Marine Genomics xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Marine Genomics

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

Method paper

Circadian signaling in the Northern krill *Meganyctiphanes norvegica*: *In silico* prediction of the protein components of a putative clock system using a publicly accessible transcriptome

Andrew E. Christie*, Andy Yu, Micah G. Pascual

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

ARTICLE INFO

Keywords:
Bioinformatics
Circadian rhythm
Transcriptome shotgun assembly
Crustacea
Malacostraca
Euphausiacea

ABSTRACT

The Northern krill *Meganyctiphanes norvegica* is a significant component of the zooplankton community in many regions of the North Atlantic Ocean. In the areas it inhabits, *M. norvegica* is of great importance ecologically, as it is both a major consumer of phytoplankton/small zooplankton and is a primary food source for higher-level consumers. One behavior of significance for both feeding and predator avoidance in *Meganyctiphanes* is diel vertical migration (DVM), *i.e.*, a rising from depth at dusk and a return to depth at dawn. In this and other euphausiids, an endogenous circadian pacemaker is thought, at least in part, to control DVM. Currently, there is no information concerning the identity of the genes/proteins that comprise the *M. norvegica* circadian system. In fact, there is little information concerning the molecular underpinnings of circadian rhythmicity in crustaceans generally. Here, a publicly accessible transcriptome was used to identify the molecular components of a putative *Meganyctiphanes* circadian system. A complete set of core clock proteins was deduced from the *M. norvegica* transcriptome (clock, cryptochrome 2, cycle, period and timeless), as was a large suite of proteins that likely function as modulators of the core clock (*e.g.*, doubletime), or serves as inputs to it (cryptochrome 1) or outputs from it (pigment dispersing hormone). This is the first description of a "complete" (core clock through putative output pathway signals) euphausiid clock system, and as such, provides a foundation for initiating molecular investigations of circadian signaling in *M. norvegica* and other krill species, including how clock systems may regulate DVM and other behaviors.

1. Introduction

Krill, small shrimp-like members of the crustacean order Euphausiacea, are found throughout the world's oceans (*e.g.*, Everson, 2008). In Northern and Southern seas, these crustaceans are of critical importance ecologically (*e.g.*, Miller and Hampton, 1989; Tarling et al., 2010), as in many regions they constitute a major portion of the biomass (*e.g.*, Ross and Quetin, 1988; Simard and Lavoie, 1999), and occupy a key tropic position, consuming both phytoplankton and small zooplankton and converting these microscopic foodstuffs into forage of a size suitable for larger, higher-level consumers, including a wide array of fishes, seabirds and marine mammals (*e.g.*, Schmidt, 2010; Schmidt and Atkinson, 2016; Simard and Harvey, 2010). Global climate change, ocean acidification and a variety of other environmental and anthropogenic stressors, including commercial fishing for members of Euphausiacea (*e.g.*, Nicol and Foster, 2016), have contributed to dramatic changes in the abundance and/or distribution of krill, often with profound consequences for the ecology of the affected areas (*e.g.*, Flores et al., 2012). For example, in portions of the Southern Ocean, changes in krill abundance are hypothesized to have contributed to changes in the population sizes of a variety of krill-dependent predators (*e.g.*, Trivelpiece et al., 2011).

Diel vertical migration (DVM) within the water column is a well-known behavior in many euphausiids (*e.g.*, Godlewska, 1996; Kaartvedt, 2010; Liu and Sun, 2010). Simplistically speaking, DVM in krill consists of assent from depth at dusk for surface feeding during the night, and a return to depth at dawn (*e.g.*, Kaartvedt, 2010).The evolution of this migratory behavior is attributed to the trade off between

Abbreviations: CKII, casein kinase II; CLK, clock; CWO, clockwork orange; CRY1, cryptochrome 1; CRY2, cryptochrome 2; CYC, cycle; DVM, diel vertical migration; DBT, doubletime; JET, jetlag; Megno, *Meganyctiphanes norvegica*; PDP1, PAR-domain protein 1; PER, period; PDF, pigment dispersing factor; PDFR, pigment dispersing factor receptor; PDH, pigment dispersing hormone; PDHR, pigment dispersing hormone receptor; PP1, protein phosphatase 1; PP2A, protein phosphatase 2A; SGG, shaggy; SLIMB, supernumerary limbs; TIM, timeless; TSA, transcriptome shotgun assembly; VRI, vrille

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

http://dx.doi.org/10.1016/j.margen.2017.09.001
Received 19 July 2017; Received in revised form 1 September 2017; Accepted 2 September 2017

feeding success and risk for predation, both of which are highest in near surface waters, particularly during daylight hours (*e.g.*, Kaartvedt, 2010). Given the biomass of krill in many locations, the DVM of these animals, and other vertically migrating zooplankton, is recognized as a major contributor to carbon transfer between surface waters and the deep ocean (Ducklow et al., 2001).

In many planktonic animals, including a variety of crustacean species, DVM has been shown to be under circadian control (*e.g.*, Cohen and Forward, 2009). Interestingly, in krill (*e.g.*, *Euphausia superba*), which have been shown to possess robust circadian rhythms in several areas of their physiology (*e.g.*, De Pittà et al., 2013; Teschke et al., 2011), it is still unclear as to the extent to which a circadian pacemaker controls DVM, and how it is modulated by environmental variables and/or interactions with other physiological/behavioral control systems (Gaten et al., 2008). One species of krill that exhibits pronounced DVM is the Northern krill *Meganyctiphanes norvegica* (*e.g.*, Kaartvedt, 2010), a native of the North Atlantic Ocean and Mediterranean Sea (*e.g.*, Tarling et al., 2010). Interestingly, the DVM of *M. norvegica* appears quite flexible in its expression, with a number of environmental and physiological variables appearing to play significant roles in modulating it (*e.g.*, Kaartvedt, 2010). Given the apparent complexity of its DVM, *M. norvegica* represents an intriguing species in which to investigate the molecular underpinnings of circadian control, and the mechanisms through which environmental and physiological factors may modulate the expression of its diel migratory behavior.

