteomics 12, 16–44.

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, 868.

Nusbaum, M.P., Blitz, D.M., Swensen, A.M., Wood, D., Marder, E., 2001. The roles of co-transmission in neural network modulation. Trends Neurosci. 24, 146–154.

O'Grady, J.F., Hoelters, L.S., Swain, M.T., Wilcockson, D.C., 2016. Identification and temporal expression of putative circadian clock transcripts in the amphipod crustacean *Talitrus saltator*. PeerJ. 4, e2555.

Otopalik, A.G., Lane, B., Schulz, D.J., Marder, E., 2017. Innexin expression in electrically coupled motor circuits. Neurosci. Lett (In press).

Overton, P., 2017. Maine Lobster Catch Tipped the Scale at a Record 130 million Pounds in 2016. Portland Press Herald. http://www.pressherald.com/2017/03/03/maine-lobster-landings-set-records-in-2016/, Accessed date: 1 August 2017.

Ozkaya, O., Rosato, E., 2012. The circadian clock of the fly: a neurogenetics journey through time. Adv. Genet. 77, 79–123.

Pegoraro, M., Gesto, J.S., Kyriacou, C.P., Tauber, E., 2014. Role for circadian clock genes in seasonal timing: testing the Bünning hypothesis. PLoS Genet. 10, e1004603.

Perrigault, M., Tran, D., 2017. Identification of the molecular clockwork of the oyster *Crassostrea gigas*. PLoS One 12, e0169790.

Roncalli, V., Christie, A.E., Sommer, S., Cieslak, M.C., Hartline, D.K., Lenz, P.H., 2017. A deep transcriptomic resource for the copepod crustacean *Labidocera madurae*: a potential indicator species for assessing near shore ecosystem health. PLoS One 12, e0186794.

Rosato, E., Trevisan, A., Sandrelli, F., Zordan, M., Kyriacou, C.P., Costa, R., 1997. Conceptual translation of timeless reveals alternative initiating methionines in *Drosophila*. Nucleic Acids Res. 25, 455–458.

Rozen, S., Skaletsky, H., 2000. Primer3 on the WWW for general users and for biologist

programmers. Methods Mol. Biol. 132, 365–386.
Sbragaglia, V., Lamanna, F., Mat, A.M., Rotllant, G., Joly, S., Ketmaier, V., de la Iglesia, H.O., Aguzzi, J., 2015. Identification, characterization, and diel pattern of expression of canonical clock genes in *Nephrops norvegicus* (Crustacea: Decapoda) eyestalk. PLoS One 10, e0141893.
Schoenhard, J.A., Eren, M., Johnson, C.H., Vaughan, D.E., 2002. Alternative splicing yields novel BMAL2 variants: tissue distribution and functional characterization. Am. J. Phys. Cell Phys. 283, C103–C114.
Schulz, D.J., Lane, B.J., 2017. Homeostatic plasticity of excitability in crustacean central pattern generator networks. Curr. Opin. Neurobiol. 43, 7–14.
Shafer, O.T., Yao, Z., 2014. Pigment-dispersing factor signaling and circadian rhythms in insect locomotor activity. Curr. Opin. Insect Sci. 1, 73–80.
Shin, S.W., Zou, Z., Saha, T.T., Raikhel, A.S., 2012. bHLH-PAS heterodimer of methoprene-tolerant and Cycle mediates circadian expression of juvenile hormone-induced mosquito genes. Proc. Natl. Acad. Sci. U. S. A. 109, 16576–16581.
Staiger, D., Shin, J., Johansson, M., Davis, S.J., 2013. The circadian clock goes genomic. Genome Biol. 14, 208.
Stein, W., 2009. Modulation of stomatogastric rhythms. J. Comp. Physiol. A Neuroethol. Sens. Neural Behav. Physiol. 195, 989–1009.
Strauss, J., Dircksen, H., 2010. Circadian clocks in crustaceans: identified neuronal and cellular systems. Front. Biosci. (Landmark Ed). 15, 1040–1074.
Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S., 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. Mol. Biol. Evol. 30, 2725–2729.
Tanaka, Y., Suetsugu, Y., Yamamoto, K., Noda, H., Shinoda, T., 2014. Transcriptome analysis of neuropeptides and G-protein coupled receptors (GPCRs) for neuropeptides in the brown planthopper *Nilaparvata lugens*. Peptides 53, 125–133.
Tassone, E.E., Geib, S.M., Hall, B., Fabrick, J.A., Brent, C.S., Hull, J.J., 2016. *De novo* construction of an expanded transcriptome assembly for the Western tarnished plant bug, *Lygus hesperus*. Gigascience 5, 6.
Tauber, E., Zordan, M., Sandrelli, F., Pegoraro, M., Osterwalder, N., Breda, C., Daga, A., Selmin, A., Monger, K., Benna, C., Rosato, E., Kyriacou, C.P., Costa, R., 2007. Natural selection favors a newly derived timeless allele in *Drosophila melanogaster*. Science 316, 1895–1898.
Tilden, A.R., McCoole, M.D., Harmon, S.M., Baer, K.N., Christie, A.E., 2011. Genomic identification of a putative circadian system in the cladoceran crustacean *Daphnia pulex*. Comp. Biochem. Physiol. Part D Genomics Proteomics. 6, 282–309.
Tomioka, K., Matsumoto, A., 2010. A comparative view of insect circadian clock systems. Cell. Mol. Life Sci. 67, 1397–1406.
Tribolium Genome Sequencing Consortium, 2008. The genome of the model beetle and pest *Tribolium castaneum*. Nature 452, 949–955.
Uryu, O., Kamae, Y., Tomioka, K., Yoshii, T., 2013. Long-term effect of systemic RNA interference on circadian clock genes in hemimetabolous insects. J. Insect Physiol. 59, 494–499.
Wildt, M., Goergen, E.M., Benton, J.L., Sandeman, D.C., Beltz, B.S., 2004. Regulation of serotonin levels by multiple light-entrainable endogenous rhythms. J. Exp. Biol. 207, 3765–3774.
Yang, J.S., Dai, Z.M., Yang, F., Yang, W.J., 2006. Molecular cloning of clock cDNA from the prawn, *Macrobrachium rosenbergii*. Brain Res. 1067, 13–24.
Yoshii, T., Hermann-Luibl, C., Helfrich-Förster, C., 2015. Circadian light-input pathways in *Drosophila*. Commun. Integr. Biol. 9, e1102805.
Yuan, Q., Metterville, D., Briscoe, A.D., Reppert, S.M., 2007. Insect cryptochromes: gene duplication and loss define diverse ways to construct insect circadian clocks. Mol. Biol. Evol. 24, 948–955.
Zhan, S., Merlin, C., Boore, J.L., Reppert, S.M., 2011. The monarch butterfly genome yields insights into long-distance migration. Cell 147, 1171–1185.
Zhang, L., Hastings, M.H., Green, E.W., Tauber, E., Sladek, M., Webster, S.G., Kyriacou, C.P., Wilcockson, D.C., 2013. Dissociation of circadian and circatidal timekeeping in the marine crustacean *Eurydice pulchra*. Curr. Biol. 23, 1863–1873.
Zhu, H., Yuan, Q., Briscoe, A.D., Froy, O., Casselman, A., Reppert, S.M., 2005. The two CRYs of the butterfly. Curr. Biol. 15, R953–R954.