The lack of knowledge concerning circadian control in krill, and for that matter, crustaceans in general, is due, at least in part, to the relative paucity of sequence information that is currently available concerning the identity of the genes and proteins that underlie timekeeping systems in these animals (Chen et al., 2017; Christie et al., 2013a, 2017a; 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, investigations directed at identifying complete sets of circadian signaling system genes/proteins have been attempted for just a handful of crustacean species (Christie et al., 2013a, 2017a; Nesbit and Christie, 2014; O'Grady et al., 2016; Roncalli et al., 2017; Sbragaglia et al., 2015; Tilden et al., 2011). The recent (2016) public deposition of an ~400,000-sequence transcriptome for *M. norvegica* (BioProject No. PRJNA324094; Blanco-Bercial and Maas, 2017), which is the largest publicly accessible transcriptomic dataset currently extant for any member of the Euphausiacea, provides a powerful resource for beginning to investigate circadian rhythmicity in this species, including its role in DVM. As the first step in understanding the molecular underpinnings of circadian rhythmicity in *M. norvegica*, homology-based BLAST searches of the above mentioned transcriptome were used to identify genes/proteins that may contribute to the circadian clock of this species, *i.e.*, core clock, clock-associated, clock input pathway and clock output pathway genes/proteins.

2. Materials and methods

2.1. Transcriptome mining

Searches of the *M. norvegica* transcriptome were conducted using methods modified from a protocol used previously for the discovery of circadian genes/proteins in other crustaceans (Christie et al., 2013a, 2013b, 2017a; Nesbit and Christie, 2014; Roncalli et al., 2017; Tilden et al., 2011). Specifically, the database of the online program tblastn (National Center for Biotechnology Information, Bethesda, MD; http://blast.ncbi.nlm.nih.gov/Blast.cgi) was set to "Transcriptome Shotgun Assembly (TSA)" and restricted to data from "*Meganyctiphanes norvegica* (taxid:48144)". Known circadian proteins, primarily those from the fruit fly *Drosophila melanogaster*, 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 Table 1.

2.2. Confirmation of protein identifications

A workflow developed to vet the identification of a variety of proteins, including those involved in circadian signaling (*e.g.*, Christie et al., 2013a, 2013b, 2017a; Nesbit and Christie, 2014; Roncalli et al., 2017; Tilden et al., 2011), was used to characterize the sequences deduced from the *M. norvegica* 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 functional start methionine and are flanked on their C-termini by a stop codon.Proteins described here as "partial" lack a start methionine (referred to as C-terminal partial proteins), a stop codon (referred to as N-terminal partial proteins), or both of these features (referred to as internal fragment 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 *Meganyctiphanes* protein was used as the input query in a BLAST search of annotated *Drosophila* protein dataset present in FlyBase (version FB2016_05; Attrill et al., 2016). It should be noted that for cryptochrome 1 (CRY1) and cryptochrome 2 (CRY2), proteins from the monarch butterfly *Danaus plexippus* were used as the original query sequences, as this species possesses both proteins, while *D. melanogaster* lacks CRY2 (*e.g.*, Yuan et al., 2007); for the putative *M. norvegica* 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 *Meganyctiphanes* sequence was subsequently determined by conducting a BLAST search of the non-redundant arthropod protein dataset curated at NCBI (taxid:6656) using each of the deduced *M. norvegica* proteins as the input query (searches conducted on or before June 17, 2017).Finally, protein structural motifs were analyzed for each of the *M. norvegica* proteins using the online program Pfam (http://pfam.xfam.org/; Finn et al., 2016) version 29.0.

To determine amino acid identity/similarity between proteins, the sequences in question were aligned using the online program MAFFT version 7 (http://mafft.cbrc.jp/alignment/software/; Katoh and Standley, 2013), and amino acid identity/similarity subsequently determined using the alignment output. Specifically, percent identity was calculated as the number of identical amino acids divided by the total number of residues in the longest sequence ($\times 100$). Amino acid similarity was calculated as the number of identical and similar amino acids divided by the total number of residues in longest sequence ($\times 100$).

3. Results

3.1. In silico identification of circadian transcripts/proteins

The fruit fly *D. melanogaster*, a member of the Arthropoda, like *M. norvegica*, has arguably the most thoroughly investigated circadian signaling system in the animal kingdom (*e.g.*, Allada and Chung, 2010; Hardin, 2011; Mendoza-Viveros et al., 2017; Ozkaya and Rosato, 2012; Yoshii et al., 2015), though its clock is not necessarily the same as the "ancestral-type" that has been proposed for insects (*e.g.*, Tomioka and Matsumoto, 2010; Yuan et al., 2007). Here, the molecular components that make up the *Drosophila* clock were used as the primary template for circadian gene/protein discovery in *Meganyctiphanes*, the cryptochromes being the exception, with proteins from the monarch butterfly *D. plexippus* used for those searches.

Regardless of species, circadian systems are typically thought to consist of four components (*e.g.*, Allada and Chung, 2010): the core clock (genes/proteins that establish the basic molecular cascade required for ~24 h cyclical timing), clock-associated proteins (which modulate core clock function), clock input pathway proteins (a means for environmental input to the core clock) and clock output pathway proteins (hormones/receptors that allow for timing information to be transmitted from clock cells to effect physiology and behavior). Here,